

Depressive Belastungen unter griechischen
Schulkindern
Eine populationsbeschreibende Studie.

Dissertation
zur Erlangung des akademischen Grades
Doktorin der Philosophie
(Dr. phil.)

vorgelegt von
Dipl. Psychologin
Christina Kampisiou

Erstgutachter: Prof. Dr. Erich Perlwitz

Zweitgutachter: Prof. Dr. Anna Auckenthaler

Datum der Disputation: 17.07.2008

*Mit Dankbarkeit an allen meinen Eltern,
meinen lieblichen – Anna und Giorgo –
und meinem geistigen Vater Prof. Dr. Perlwitz*

Inhaltsverzeichnis

Kapitel	Seite
1. Die Verdeutlichung gesellschaftlicher Notstände über eine durch Daten aus Stichproben vermittelte Populationsbeschreibung	1
2. Der aktuelle Forschungsstand über Depressionen bei griechischen Kindern und Jugendlichen	10
3. Bestimmung des Objekts einer Untersuchung zur Auftretenshäufigkeit depressiver Störungen bei griechischen Schulkindern	24
3.1 Depressive Störungen bei Kindern und Jugendlichen als historischen Gegenstand	25
3.2 Definition und Klassifikation depressiver Störungen bei Kindern und Jugendlichen	29
3.3 Diagnostik affektiver Störungen	43
3.4 Gegenstand der auf die Erfassung von Prävalenzen für affektive Störungen bei Kindern und Jugendlichen gerichteten Untersuchungen	49
3.5 Bestimmungen des Gegenstandes einer epidemiologischen Untersuchung ohne unmittelbare Bezugnahme auf empirisch verlässliche Diagnosen	51
3.6 Die Bestimmung des Untersuchungsgegenstandes für die durchzuführende epidemiologische Untersuchung mit griechischen Schulkindern	60
4. Der unter gegebenen Restriktionen gewählte Ansatz zur Suche nach Hinweisen zum potentiellen Auftreten depressiver Störungen bei griechischen Schulkindern	63
4.1 Von der Ausgangsproblematik zur Bestimmung depressionsspezifischer Informationsträger als Untersuchungsobjekte	64
4.2 Die auf den eingesetzten Untersuchungsgegenstand und auf die dazugehörige Forschungsfrage zu beziehende Durchführung der Datenaufnahme	71
5. Die Realisierung der empirischen Studie zum Auftreten depressiver Belastungen unter griechischen Schulkindern	76
5.1 Die Durchführung der Studie	76
5.2 Die untersuchte Stichprobe	80
5.2.1 Die Stichprobe der Schülerinnen und Schüler	80
5.2.2 Die Stichprobe der Lehrkräfte	86
6. Die Messqualitäten der in dieser Untersuchung verwendeten Untersuchungsinstrumente	88
6.1 Kriterien der Bewertung von Messqualitäten psychologischer Untersuchungsinstrumente	89
6.2 Die Messqualitäten des DTK	91

6.2.1	Messqualitäten des DTK in Studien außerhalb der griechischen Untersuchung 2034GR	91
6.2.2	Messqualitäten des DTK bei seiner griechischen Anwendung in der Untersuchung 2034GR	94
6.2.2.1	Itemanalyse	94
6.2.2.2	Interne Konsistenz der DTK-Subskalen	101
6.2.2.3	Faktorielle Struktur des DTK	106
6.2.3	Zur Vertrauenswürdigkeit der mit dem DTK in dieser griechischen Untersuchung 2034GR erhobenen Daten	109
6.3	Messqualitäten des CDI	110
6.3.1	Messqualitäten des CDI in Studien außerhalb der griechischen Untersuchung 2034GR	111
6.3.2	Messqualitäten des CDI bei seiner griechischen Anwendung in der Untersuchung 2034GR	114
6.3.2.1	Itemanalyse der griechischen Version des CDI	114
6.3.2.2	Interne Konsistenz	118
6.3.2.3	Faktorenstruktur der griechischen Form des CDI	120
6.3.3	Zur Vertrauenswürdigkeit der mit dem CDI in dieser griechischen Anwendung erhobenen Daten	124
6.4	Messqualitäten des ESDM	124
6.4.1	Die Skala ESDM und ihre Bewährung in voran gegangenen Untersuchungen	124
6.4.2	Messqualitäten des ESDM bei seiner griechischen Anwendung in der Untersuchung 2034GR	126
6.4.2.1	Itemanalyse	127
6.4.2.2	Interne Konsistenz der ESDM-Skala	130
6.4.2.3	Faktorielle Struktur des ESDM	131
6.4.3	Zur Vertrauenswürdigkeit der mit dem ESDM in der griechischen Anwendung von Kampisiou (2034) erhobenen Daten	134
6.5	Messqualitäten des SDQ	135
6.5.1	Messqualitäten des SDQ in voran gegangenen Untersuchungen	135
6.5.2	Messqualitäten des SDQ bei seiner griechischen Anwendung in der Untersuchung 2034GR	136
6.5.2.1	Itemanalyse des SDQ	136
6.5.2.2	Interne Konsistenz	140
6.5.2.3	Faktorenanalyse aus der Anwendung des SDQ innerhalb dieser Untersuchung	142
6.5.3	Zur Vertrauenswürdigkeit der mit dem SDQ in dieser griechischen Anwendung erhobenen Daten	146
6.6	Zur Frage nach der Messqualitäten der eingesetzten Fragenliste FIL	146
7.	Analyse der mit dem DTK aufgenommenen Daten	148
7.1	Der DTK als Messverfahren zur Bestimmung von Hinweisen auf unterschiedliche Ausprägungen depressiver Störungen	162

7.2	Die in der Untersuchung 2034GR verfügbare Datenbasis der Analysearbeit zum DTK	167
7.3	Analyse der über die Skala 1 des DTK in 2034 GR erhobenen Daten	168
7.3.1	Verteilung der über die Skala 1 des DTK bekannt werdenden skalenspezifischen Gesamtpunktwerte (S1-GPW)	169
7.3.2	Der individuelle Gesamtpunktwerte aus der Skala 1 des DTK als Ausprägung des Skalenkonstruktes in der Persönlichkeit (S1-GPW)	173
7.3.3	Die mit den Gesamtpunktwerten auf Skala 1 verbundenen skalenspezifischen Informationen	176
7.3.3.1	Der Gesamtpunkt看wert aus Skala 1 des DTK als Hinweis auf individuelle Ausprägungen von Symptomen nachdem Skalenkonstrukt	176
7.3.3.2	Modell einer verteilungsinternen Bestimmung von Trennwerten	179
7.3.3.3	Informationen aus den Gesamtpunktwerten auf Skala 1 im DTK über das Auftreten klinisch ausgeprägter depressiver Störungen	182
7.3.3.4	Der nach Analyse der Gesamtpunktwerte aus Skala 1 des DTK vorliegenden Informationsstand	184
7.4	Analyse der über die Skala 2 des DTK erhobenen Daten	187
7.4.1	Verteilung der über die Skala 2 des DTK bekannt werdenden skalenspezifischen Gesamtpunktwerte (S2-GPW)	188
7.4.2	Informationen aus den Gesamtpunktwerten der DTK-Skala 2 als Hinweise auf depressive Störungen	193
7.4.2.1	Das Phänomen der hohen Selbstzuschreibung von Symptomen aus der Itemliste der Skala 2 des DTK	194
7.4.2.2	Hinweise der GPW aus Skala 2 des DTK auf depressive Störungen in der untersuchten Stichprobe	196
7.4.2.3	Der nach Analyse der Gesamtpunktwerte aus Skala 2 des DTK vorliegenden Informationsstand	199
7.5	Analyse der über die Skala 3 des DTK in 2034GR erhobenen Daten	200
7.5.1	Verteilung der über die Skala 3 des DTK bekannt werdenden skalenspezifischen Gesamtpunktwerte (S3-GPW)	200
7.5.2	Aus den Gesamtpunktwerten der DTK-Skala 3 ableitbare Hinweise auf das mutmaßliche Auftreten depressiver Störungen in der untersuchten Stichprobe	204
7.5.2.1	Informationen über die Selbstzuschreibung von Symptomen aus der Itemliste der Skala 3	204
7.5.2.2	Hinweise auf das mutmaßliche Auftreten störungsrelevanter Ausprägungen des Konstruktes „Müdigkeit/autonome Reaktionen“	205
7.5.2.3	Informationen aus den Gesamtpunktwerten auf Skala 3 im DTK über Hinweise auf klinisch ausgeprägte depressive Störungen in der untersuchten Stichprobe	206
7.6	Eine Zusammenlegung von zwei Skalen des DTK als zusätzliche Datenbasis: Skala 1 plus Skala 2	207
7.6.1	Die Verteilung der DTK Gesamtpunktwerte aus einer Skalenkombination unter der Bezeichnung „Depressive Belastung“ (DTK-Skala 1 plus 2)	208
8.	Verteilung der DTK-Gesamtpunktwerte in verschiedenen Teilgruppen der untersuchten Stichprobe 2034GR	222
8.1	Geschlechts- und Alterseffekte in Punktwerten aus der Anwendung des DTK	223
8.2	Alters- und Geschlechtseffekte in den DTK-Daten aus der griechischen Untersuchung 2034GR	224
8.2.1	Geschlechts- und Alterseffekte in den Gesamtpunktwerten zur Skala „Dysphorie/Selbstwertprobleme“ (DTK-Skala 1) in der Untersuchung 2034GR	224
8.2.2	Geschlechts- und Alterseffekte in den Gesamtpunktwerten aus der Skala „Agitiertes Verhalten“ (DTK-Skala 2)	

in der Untersuchung 2034 GR)	227
8.2.3 Geschlechts- und Alterseffekte in der Produktion von Gesamtpunktwerten auf der DTK-Skala Müdigkeit/autonome Reaktionen“ in der Untersuchung 2034GR	229
8.3 Geschlechtsspezifische Variationen des nach den DTK-Ergebnissen abzuschätzenden mutmaßlichen Auftretens depressiver Störungen in der untersuchten griechischen Stichprobe	231
8.4 Mit dem Alter auftretende Variationen des nach Ergebnissen aus dem DTK abzuschätzenden Auftretens depressiver Störungen in der untersuchten Stichprobe	233
8.5 Variationen des geschätzten Auftretens depressiver Störungen im Zusammenhang mit Interaktionen von Alter und Geschlecht	235
8.6 Das Wiederholen von Schulklassen als potentieller Risikofaktor für das Auftreten von depressiven Störungen	242
8.7 Die Bedeutung des Wohnortes für die Variation von DTK Gesamtpunktwerten	248
9. Analyse der mit dem CDI aufgenommenen Daten	265
9.1 Das CDI als Messverfahren zur Erfassung depressiver Störungen bei Kindern und Jugendlichen	265
9.2 Die Datenbasis der nachfolgenden Analysearbeit zum CDI	275
9.3 Verteilung der CDI-Gesamtpunktwerte	278
9.4 Die CDI-Gesamtpunktwerte als Informationen über das Vorliegen von Hinweisen auf das Bestehen subklinischer und klinischer Ausprägungen an depressiven Störungen in der untersuchten Stichprobe griechischer Schulkinder	286
10. Beziehungen zwischen CDI-Gesamtpunktwerten und verschiedenen Teilstichproben in dieser Studie	290
10.1 Geschlechts- und Alterseffekte in Punktwerten aus der Anwendung des CDI	290
10.2 Geschlechts- und Alterseffekte in den CDI-Gesamtwerten aus der griechischen Untersuchung 2034GR	291
10.3 Variationen der Gesamtpunktwerte nach dem Geschlecht der Schulkinder	293
10.4 Variationen der CDI Gesamtpunktwerte nach dem Alter der untersuchten Kinder	294
10.5 Variationen der CDI-GPW aus der Interaktion von Alter und Geschlecht der untersuchten Kinder	297
10.6 Das Wiederholen einer Klasse als potentieller Risikofaktor einer erhöhten depressionsspezifischen Vulnerabilität	304
10.7 Der Wohnort der Kinder als möglicher Faktor für Variationen von CDI-Gesamtwerten	306
10.8 Die Befundlage nach der Analyse der Beziehungen zwischen CDI-Gesamtpunktwerten und verschiedenen Teilstichproben in dieser Studie	316
11. Analyse der mit dem SDQ gewonnenen Daten	318
11.1 Das SDQ als Screening Verfahren zur Identifizierung psychisch belasteter Kinder	319
11.2 Die aktuelle Datenbasis der nachfolgenden Analysearbeit aus der Anwendung des SDQ in der Studie 2034GR	323
11.3 Verteilung der Werte der SDQ-Skala „Total difficulties“ in der Untersuchung 2034GR	324
11.4 Die Interkorrelationen zwischen den SDQ Subskalen in den Daten der Stichprobe 2034GR	326
11.5 Die verfahrensinternen korrelativen Beziehung der SDQ Items	335
11.6 Die empirisch gesicherten Beziehung zwischen SDQ, DTK und CDI in der Untersuchung 2034GR:	

Aspekte der konvergenten Validität	339
11.7 Die SDQ-Gesamtwerte als Information über das potentielle Vorliegen subklinischer und klinischer psychischer Beeinträchtigungen in der untersuchten Stichprobe	343
12. Das Auftreten depressiver Symptome bei Schulkindern in der Perspektive von Lehrkräften: Analyse der mit dem ESDM aufgenommenen Daten	346
12.1 Die Lehrkräfte als Informationsquellen zum Auftreten depressiver Störungen bei Schulkindern	346
12.2 Die Skala ESDM als Verfahren zur Erfassung depressiver Symptome aus der Perspektive der Lehrkräfte	352
12.3 Die Datenbasis der Analysearbeit zum ESDM	354
12.4 Verteilung der ESDM-Punktwerte	355
12.5 Die Informationen über das Auftreten depressiver Störungen nach den Einschätzungen der Lehrkräfte und den Selbstauskünften ihrer Schul Kinder	365
12.5.1 Korrelationen auf der Ebene von Gesamtpunktwerten aus den eingesetzten Verfahren	366
12.5.2 Übereinstimmung zwischen Urteilen der Lehrkräfte im ESDM und Selbstangaben der Kinder zu ausgewählten CDI Items	381
12.6 Weitere Einflussfaktoren auf die Lehrurteile im ESDM	391
12.6.1 Potentieller Einfluss des Geschlechts der Lehrkräfte auf die Urteile in der Skala ESDM	392
12.6.2 Die Unterrichtsfächer der Lehrkräfte als potentielle Determinanten oder Moderatoren von urteilen über die Skala ESDM	396
12.6.3 Geschlecht der Kinder	399
12.6.4 Alter oder schulische Klassenstufe der Kinder	405
12.6.5 ESDM Urteile der Lehrkräfte und die von ihnen abgegebenen Leistungsbeurteilungen für die Kinder in ihrem Unterricht	407
12.6.6 Beziehungen zwischen ESDM Urteilen und anderen von den Lehrkräften gegebenen Informationen über die Kinder in ihrem Unterricht	408
12.6.7 ESDM Urteile der Lehrkräfte und Frequenz der Kinder in den Schulklassen	412
12.7 Aspekte der Bewährung des ESDM in der griechischen Untersuchung 2034GR	415
13. Der aus den Teilergebnissen der durchgeführten Untersuchung zu generierende Gesamtbefund	417
13.1 Befunde in der Perspektive der Datenaufnahme als kommunikativer Prozess	420
13.2 Befunde über die Bewährung griechischer Versionen depressionsspezifischer Messinstrumente	428
13.3 Hinweise zum Auftreten depressiver Symptome und depressiver Störungen in der untersuchten Stichprobe: Hinweise über Daten aus einem Messverfahren	430
13.4 Hinweise zum Auftreten depressiver Symptome und depressiver Störungen in der untersuchten Stichprobe: Hinweise über Datenkombinationen aus zwei Messverfahren	449
13.5 Auf personenspezifische Merkmale bezogene Hinweise zum Auftreten depressiver Symptome und depressiver Störungen in der untersuchten Stichprobe	454
13.5.1 Geschlechtsspezifische Variationen von Hinweisen auf depressive Störungen	454

13.5.2	Geschlechtsspezifische Variationen der Hinweise auf depressive Störungen	455
13.5.3	Kombinierte alters- und geschlechtsspezifische Variationen der Hinweise auf depressive Störungen	456
13.5.4	Schulische Merkmale und auf sie bezogene Variationen der Hinweise auf depressive Störungen	456
13.5.5	Wohnort und Variationen der Hinweise auf depressive Störungen	458
13.6	Beantwortung der Frage nach dem Wissen von Lehrkräften über das Auftreten selbstberichteter depressiver Belastungen unter ihren Schulkindern	461
13.7	Faktizität, Nutzbarkeit und potentielle Nützlichkeit der Untersuchungsergebnisse aus der Studie 2034GR	469
	Literaturverzeichnis	475
	Anhang	
	Tabellen- und Abbildungsverzeichnis	491
	Danksagung	497
	Erklärung	499

Kapitel 1

Die Verdeutlichung gesellschaftlicher Notstände über eine durch Daten aus Stichproben vermittelte Populationsbeschreibung

Eine neben anderen konstitutiv bedeutsame Qualität menschlicher Gesellschaften ist die Kommunikation ihrer Angehörigen untereinander. Folgt man Watzlawick, Beavin und Jackson (1969), die Kommunikation als weitgehend identisch mit Interaktion beschreiben, ist die zwangsläufig eine der weiteren Qualitäten. Aus beiden folgert für Gesellschaften der Fluss von Informationen, und wenn zutrifft, dass Kommunikation eine konstitutive Qualität ist, sollte Informationsfluss auch als ein primäres Merkmal menschlicher Gesellschaften gesehen werden. Allen dreien gemeinsam ist deren spezifische funktionale Relation zu ihren Gegenständen. Zu denen gehören auch die Gefahren, von denen Menschen und die Gesellschaften, in denen sie leben, potentiell und aktuell bedroht sein können. Solche Gefahren sind in allen ihren verschiedenartigen Varianten spezielle Anlässe für gesellschaftliche Kommunikation und Interaktion, die jeweils über den sie kennzeichnenden Informationsfluss sozial reguliert und gesteuert werden.

Eines der zur Veranschaulichung der existentiellen Komplexität des entstehenden gesellschaftlichen Informationsflusses über Gefahren und dessen Bezüge sowohl zum Erkennen von Möglichkeiten der wirksamen Bekämpfung von Gefahren als auch zur Planung, Organisation und Realisierung von gesellschaftlicher Gegenwehr geeigneten und eher zufällig ausgewählten historischen Beispiels bietet das im 19. Jahrhundert ausgebrochene Wüten der Cholera in Hamburg. Die ersten damals auftretenden Erkrankungen mögen durchaus als Einzelschicksale verkannt worden sein. Die Unterscheidung einer als individuelles Schicksal anzusehenden Erkrankung einer Persönlichkeit von dem aufkeimenden Ausbruch einer Epidemie ist an die Wahrnehmung der anscheinend blindwütig von Individuum zu Individuum springenden tödlichen Seuche gebunden. Das sind erste und sicherlich auch noch fehlerhafte Stichprobenbeschreibungen durch praktizierende Ärzte, die eine Häufung der Erkrankung in Wohnblöcken, Straßenzügen und Stadtvierteln erkannten. Aus vielen solchen Beschreibungen bildete sich die Möglichkeit des Erkennens der von der Cholera aktuell heimgesuchten Population von Menschen. Die erwies sich über die Populationsbeschreibung als regional abgrenzbar. Erstaunlicherweise nicht betroffen schienen ländliche Randgebiete Hamburgs zu sein. Über die mit der Zahl der Erkrankten vollzogene Populationsbeschreibung ergab sich aber auch die Bestätigung des von einzelnen Ärzten beobachteten Phänomens, dass die Auftretenshäufigkeit der Erkrankungen in unverständlicher Weise den politischen Grenzziehungen folgte. So häuften sich in Straßen, die auf der einen Seite zu Altona und auf der anderen Seite zu Hamburg gehörten, die Erkrankungen und die Zahl der Todesfälle nur auf der Hamburger Straßenseite. Auf einer der Altonaer Straßenseiten trat beispielsweise nicht ein einziger Erkrankungsfall auf, während auf der anderen Straßenseite mehr und mehr der vom Unglück heimgesuchten Straßenachbarn verstarben. Diese, durch die Beschreibung des Krankenstandes und der Todesfälle in den benachbarten städtischen Populationen empirisch unabweisbar und somit offenkundig gewordene Tatsache wurde in dem damaligen durchaus vielschichtigen sozialen und gesellschaftlichen Informationsfluss zu einer viel beachteten Information. Die Aufmerksamkeit, die sie auf sich zog, rührte vor allem aus dem Erleben der kognitiven Dissonanz her, die sich gegen eine rationale Aufklärung des ihr innewohnenden Widerspruchs sperrte. Die Suche nach Antworten auf die Frage, warum die eine Straßenseite von der Seuche heimgesucht wird, die aber nicht auf die andere Straßenseite überspringt, trug mit dazu bei, die in dieser Zeit aus unterschiedlichen Quellen kommende

Trinkwasserversorgung in Hamburg und Altona zu verdächtigen, ursächlich an der Epidemie beteiligt zu sein. Die am Kriterium der auftretenden Fälle von Erkrankungen in Häusern, Straßenzügen, Stadtvierteln und Stadtgebieten orientierte Populationsbeschreibung half zum einen, die Epidemie zu erkennen. Das führte zum gesellschaftlichen Aufmerken und zu Interventionsbemühungen. Die Beschreibung des Fortganges der Epidemie veranlasste die Steigerung der Bemühungen um ihre Eindämmung und Überwindung. Die aber wurde durch die aus vergleichenden Populationsbeschreibungen zwischen Altona und Hamburg erwachsende und dann vor allem auch gesellschaftlich beachtete kognitive Dissonanz und der gelungenen Analyse ihres Informationsgehalts gefördert. Die hier ansetzende Darstellung und Erörterung einer speziellen empirischen Arbeit folgt dem Grundgedanken, dass Populationsbeschreibungen nutzbare Mittel zur Information der Gesellschaft über sie potentiell oder akut bedrohende Gefahren sind. Sie sind nutzbar. Als nützlich können sie sich aber erst erweisen, wenn sie in angemessener Weise gesellschaftlich beachtet werden. Angemessen beachtet sind sie, wenn das ihnen innewohnende Potential zur Minderung von Gefahren zum einen erkannt und zum anderen auch genutzt worden ist.

Gefahren, von denen in Gesellschaften lebende Menschen bedroht werden, treten permanent und auch in variierenden Kombinationen auf. Die Kommunikation über sie ist in nahezu allen bestehenden Gesellschaften institutionalisiert. Das Auftreten von vielen potentiellen Gefahren nötigt Individuen wie auch Gesellschaften zur fortlaufenden Hierarchisierung von Gefahren, zur Setzung von Prioritäten in der Gefahrenbekämpfung und zur Aktualisierung des erforderlichen Wissens, der Wertungen und Entscheidungen. Was aktuell als Gefahr gesehen und beachtet wird, unterliegt der Entwicklung der damit verbundenen regulativen Prozesse.

Die hier begonnene Erörterung ist auf eine empirische Untersuchung bezogen, die sich dem Auftreten depressiver Störungen unter griechischen Schulkindern zuwendet und insbesondere auf die stichproben- und populationsorientierte Erfassung der Verbreitung dieser Störung eingeht. Damit versucht sie mittelbar, die bestehenden Risiken für das Eintreten einer einerseits individuellen und andererseits zugleich auch gesellschaftlichen Betroffenheit abzuschätzen. Depressive Störungen gefährden Kinder in ihrer physischen und psychischen Gesundheit und damit sowohl in ihrem alltäglichen Leben als auch in ihrer Entwicklung (WHO, *Bericht über die gesundheitliche Situation der jungen Menschen in der Europäischen Union*, 2000; online Dokument). Beides hat risikofolle Auswirkungen auf Familien, soziale Umgebung und Gesellschaft (s. auch Laux, 2002). Den nachfolgenden Ausführungen ist das Ziel gesetzt, die an einer Stichprobe von 2034 griechischen Schulkindern durchgeführte Untersuchung in ihren Voraussetzungen, ihrem soziokulturellen und wissenschaftlichen Hintergrund, ihrem Ansatz, in ihrer Durchführung und in ihren Ergebnissen und Befunden darzustellen. Die Untersuchung arbeitet mit dem epidemiologisch eingeführten und bewährten Mittel der durch hinreichende Stichprobendaten empirisch begründeten prospektiven Populationsbeschreibung (s. Bortz und Döring, 2002). Die von Daten aus dazu geeigneten Stichproben auf Merkmale von Populationen schließende Methodik ist spätestens seit der Mitte des 20. Jahrhunderts in sehr unterschiedlichen Anwendungsbereichen verstärkt entwickelt und eingesetzt worden. Ihre Nutzbarkeit erweist sich in vielen Industrieländern, durch hohe Medienbeachtung publikumswirksam demonstriert, aber ungeachtet dessen auch methodisch überzeugend, alljährlich mit den Vorhersagen der Ergebnisse nationaler politischer Wahlen.

Die auf Untersuchungen von Stichproben beruhende Vorhersage von Eigenheiten einer Population ist, um hinreichend treffsicher sein zu können, auf die Einhaltung entsprechender methodischer Standards angewiesen (s. auch Bortz und Döring, 2002). Eine der methodisch unabwiesbaren Voraussetzung ist die Möglichkeit zur Inanspruchnahme einer empirisch

nutzbaren Relation zwischen spezifischen Stichprobendaten und dem jeweils vorherzusagenden Merkmalen der Population. Ausgehend von den Stichproben sollte bekannt geworden sein, welche Eigenheiten der Stichproben für eine Vorhersage auf Eigenheiten der Population mit welchen Wahrscheinlichkeiten für das Erreichen der angestrebten Treffsicherheit genutzt werden können. Für die Vorhersage politischer Wahlergebnisse sind inzwischen Strategien der Datenerhebung bekannt, die so wirksam sind, dass bereits Daten aus Stichproben mit weniger als 2000 Mitgliedern hinreichen, um das Wahlverhalten von Millionen an Wählern mit Abweichungen von weniger als 1% treffen zu können. Vorhersagen mit Daten aus Stichproben von Schulkindern auf die Prävalenzraten für das Auftreten depressiver Störungen in Populationen von Schulkindern erreichen diese Treffsicherheit bislang nicht. Ohne Vorwegnahme von ausstehenden Antworten auf die Frage, ob dies möglich sein kann, sind mehrere methodische Defizite dafür erkennbar, dass auch die Annäherung an treffsichere Vorhersagen sich bereits als schwierig erweist. Behindernd für die Steigerung der Treffsicherheit von Vorhersagen auf die Prävalenzraten für das Auftreten depressiver Störungen bei Schulkindern in Populationen von nationalem Umfang ist bereits die zumeist unüberwindbare Schwierigkeit, den Aufwand für die depressionsspezifische Diagnose aller Schulkinder einer größeren Stichprobe zu erbringen. Dieser Gegebenheit ist die Bevorzugung von empirischen Untersuchungen zuzuschreiben, die von Diagnosen absehen und dann versuchen, über den Einsatz von Screening-Instrumenten zur Aufnahme von Indizien für das Vorliegen depressiver Störungen zu kommen. Dementsprechend können solche Untersuchungen auch nur durch Indizien begründete und mehr oder weniger zutreffende Abschätzungen des Auftretens depressiver Störungen in Populationen bieten. Die Einführung und Nutzung diagnostischer Interviews zur Untersuchung größerer Stichproben ist als Versuch zu sehen, höhere Gewissheit über die Prävalenzraten auch von depressiven Störungen bei Kindern im Schulalter zu erreichen. Wenn der Unterschied zwischen Diagnosen und diagnostischen Interviews eine angemessene Beachtung findet und die inzwischen verfügbaren empirischen Daten analysiert und bewertet werden, ist erkennbar, welche Probleme in dem betrachteten Tätigkeitsbereich weiterhin ungelöst sind. Als ein methodisch relativ unerhebliches, in der empirischen Praxis dagegen meist unüberwindbares Problem bleibt der auch bei diagnostischen Interviews relativ hohe Arbeitsaufwand. Zur Durchführung der im Folgenden zu erörternden Untersuchung standen weder die für depressionsspezifische Diagnosen von 2034 Schulkindern noch die für 2034 diagnostische Interviews erforderlichen Kapazitäten zur Verfügung. Insoweit waren die Bedingungen unter denen die Untersuchung zu führen war, nicht wesentlich anders als die für vergleichbare Vorhaben. Dennoch waren die nur begrenzt einsetzbaren Arbeitskapazitäten nicht der Hauptgrund für die Wahl der eingesetzten Untersuchungsmethode. Drei weitere Argumente wurden dabei einbezogen. Zum einen sind Screening Untersuchungen zurzeit und möglicherweise auch weiterhin nicht ersetzbar. Zum anderen verlangt eben dies weitere Bemühungen um die Erhöhung der Treffsicherheit von Vorhersagen aus solchen Stichproben auf Populationen. Schließlich aber ist eine relative Dringlichkeit des Erfassens von verwertbaren Indizien für das Auftreten depressiver Störungen unter griechischen Schulkindern gesehen worden. Die hier dann zu erörternde Untersuchung versucht, empirisch verlässliche Indizien für die Häufigkeit des Auftretens depressiver Störungen unter griechischen Schulkindern zu erfassen. Sie soll weiterhin zu prüfen versuchen, ob an einer Stichprobe auch Daten zu gewinnen sind, die als Indikatoren für das Auftreten depressiver Störungen dienen könnten. Zwischen Indizien und Indikatoren wird in diesem Zusammenhang ein qualitativer Unterschied postuliert. Während Indizien mittelbar und unter Berücksichtigung der sie qualifizierenden Argumente für das Auftreten von depressiven Störungen in der Population sprechen, sollten für Daten, die als Indikatoren nutzbar sind, über Wahrscheinlichkeiten definierte Relationen zwischen ihrem Auftreten in Stichproben und dem Auftreten depressiver Störungen in Populationen nachgewiesen sein.

Nach den genannten Zielsetzungen ist die hier vorgelegte Arbeit eine vor allem epidemiologische Studie zum Auftreten depressiver Verstimmungen unter griechischen Schulkindern. Diese Benennung des Untersuchungsgegenstandes ist ein Terminus, mit dem das Auftreten unterschiedlicher Ausprägungen des Schweregrades depressiver Symptome unterstellt wird. Anzunehmen ist danach, dass in einer entsprechend großen Stichprobe an Schulkindern der Ausprägungsgrad der jeweiligen individuellen Belastung vom Fehlen depressiver Symptome bis hin zu klinischen Ausprägungen variieren wird. Diese Annahme korrespondiert mit einem, inzwischen wohl schon als historisch zu bezeichnenden Verständnis psychischer Störungen, nach dem die jeweils klassifizierte Störung an die Existenz eines allgemein im Normbereich liegenden Funktionsbereiches gebunden ist. In diesem Sinne beschrieb Ernst Kretschmer (1888-1964) beispielsweise innerhalb der von ihm entwickelten Typenlehre die von ihm beachteten psychiatrischen Störungsbilder, so etwa auch die Schizophrenie, als auf einer Dimension positioniert, deren einer Pol im Bereich allgemeiner und lebensnotwendiger psychischer Funktionalität liegt und deren Gegenpol von schweren psychiatrisch relevanten Störungen besetzt ist. Der im Bereich der Populationsmehrheit nach Kretschmer auftretende schizoide Typ kann im Verlauf seines Lebens eine, auf der selben Dimension liegende Schizophrenie entwickeln. Auch Pawlow (1849-1936) strukturierte seine Typenlehre unter Verwendung von Dimensionen. Die Summe physiologischer Kennwerte entschied über die individuelle Platzierung auf der Typenskala. So verfuhr dann auch Eysenck (1953) der explizit darauf verwies, dass jede Persönlichkeit auf den von ihm unterschiedenen Skalen, so auf der Skala zwischen Extraversion und Introversion zu positionieren sei. Seine über ein Koordinatensystem dargestellte Typologie unterschied sich von anderen durch die Lage des Nullpunktes in der Mitte der Dimensionen. Auch den im DSM-IV und im ICD 10 beschriebenen Störungsbildern unterliegt eine mehr oder weniger verdeutlichte und fixierte Vorstellung von der Dimensionalität der Ausprägung von Störungsbildern. Das trifft auch auf die depressiven Störungen zu. Leichtere Ausprägungen werden als alltäglich mögliche und durchaus erwartbare Episoden gesehen, die auch relativ schnell wieder in ein positives Lebensgefühl umschlagen, eben dies fehlt dann den schwereren Ausprägungen. Für alle von Stichprobendaten ausgehenden Populationsbeschreibungen zum Auftreten psychischer Störungen verlangen dimensional orientierte Störungsbilder auch an Positionen auf solchen Dimensionen orientierte Beschreibungen von Stichproben und Populationen. Das schafft für alle auf psychische Störungen gerichtete Untersuchungen gewichtige Messprobleme, weil die Ausprägung der sie konstituierenden Symptome nicht selbst gemessen werden kann und die Anwendung mittelbarer Meßmethoden verlangt.

Zu einer auf das Auftreten depressiver Symptome und Störungen bezogenen Populationsbeschreibung gehört dann auch die Beschreibung der Ausprägungsgrade an individuellen Belastungen. Das ist als Forderung fast problemlos zu begründen, um so problematischer wird die Realisierung eines solchen Vorhabens, weil kein eingeführtes Maß für die Ausprägung depressiver Symptome verfügbar ist, das objektiv, reliabel und valide genannt werden könnte. Eine Annäherung an die verlässliche Erfassung depressiver Symptome wird über die Manuale der DSM- und der ICD-Ausgaben zu erreichen versucht. Auch die hier geführten Erörterungen und die von ihnen beschriebene empirische Studie sind auf diese Manuale bezogen und, soweit Ausnahmen nicht ausdrücklich angemerkt werden, speziell auf die Ausgaben DSM-IV und ICD-10. Danach ist dann von depressiven Störungen mit klinischer Ausprägung zu sprechen, wenn entsprechende und auf die Manuale bezogene Diagnosen vorliegen. Die Verwendung der in den genannten Handbüchern klassifikatorisch festgelegten Begriffe soll an das Vorliegen von entsprechenden Diagnosen gebunden bleiben, die in ihrer empirischen Absicherung den dort beschriebenen Störungsbildern entsprechen. In der hier zu beschreibenden Studie werden, den Erwartungen vom Auftreten sehr unterschiedlicher Ausprägungen an

depressiven Symptomen entsprechend, auch unterschiedliche Benennungen der jeweils aufgefundenen individuellen Betroffenheit oder Belastung gesucht werden müssen. Nahe liegend wäre, als Grundlage dafür die Daten zu verwenden, die durch den Einsatz von depressionspezifischen Messverfahren erhoben und auch klassifiziert werden können. Solche Daten wären, weil sie nicht unmittelbar Depressivität oder Depression messen, als Indizien für das Vorliegen depressiver Störungen anzusehen, wenn dafür valide Argumente vorliegen sollten. Bei eingeführten depressionspezifischen Fragebögen, so beispielsweise beim CDI (Kovacs, 1992; 2004) werden die verwendeten Fragen als mehr oder weniger inhaltlich identische verbale Repräsentationen von Inhalten depressiver Symptome angesehen. Die dann von den befragten Persönlichkeiten angegebenen unterschiedlichen Summen von Zustimmung zum Auftreten solcher Inhalte lassen sich als Indizien für den Ausprägungsgrad der psychischen Belastung setzen und verwenden. Welche Nützlichkeit das hat, ließe sich beispielsweise durch den Vergleich der Rangreihe von solchen Summen als zu wichtenden Indizien mit den Rangreihen diagnostisch präzisierter Ausprägungen von depressiven Störungen vergleichen. Der Vergleich der Messreihen kann mit elaborierteren statistischen Verfahren durchgeführt werden, das Ziel indessen bleibt gleich. Solange depressive Symptome nicht unmittelbar gemessen werden können, bleibt nur die Möglichkeit, die Messreihen aus unterschiedlichen mittelbaren Messverfahren miteinander auf den Grad der Übereinstimmung zu prüfen, ohne dass überhaupt Gewissheit darüber erlangt wird, welches Messverfahren valide arbeitet. Die auch in dieser Arbeit im weiteren Verlauf der Erörterungen aus unterschiedlichen Gründen eingenommene Position, die psychiatrische Diagnose sei der für alle epidemiologischen Studien zum Auftreten psychischer Störungen verbindliche externe Referenzrahmen ist wissenschaftstheoretisch insoweit relativierbar, als auch Diagnoseverfahren den messtheoretisch gebräuchlichen Kriterien wie Objektivität, Reliabilität und Validität unterliegen, sobald sie als Verfahrensweisen betrachtet werden, die Gewissheit über die Merkmale von organismischen Zuständen erbringen sollen. Die Öffnung solcher verfahrensspezifischen Perspektiven anlässlich der ersten und noch sehr allgemeinen Zielbestimmung für eine in ihrem Ansatz zu beschreibende epidemiologische Studie dient zunächst der Begründung für die Einführung zweier einander wechselseitig durchdringender Teilaspekte der zu regulierenden Tätigkeit. In der Handlungsregulationstheorie nach W. Hacker (1986) werden den inhaltlich als Vorwegnahmen des mit der Tätigkeit angestrebten Zustandes definierten Zielsetzungen zwei Subkategorien zugeordnet, die, als >Mittel< und als >Wege< bezeichnet, oberbegrifflich die Notwendigkeit repräsentieren, aus der allgemeinen Zielsetzung die Vorgehensweise zur Realisierung der Tätigkeit zu entwickeln. Die Zielsetzungen für die hier zu erörternde Studie geben eine erste, noch recht allgemeine inhaltliche Bestimmung dessen, was erreicht werden soll. Die bei Hacker als >Mittel< und >Wege< bezeichneten Planungsaspekte, fasst hier der Begriff von einer für diese Untersuchung durch Inanspruchnahme der allgemeinen Methodik empirischer Untersuchungen zu entwickelnde >spezifische Methodik< zusammen. Spezielle Ziele werden durch für sie spezifisch geeignete Verfahrensweisen zu erreichen versucht, welche das sind findet sich in einer damit korrespondierenden spezifischen Methodik erläutert. Für die wird aus der Handlungsregulationstheorie von W. Hacker neben anderem explizit auch der Aspekt der hierarchisch-sequentiellen Regulation auf die empirische Tätigkeit in all ihren Teilbereichen übertragen. Sowohl die Planung, als auch die Durchführung der Studie und die Ergebnisfeststellung werden zum einen durchgängig hierarchisch geordnet, zum anderen in dieser Ordnung während des Verlaufes der Studie und innerhalb ihrer Erörterung sequentiell geöffnet und realisiert.

Die nun folgenden Ausführungen wenden sich zunächst den allgemeinen Inhalten der anzusetzenden Studie und der für das Erreichen ihrer Ziele anzusetzenden spezifischen Methodik zu.

Die Einbeziehung von allgemeinen methodischen Aspekten in die Erörterung der dieser Studie gesetzten Arbeitsziele kann in sehr allgemeiner Weise als Folge der wechselseitigen Durchdringung von inhaltlich bestimmten Zielangaben und den zum Erreichen dieser Ziele einzusetzenden spezifischen Methodik begründet werden. Exemplarisch erkennbar wird die wechselseitige Durchdringung, wenn der allgemeinen Zielsetzung einer über die Daten aus einer Stichprobe zu erarbeitenden Populationsbeschreibung zum Auftreten depressiver Symptome der Hinweis auf die generelle Beschränkung der Möglichkeiten zum empirisch verlässlichen Erfassen solcher Symptome gegenüber gestellt wird. Wenn für die anzugehende Studie das Erstellen von medizinischen Diagnosen ausdrücklich ausgeschlossen wird, ergibt sich aus dieser methodischen Entscheidung, dass dann auch nur Indizien für das Auftreten von depressiven Symptomen erfassbar sein werden. So erklärt sich zum einen, warum der für diese Arbeit gesetzte Titel fast zwangsläufig zu wählen war. Für das Auftreten von depressiven Symptomen lassen sich nach bestehender messtheoretischer Sachlage nur Indizien sammeln. Daraus folgt zum anderen, dass die anzusetzende Studie den Widerspruch nicht aufzuheben vermag, durch den sie selbst notwendig geworden ist, den Widerspruch zwischen dem mutmaßlichen Auftreten depressiver Störungen unter griechischen Schulkindern und der damit verbundenen Gefährdung einerseits und den dazu bestehenden Defiziten an empirisch gesicherten Wissen andererseits. Die Untersuchung kann sich der Zielsetzung, empirisch gesichertes Wissen zu schaffen nur insoweit annähern, als die Aussagekraft der mit ihr gesammelten Indizien das zulassen. Das lässt aber auch die Möglichkeiten der Studie erkennen. Sie kann bestehende Wissensdefizite verringern helfen.

Den Gegenstand epidemiologischer Arbeiten stellt Shelley Taylor (1999, S. 8) wie folgt dar: „Changing patterns of illness have been charted and followed by the field of epidemiology, a discipline closely related to health psychology in its goals and interests (N.E. Miller, 1992). Epidemiology is the study of the frequency, distribution, and causes of infectious and non-infectious in a population, based on an investigation of the physical and social environment. For example epidemiologists not only studied who has what kind of cancer, but address questions such as why is some cancers prevalent in particular geographic areas than other cancers are. In the context of epidemiologic statistics we will see frequent use of two important terms: morbidity and mortality.”

S. Taylor nennt Krankheiten und die Begleitumstände ihres Auftretens als den Gegenstand epidemiologischer Arbeit. Indem daraus folgt, dass dann alle Untersuchungen, die zwar von Stichproben auf Populationen zu schließen versuchen, aber nicht das Auftreten von Erkrankungen erfassen, auch keine epidemiologische Untersuchungen wären, lässt sich die Notwendigkeit der Nutzung eines für diesen Bereich verwendbaren Begriffes erkennen. Bortz und Döring (2002) verwenden einen solchen Begriff, indem sie, die unterschiedlichen Gegenstände solcher Studien ausblendend, statistisch orientiert von Populationsbeschreibungen sprechen. Dieser Begriffsbestimmung wird in diesen Erörterungen in der Weise gefolgt, dass empirische Populationsbeschreibungen als Oberbegriff für alle Untersuchungsformen verwendet wird, die das tun, was der Begriff angibt. Epidemiologische Studien sind dann ein Spezialfall populationsbeschreibender empirischer Studien.

Die von Shelley Taylor zitierte Bestimmung epidemiologischer Studien ist, sobald Übereinstimmung über die Verwendung des Begriffes der Krankheit besteht, relativ problemlos anwendbar. Für die vielen verschiedenen Formen der Beeinträchtigungen durch psychische Störungen gilt das, zumindest in deutschsprachigen Ländern nicht oder nur bedingt. Anlass für diese Einschränkungen ist ein in der zweiten Hälfte des 20. Jahrhunderts verstärkt beobachtbarer Wandel in der oberbegrifflichen Einordnung von gesundheitlichen Belastungen durch psychische Beeinträchtigungen. In der deutschsprachigen Ausgabe des DSM-IV findet sich ein Nachhall der

grundlegenden Erörterungen zum Unterschied zwischen Krankheit und psychischen Störungen in der „Einführung zur deutschen Ausgabe: Zur Situation der operationalisierten Diagnostik in der deutschsprachigen Psychiatrie“ (Saß, Wittchen, Zaudig, 1996, S. IX - XXIV). Der Titel des Manuals verweist auf die oberbegriffliche Zuordnung all der im DSM-IV beschriebenen psychischen Beeinträchtigungen, indem es als „Manual Psychischer Störungen“ bezeichnet wird. Dann müsste folgerichtig im deutschsprachigen Raum für die Epidemiologie eine Veränderung in der Bestimmung ihres Arbeitsbereiches vorgenommen werden. Alternativ dazu könnten auch alle epidemiologisch mit psychischen Störungen befassten Untersuchungen als Populationsbeschreibungen zum Auftreten depressiver Störungen bezeichnet werden.

Für das erstgenannte Verständnis wäre neu zu bestimmen, dass epidemiologische Arbeiten auch auf psychische Störungen gerichtet sein können. Wenn C. Essau (2002, S. 45) gefolgt wird, ist dies längst geschehen, da sie, unter Berufung auf Lilienfeld & Lilienfeld (1980), Epidemiologie als „...Untersuchung der Häufigkeit einer Störung und ihrer Verteilung innerhalb unterschiedlicher Populationen...“ kennzeichnet. Als Hinweis auf eine entsprechende Praxis könnte der Titel einer Studie von Lewinsohn, P.M. & Seeley, J.R. (1998) gesehen werden. Sie berichten über „Major depressive disorder in older adolescents: Prevalence, risk factors, and clinical implications.“ Psychische Störungen werden auch in anderen Publikationen als Gegenstand der Epidemiologie betrachtet, so beispielsweise bei G. Laux (2002, S. 1 – 5). Die Darstellungen dort beziehen sich vor allem auf die Erfassung von Prävalenzraten für diagnostizierte depressive Störungen. Diese zweifellos verlässliche Erhebungsstrategie wird aber auch durch sich selbst in ihren Möglichkeiten begrenzt. Als auftretend können nur die Fälle psychischer Störungen erfasst werden, die über Kontakte zu diagnostisch arbeitenden Personen und Institutionen entsprechend diagnostiziert worden sind. Eine dem entgegenwirkende epidemiologische Strategie ist zunächst die als screening bekannt gewordene Vorgehensweise. Über möglichst ökonomisch einsetzbare Erhebungsmittel, wie etwas Fragebögen, wird versucht, relevante Informationen über Persönlichkeiten zu erfassen, die nach der jeweiligen Datenlage mit höherer Wahrscheinlichkeit als psychisch gestört gelten sollten. Solche Verfahren ersetzen nicht die Diagnose, können aber in vielfältiger Weise nutzbar gemacht werden. Eine für epidemiologische Untersuchungen bedeutungsvolle Entwicklung ist der Versuch der Konstruktion psychometrischer Verfahren, die Diagnosen in etwa gleichzusetzen wären. Das müssten Verfahren mit einer für klinische Ansprüche hinreichenden Sensibilität gegenüber psychischen Störungen und somit also auch Verfahren mit hoher Validität sein.

Bezogen auf die Erfassung depressiver Symptome bei Kindern und Jugendlichen ordnet Essau (2002, S. 29) solche Vorgehensweisen wie folgt ein: „Will man Depression bei Jugendlichen untersuchen, sind altersangemessene Erhebungsinstrumente mit guten psychometrischen Eigenschaften erforderlich, um Symptome, ihre Dauer, ihren Schweregrad und ihr erstmaliges Auftreten zu erfassen. Gute Erhebungsinstrumente sind ebenfalls wichtig für die Erfassung psychosozialer Probleme, die mit Depression einhergehen, sowie zur Evaluation des Behandlungserfolgs. Eine weitere Funktion der Erhebungsinstrumente ist das „Screening“, mit Hilfe dessen die Population einer Studie schnell, ökonomisch und valide in wahrscheinlich „gesunde“ und wahrscheinlich „kranke“ Gruppen unterteilt werden kann. Zur Erfassung von Depressionen bei Jugendlichen wurden im Allgemeinen strukturierte diagnostische Interviews, Selbstbeurteilungs-Fragebögen und Rating-Skalen eingesetzt.Die Einbeziehung der psychosozialen Beeinträchtigung ist von Bedeutung, da die Erfassung der Symptome oft einer gewissen Willkür unterliegt. Es fehlt noch immer ein „goldener Standard“, anhand dessen Daten aus verschiedenen Quellen validiert und die Unterschiede zwischen diesen Quellen gehandhabt werden können. ...Die Entwicklung differenzierterer Taxonomien psychischer Störungen und expliziter diagnostischer Kriterien erforderte und führte folglich auch zu einer eher

standardisierten Herangehensweise zur Erfassung psychiatrischer Symptome (Essau et al. 1997b; Tab. 2.1). In diagnostischen Interviews werden spezifische Symptom- und Prüffragen eingesetzt, detaillierte Codierungsregeln und diagnostische Algorithmen. In jeder diagnostischen Kategorie werden Fragen formuliert, um die Symptome der Störung, ihre Dauer und Ausschlusskriterien zu evaluieren. Alle diagnostischen Interviews beinhalten Listen von Zielverhalten, Symptomen und Richtlinien für die Durchführung von Interviews sowie die Aufzeichnung der Antworten und erlauben eine direkte Ableitung von Diagnosen (Essau et al. 1997b).“

Die in dem letzten der zitierten Sätze enthaltene Aussage ist für epidemiologische Untersuchungen bedeutsam und folglich auch Ansatzpunkt für kritische Nachfragen. Diagnostische Interviews, wie etwa das Verfahren CAPA nach Angold et al. (1995) sind relativ zeitaufwendig und begegnen bei ihrer Durchführung einigen verfahrensspezifischen Problemen. Für epidemiologische Arbeiten wird aufmerksam zu verfolgen und zu prüfen sein, ob die Identifizierungsleistungen der diagnostischen Interviews die von Selbstangaben aufnehmenden Fragebögen wie die von Angold et al. (1987), Kovacs (1992), Reynolds (1987), oder Rossmann (1993) zu übertreffen vermögen. Das scheint bisher noch nicht sicher zu sein. Für epidemiologische Arbeiten bleibt ungeachtet der künftigen Entwicklung dieser Qualitätsmerkmale zu entscheiden, ob die Ökonomie der Anwendung von Fragebögen durch andere Messmethoden erreicht oder kompensiert werden kann.

Die Sichtung der zur Beantwortung der Frage nach den Objekten epidemiologischer Studien einschlägigen deutschsprachigen Literatur stützt die Feststellung, dass psychische Störungen ein notwendiger Gegenstand epidemiologischer Arbeiten sind. Dann sind aber auch, so wird hier gefolgert, Nachforschungen zum Ausmaß depressiver Verstimmungen bei Kindern und Jugendlichen ein legitimer Gegenstand epidemiologischer Arbeit. Dass dies so verstanden werden kann, war hier festzustellen und zu begründen.

Welche Bedeutung einer solchen Einordnung zukommt, ist von dem Bezugssystem abhängig, in dem die Begriffe verwendet werden. Für die hier zu beschreibende Studie entstünden keine unmittelbaren Folgen, wenn sie als populationsbeschreibende Studie eingeordnet wird. Das ersparte die Auseinandersetzung mit dem Krankheitsbegriff und seiner verschiedenartigen Verwendung in unterschiedlichen Ländern. Eine Positionierung populationsbeschreibender Studien findet sich bei Bortz und Döring (2002), dort dann auf S. 113 und 114. Die Autoren gehen auf den hier zuvor aufgenommenen Sachverhalt unter dem Aspekt der „Evaluation durch Populationsbeschreibung“ ein.

„Wie auf S. 55 bereits dargelegt, gehören populationsbeschreibende Untersuchungen nicht zu den hypothesenprüfenden Untersuchungen im engeren Sinne; sie sind damit für summative Evaluationsstudien ebenfalls nur von nachgeordneter Bedeutung. *Prävalenz und Inzidenz*: Für die Vorbereitung eines größeren Interventionsprogramms sind Populationsdeskriptionen allerdings unverzichtbar. Hier hat die auf Stichproben (bzw. bei kleineren Zielpopulationen auch auf einer Vollerhebung) basierende Populationsbeschreibung die Funktion, die Notwendigkeit einer Maßnahme zu begründen. Von besonderer Bedeutung sind populationsbeschreibende Untersuchungen für die *epidemiologische Forschung*. Die allgemeine Epidemiologie hat zum Ziel, die Verteilung und Verbreitung von Krankheiten und deren Determinanten in der Bevölkerung zu untersuchen (Pflanz, 1973). Eine wichtige epidemiologische Maßzahl ist die Prävalenz (oder Prävalenzrate), die angibt, wie viele Personen (bzw. welcher Anteil) einer Zielpopulation zu einem bestimmten Zeitpunkt (Punktprävalenz) bzw. über eine bestimmte Zeitspanne (Periodenprävalenz) an einer bestimmten Krankheit leiden. Das Krankheitsgeschehen wird durch die *Inzidenz*(rate) charakterisiert, mit der die Anzahl (der Anteil) der Neuerkrankungen während eines bestimmten Zeitraumes erfasst wird. Von der Prävalenz und Inzidenz einer

Krankheit (z.B. Grippeepidemie) hängt es ab, welche Maßnahmen zum Schütze der Bevölkerung ergriffen werden müssen bzw. wie umfanglich diese Maßnahmen sein sollten (z.B. Herstellung, Verbreitung und Vorratshaltung von Impfstoffen). Die Wirksamkeit derartiger Maßnahmen zu überprüfen, wäre dann wiederum eine Evaluationsaufgabe. Im allgemeinen Verständnis beziffert die Prävalenz die Verbreitung eines bestimmten Notstandes, einer Störung oder eines Phänomens und die Inzidenz das Neuauftreten des fraglichen Sachverhaltes. Hierzu zwei Beispiele: In einer „Schlafstadt“ außerhalb der Stadtgrenzen sollen die Angebote zur Gestaltung jugendlicher Freizeit verbessert werden (Jugendclubs, Diskotheken, Sporteinrichtungen etc.). Man verspricht sich von dieser Maßnahme eine Reduktion der hohen Gewalt und Kriminalitätsraten. Hier müsste in einer stichprobenartigen Vorstudie zunächst geklärt werden, wie prävalent das Bedürfnis nach neuen Freizeitmöglichkeiten unter den Jugendlichen ist und welche Maßnahmen die Jugendlichen für geeignet halten, ihre Situation zu verbessern. Ohne diese Vorstudie bestünde die Gefahr, dass die im Zuge der Maßnahme geplanten Neueinrichtungen von den Jugendlichen nicht genutzt werden, weil man generell nicht an organisierter Freizeitgestaltung interessiert ist oder weil die neuen Möglichkeiten nicht den Wünschen der Jugendlichen entsprechen. Die Maßnahme wäre also eine Fehlinvestition und würde das eigentliche Ziel, die Jugendkriminalität zu reduzieren, nicht erreichen.“

Aus diesem recht umfanglichen Zitat kommt für die hier dann nachfolgend zu beschreibende eigene Untersuchung einem Satz besondere Bedeutung zu. Der wird nachfolgend erneut zitiert. „Im allgemeinen Verständnis beziffert die Prävalenz die Verbreitung eines bestimmten Notstandes, einer Störung oder eines Phänomens und die Inzidenz das Neuauftreten des fraglichen Sachverhaltes.“ Hinzuzusetzen wäre, dass mit der Bezifferung eines bestimmten Notstandes die Aufgabe, die sich daraus ergibt, in ihrem Umfang verdeutlicht wird. Eben das wird mit der dann nachfolgend zu beschreibenden eigenen Untersuchung für den bestimmten Notstand der depressiven Belastung von Kindern und Jugendlichen in Griechenland zu erreichen versucht. Daraus wird dann erneut der Widerspruch erkennbar, der zwischen Zielsetzung und Realisierungsmöglichkeiten entsteht. Wiederholt verwiesen wurde darauf, dass bei Ausschluss der medizinischen Diagnose auch keine Gewissheit über die Prävalenz depressiver Störungen entstehen kann. Ein Nebenaspekt dieses Widerspruchs ist, dass nicht nur die Prävalenz von depressiven Störungen von fachlichem und gesellschaftlichem Interesse ist, sondern darüber hinaus auch die Verteilung depressiver Symptome in Stichproben und Zielpopulationen. Weiterhin ist für die wissenschaftliche Positionierung von Präventionen nicht allein Wissen über das Auftreten depressiver Störungen bedeutungsvoll, viel zu selten wird hervorgehoben, wie wichtig das Wissen über das Nichtauftreten von Störungen sein kann. Medizinische Diagnosen vermitteln aber so gut wie keine Informationen über die Ausprägung der Freiheit von depressiven Symptomen. Insoweit ist dann die diesem Textabschnitt vorangestellte Überschrift durchaus programmatisch. Versucht wird, die Verdeutlichung spezieller gesellschaftlicher Notstände durch die über eine Stichprobe vermittelte Populationsbeschreibung zum Auftreten depressiver Symptome unter griechischen Kindern und Jugendlichen zu erreichen.

Kapitel 2

Der aktuelle Forschungsstand über Depressionen bei griechischen Kindern und Jugendlichen

Das durchzuführende Untersuchungsvorhaben soll die in einem speziellen Zusammenhang aufgeworfene Frage nach dem Auftreten depressiver Symptome unter griechischen Schulkindern möglichst verlässlich beantworten. Obschon zumindest nach dem erreichten Stand der Erörterungen noch nicht durchgeführt, geht das Vorhaben selbst bereits auf Voraussetzungen zurück. Die könnten, als selbstverständlich angenommen, ungenannt bleiben. Implizit in Anspruch genommener Voraussetzungen bleiben aber zunächst unerkannt. Explizit benannte Voraussetzungen erlauben deren Einbeziehung in die Kontrolle der Vorgehensweise. Die Gesamtheit der impliziten und expliziten Voraussetzungen einer empirischen Untersuchung werden hier als deren Ansatz bezeichnet. In dem Ansatz der dann nachfolgend zu beschreibenden Untersuchung sind mehrere explizit hervorgehobene Voraussetzungen aufgenommen worden. Sie betreffen primär das Verständnis von populationsbeschreibenden Untersuchungen, von depressiven Belastungen und deren sozialen oder gesellschaftlichen Implikationen. Alle diese explizit aufzuführenden Voraussetzungen sind Inhalte der Vorarbeiten zu dieser Untersuchung gewesen. Im folgenden werden sie als explizite Inhalte des erarbeiteten Untersuchungsansatzes aufgeführt:

1. Mit W. Schönplugg (2004) wird unterstellt, dass die von der Psychologie wissenschaftlich erfassten Gegenstände, so dann auch das Phänomen depressiver Belastungen, mit den soziokulturellen und gesellschaftlichen Verhältnissen der jeweiligen Gemeinschaften in wechselseitiger Abhängigkeit stehen.
2. Damit verbunden ist auch die Möglichkeit des zeitübergreifenden Verlustes von bereits gesellschaftlich verfügbar gewesenem psychologischen Wissen (Verlusthypothese).
3. In dieser historischen Perspektive sind depressive Belastungen ein die Menschen zu jedem Zeitpunkt in der Menschheitsgeschichte betreffendes Phänomen. Für die griechische Antike ist das historisch mit den Schriften vieler Philosophen belegt, so beispielsweise in „*Περί Ευθυμίας*“ („Über Euthymie“) von Plutarch.
4. Dessen Wahrnehmung, Bewertung und interpretative Einbindung in Theorien variierte und verändert sich weiterhin.
5. Der Sachverhalt einer in ihren jeweiligen Kennwerten unterschiedlichen Hemmung der Menschen ansonsten auszeichnenden situationsadäquaten Variabilität von emotionalen und aktionalen Zuständen trifft, wie hier postuliert wird, auf bislang vertretenen Auffassungen über Depressionen zu.
6. Die in den verschiedenen Ausgaben von DSM und ICD jeweils festgelegten Symptomlisten repräsentieren fachspezifische Übereinkünfte, die der Optimierung des intervenierenden Umganges mit Depressionen dienen sollen und in ihrer jeweiligen Nützlichkeit danach zu bewerten sind.
7. Für die Psychologie wird angenommen, dass sie fachlich befähigt ist, nutzbare Informationen über Depressionen und ihr Auftreten aufzunehmen, wobei sie, wegen ihrer vielfältigen außerklinischen Zuständigkeitsbereiche entsprechend weiträumiger informiert als die Medizin.
8. Depressive Zustände sind in ihrer Gesamtheit individuelle Belastungen, die in ihrem jeweiligen Ausprägungsgrad variieren und mit einzelnen Symptomen oder als Syndrome auftreten und dann den Kriterien für depressive Störungsbilder entweder nahe kommen oder ihnen entsprechen.

9. Subjektiv können Belastungen, wenn sie eindringlich genug empfunden werden, graduelle Steigerungen erreichen und in Zustände übergehen, die in ihren Auswirkungen pädagogischen, psychologischen oder medizinischen - als persönliche Notlagen bezeichnet werden können. Suizidale Handlungen sind ein extremer Ausdruck subjektiver Not.
10. Subjektive Notlagen wegen depressiver Belastungen haben soziale Folgewirkungen.
11. Subjektive Notlagen wegen depressiver Belastungen ziehen gesellschaftliche Belastungen und Notstände nach sich.
12. Die Psychologie hat hinreichende fachliche Kompetenzen, um auf die durch depressive Phänomene unterschiedlicher Ausprägung verursachten persönlichen, soziale und gesellschaftliche Belastungen und Notstände aufmerksam machen zu können. Sie kann weiterhin zu deren Prävention, Minderung und Überwindung beitragen.

Die Psychologie in Griechenland hätte sich dieser Forderung erst seit 1984, dem Jahr der erstmaligen Besetzung eines griechischen Lehrstuhls für Psychologie, annehmen können. Festzustellen ist, dass sich um das Jahr 2000 die griechische Psychologie als eigenständige Wissenschaftsdisziplin nach einem historischen Kulturbruch in einem mit Hoffnungen zu betrachtenden Neubeginn befindet.

Der nun zu vollziehende nächster Arbeitsschritt in der Annäherung an die eigene empirische Untersuchungsarbeit ist die Darstellung des in Griechenland aktuell bereits vorfindbaren Wissens über Depressionen bei griechischen Kindern und Jugendlichen. Das ist nicht die Frage nach dem Wissen über die Möglichkeit des Auftretens von depressiven Störungen bei Kindern und Jugendlichen und auch nicht die nach Wissen über Prävalenzraten ausserhalb von Griechenland. Die Befassung mit dem Auftreten emotionaler Störungen in Griechenland muss sich, ebenso wie etwa intendierte Präventionen oder Interventionen, auf in Griechenland geführte empirische Untersuchungen stützen können. Erst die verdeutlichen die griechische Gestalt des psychologischen Gegenstandes und dessen inter- und intrapersonale Struktur.

Aus der Darstellung und Analyse griechischer Forschungen über das Auftreten depressiver Symptome im Kindes- und Jugendalter, wie auch aus deren Vergleich mit den relevanten internationalen Forschungen, resultieren dann die weiteren Zielsetzungen und Aufgaben, denen wissenschaftliches und gesellschaftliches Interesse zufallen sollte. Die nachstehenden Ausführungen haben die Darstellung des speziellen Literaturstandes um 2004 zum Ziel. Sie sollen erfassen, bestimmen und erklären helfen, wie aus dem verfügbaren griechischen und internationalen Forschungsstand über depressive Belastungen bei Kindern und Jugendlichen sich aktuell aufzunehmende Zielsetzungen ergeben, denen dann auch die hier erörterte Untersuchung zu folgen versuchen wird.

Wenn eine möglichst umfassende Darstellung der griechischen psychologischen Fachliteratur über Depression im Kindes- und Jugendalter begonnen wird, könnte die Übernahme der kategorialen Ordnung hilfreich sein, die in Deutschland, unter Anlehnung an die internationale Literatur, zur inhaltlichen Strukturierung des Forschungsbereiches gebräuchlich ist. Die beispielsweise von Groen und Petermann (2002) benutzen Begriffe sind diese: Geschichte, Beschreibung und Klassifikation, Epidemiologie, Komorbidität, Verlauf, Erklärungsansätze, psychologische Diagnostik, Intervention und Prävention. Hannelore Reicher (1998) unterschied zwischen Störungsform und Symptomatik, Auftretenshäufigkeit und Verlauf, diagnostische Zugänge, theoretische Erklärungsmodelle und empirische Forschungsergebnisse, Interventionsansätze und Prävention. Essau (2002) benutzte abweichend davon Beschreibung und

Klassifikation, Erhebungsmethoden und Diagnostik, Epidemiologie, Komorbidität, psychosoziale Beeinträchtigungen und Inanspruchnahme von Gesundheitsdiensten der psychosozialen Versorgung, Verlauf, Entstehungsmodelle von Depression, familiäre Faktoren, kognitive Faktoren, Lebensereignisse und Bewältigungsstrategien, Prävention und Intervention.

Die von den verschiedenen Autoren zur Gliederung des Forschungsbereiches benutzten Kategorien sind einerseits unterschiedlich und haben andererseits deutliche Gemeinsamkeiten. Die Darstellung all dessen, was über psychologisch relevante Aspekte des Auftretens von Depressionen bei Kindern und Jugendlichen in entsprechenden griechischen Lehrbüchern geschrieben wird, dürfte in seiner Ordnung mit den zitierten Publikationen in Übereinstimmung stehen. Unterschieden werden könnte dann zwischen Publikationen von im Ausland tätigen Psychologen einerseits und Darstellungen griechischer Autoren. Für die hier zu führende Nachforschung ist jedoch weit weniger wichtig, welche nationale Zugehörigkeit die jeweils tätig gewordenen Forscher hatten, als vielmehr die Beantwortung der Frage, ob die jeweils zu betrachtenden Untersuchungen mit dem Auftreten depressiver Störungen bei griechischen Kindern und Jugendlichen befasst waren. Wissen über das Auftreten depressiver Störungen in anderen Ländern ist unerlässlich, aber noch nicht hinreichend. Psychologisches Wissen über das Auftreten depressiver Störungen bei griechischen Kindern und Jugendlichen ist die Voraussetzung für das Eingehen auf diese Störung in Griechenland.

Demzufolge wird nun nach solchen Publikationen in der griechischen psychologischen Literatur gesucht.

Über Depressionen bei griechischen Erwachsenen dürften wissenschaftliche Publikationen von Psychiatern und von Psychologen erwartet werden. Die Erwartungen sind bei einer entsprechenden Umschau zu verifizieren. Die Studien in diesem Bereich sind relativ zahlreich und wenden sich unterschiedlichen Aspekten der depressiven Störungen zu. Untersucht wurden beispielsweise depressive Störungen im Alter (Argyriadou et al., 2001; Gournellis et al., 2001, 1999; Kontaxakis & Christodoulou, 2000; Madianos et al., 1992; Parashos et al., 2002), die saisonale Depression (Beratis et al., 1996; Simos & Dimitriou, 1995), der Zusammenhang vom Stress am Arbeitsplatz und Depressionen (Iacovides et al., 2003; Iacovides, et al., 1999 a, b; Tselebis et al., 2001), die depressiven Symptome bei psychotischen Patienten (Dimitriou, 1996; Kontaxakis, Havaki-Kontaxakis, Margariti et al., 2000; Kontaxakis, Havaki, Stamouli et al., 2000; Lyketsos et al., 1988; Lykouras et al., 2002; Lykouras et al., 2000), die Komorbidität von Depression und Persönlichkeitsstörungen (Garyfallos et al., 1999, 1994), die Depression bei schwerkranken Menschen (Sarantidis et al., 1997), die psychoanalytische Therapie der Depression (Ierodiakonou-Benou, 1998; Vaslamatzis, 1995) und adlerianische Psychotherapie bei depressiven Patienten (Zachariadis, 1991). Weiterhin finden sich Untersuchungen zu somatischen Symptomen bei Depressionen (Ierodiakonou und Iacovides, 1987; Marmanidis et al., 1994).

Ein wesentlicher Teil der psychiatrischen Forschung wandte sich den pharmakologischen Therapien und der neurophysiologischen Forschung bei depressiven Patienten zu (Dikeos et al., 1999; Dimitriou et al., 1998; Fotiou et al., 2003; Fountoulakis et al., 2001; Markianos et al., 2002; Papageorgiou et al., 2001; Papakostas et al., 1995; Paparrigopoulos, 2002; Paparrigopoulos et al., 2001; Tzanakaki et al., 2000).

Alle bis hier genannten Arbeiten befassen sich nicht mit dem Auftreten von Depressionen bei griechischen Kindern.

Die Suche nach dieser speziellen Literatur führt über eine entsprechende Literaturrecherche in einer internationalen Datenbank (PsycINFO, Ende 2004, Suchbegriffe: Depression, Child, Greece) zu keinen Hinweisen. Diese Rückmeldung kann nicht wirklich dem realen Stand entsprechen, da einige Untersuchungen nur in griechischen psychologischen Zeitschriften veröffentlicht worden sind.

Die schließlich nach Ausweitung der Suche aufgefundenen Publikationen sind sowohl im Hinblick auf den jeweils untersuchten Forschungsgegenstand, die angewandte Methodik, die verwendeten Messinstrumente und die untersuchten Stichproben wie auch mit Blick auf ihren Forschungsgegenstand recht unterschiedlich.

Explizit mit dem Auftreten depressiver Symptome bei Kindern und Jugendlichen hat sich nur eine Arbeit, nämlich die von Bibou und Kioseoglou (2001a,b) befasst. Zwei weitere Studien (Economou und Angelopoulos, 1989; Triliva und Mitarbeiter, 1998) haben den Zusammenhang zwischen depressiven Symptomen und Prüfungsangst bei Jugendlichen untersucht. Alle weiteren in Griechenland speziell zur Kindheit und Jugend publizierten psychologischen Untersuchungen befassen sich nicht vornehmlich mit dem Auftreten depressiver Störungen, sondern unterrichten über Gefühls- und Verhaltensproblemen auf allen Altersstufen, wobei Informationen über depressive Belastungen bei Kindern und Jugendlichen nur randständig gegeben werden.

Diese Studien lassen sich zwei Gruppen zuordnen. Eine Gruppe umfasst Studien, die auf von Kindern berichteten Informationen basieren. Dazu gehören die Arbeiten von Christopoulou (1996) und von Fichter, Xepapadakos, Quadflieg, Georgopoulou und Fthenakis (2004). Die zweite Gruppe besteht aus Untersuchungen, die Informationen über die psychische Gesundheit von griechischen Kindern nach Fremdbeurteilungen gesammelt haben. Als Informationsquellen dienten Motti-Stefanidi, Tsiantis & Richardson (1993), Tsiantis, Motti-Stafanidi, Richardson, Schmeck & Poustka (1994) sowie Mc Donald, Tsiantis, Achenbach, Motti-Stefanidi & Richardson (1995) die Eltern der Kinder. Lehrkräfte wurden von Motti-Stefanidi, Besevegis und Giannitsas (1996) befragt. Eltern und Lehrkräfte dienten Papatheofilou et al. (1988), Roussos, Karantanos et al. (1999) und Bibou, Kioseoglou & Stogiannidou (2001) als Informanten.

Im weiteren werden die Hauptergebnisse aus den aufgeführten Studien dargestellt.

Der bislang einzige bekannt gewordene Versuch, speziell auf Depressionen bei Kindern und Jugendlichen einzugehen, ist, wie schon erwähnt, eine Arbeit von Bibou & Kioseoglou (2001a,b). In dieser Untersuchung wurden über den Einsatz des Fragebogens „Moods and Feelings Questionnaire“ (MFQ, Angold, Costello & Pickles, 1987), Informationen über das Auftreten von Depressionen in einer Stichprobe von 610 Kindern und Jugendlichen aus Thessaloniki im Alter von 9 bis 18 Jahren zu erfassen versucht. Die Autoren fanden Hinweise auf das Bestehen einer Prävalenzrate von etwa 11,5%. Die festgestellten depressiven Störungen treten mit größerer Häufigkeit unter den Mädchen auf und nehmen allgemein mit dem Alter zu. Die Gruppe der depressiv hochbelasteten Kinder setzt sich bei den älteren Kindern der Stichprobe aus 15,8% der untersuchten Mädchen und 3,6% der beteiligten Jungen zusammen. Die Antworten, in denen sich depressiv gefährdete von weniger belasteten Kindern unterscheiden, beziehen sich vor allem auf den Verlust von Freude, das Gefühl, nicht geliebt zu sein, die Hoffnungslosigkeit und auf Suizidgedanken. Die von den Kindern über die Beantwortung der jeweiligen Items bekundete persönliche Vereinsamung, die Hinweise auf bestehende Schlafprobleme und die Angaben über Sprachstörungen weisen statistisch signifikante Unterschiede zwischen den identifizierten Gruppen auf. Ganz im Gegensatz dazu und in Kenntnis der Literaturlage eher überraschend traten keine signifikanten Gruppenunterschiede bei der Beantwortung der Frage nach der persönlichen Stimmungslage auf. „Dieses Ergebnis könnte man durch die Amphithymie und die

vorübergehenden Stimmungsschwankungen, die charakteristisch bei Kindern dieses Alters sind, interpretieren.“ (Bibou & Kioseoglou, 2001a,b, Seite 86). Die Anmerkung verweist auf ein, den Autoren zufolge, bekanntes Phänomen in den Stimmungsangaben innerhalb der untersuchten Altersklasse, die für schnell umschlagende Stimmungsänderungen bekannt seien.

Die Schulleistungen der Kinder und Jugendlichen stehen in dieser Untersuchung nicht in Zusammenhang mit dem Auftreten depressiver Symptome.

In dieser Untersuchung wurden auch depressionsspezifische Informationen von den Lehrkräften über ihre Schülerinnen und Schüler erhoben. Verwendet wurde das Instrument „Strengths and Difficulties Questionnaire“ (SDQ, Goodman, 1994). Die Ergebnisse zeigen, dass die Lehrkräfte Schwierigkeiten haben, die depressiv belasteten Kinder zu erkennen. Die berichteten Schwierigkeiten sind bei den Jungen größer als bei den Mädchen.

Ob selbstberichtetes Erleben depressiver Symptome bei Jugendlichen im Zusammenhang mit Prüfungsangst auftritt, untersuchten Economou und Angelopoulos (1989) sowie Triliva (1998).

Ziel der Forschungsarbeit von Economou und Angelopoulos (1989) war, eine empirische fundierte Entscheidung über vermutete Zusammenhänge zwischen Symptomen von Angst und Depression einerseits, der Manifestation von Aggressivität oder Feindseligkeit (hostility) andererseits und den Schulleistungen zwei Wochen vor abschließenden Schulprüfungen herbeizuführen. Die Stichprobe bildeten 754 griechische Jugendliche im Alter von 15 bis 18 Jahren aus Tripolis. Die aus 14 Items bestehende Selbstbeurteilungsskala „Delusions Symptoms States Inventory / states of Anxiety and Depression“ (DSSI / sAD, Bedford et al., 1967) wurde für die Einschätzung von depressiven Belastungen und Angst verwendet. Der Fragebogen „Hostility and Direction of Hostility Questionnaire“ (HDHQ, Foulds, 1965; Caine et al., 1967) diente zur Bestimmung von introvertierter und extravertierter Aggressivität.

Aus der Analyse der Daten ergab sich eine negative Korrelation zwischen Angst und depressiven Symptomen einerseits und Schulleistung andererseits. Diese Korrelation ist allerdings relativ niedrig.

In dieser Untersuchung berichteten 26,9% der Mädchen über depressiven Symptome und unterschieden sich damit signifikant von den anderen der Stichprobe. Viele der über depressive Belastungen berichtenden Mädchen erreichten sehr gute Leistung in ihren schulischen Abschlussprüfungen. Allgemein erweisen sich die Schulleistungen der Mädchen in dieser Untersuchung gegenüber denen der Jungen als signifikant besser bewertet. In Bezug auf die erhobenen Daten zur Aggressivität ergab sich für die Mädchen der Stichprobe gegenüber den Jungen ein höheres Aufkommen an introvertierter Aggressivität, während die Jungen vermehrt extravertierte Aggressivität bekundeten. Beide Formen von Aggressivität sind mit der Schulleistung der Jugendlichen negativ korreliert. Die Korrelationen sind aber auch in diesem Fall sehr schwach ausgeprägt. In der Veröffentlichung dieser Studie werden keine Informationen über die psychometrische Eigenschaften der verwendeten griechischen Form der Messinstrumente gegeben.

Vermutete Zusammenhänge zwischen Copingstrategien (Bewältigungsstrategien) gegenüber Angst vor Prüfungen zum einen und Depression bei Jugendlichen zum anderen untersuchten Trilivas, Vassilaki und Chimienti (1998). Ihre Studie bietet auch einige Informationen über depressive Belastungen bei griechischen Jugendlichen. Die Autoren haben eine Stichprobe von insgesamt 800 Jugendlichen im Alter von 17 und 18 Jahren auf Kreta untersucht. Verwendet wurden das „Beck Depression Inventory“ (Beck, 1978), das „State and Trait Anxiety Inventory“ (Spielberger et al., 1970) und die griechische Skala „Strategies in Coping with Problems“ (Besevegis, 1996). Alle drei Instrumente waren schon in verschiedenen griechischen Stichproben standardisiert worden. Die Autoren berichten; dass, in Übereinstimmung mit den Ergebnissen aus

der internationalen Forschung, die Gesamtscores von Mädchen sowohl im Hinblick auf depressive Belastungen als auch in Bezug auf Angst signifikant höher lagen als die der Jungen. Bei Verwendung der von Beck vorgeschlagenen Cutoff-Scores zeigten sich 41,5% der Jugendlichen als nicht oder kaum von depressiven Symptomen belastet. Desweiteren waren 25,6% der Teilnehmer der Gruppe „mild depressiv“ zuzurechnen, 11,4% gehörten zu der Gruppe „mild-to-moderate depressiv“, während für 16,4% der Jugendlichen eine „moderate-to-severe depression“ anzunehmen war. Von 5,1% der Jugendlichen wurde berichtet, sie hätten Symptome einer „severe depression“ erlebt. Die aus dieser Untersuchung bekannt werdenden Prozentsätze für das Auftreten verschiedener depressiven Belastungen und auch die für das Auftreten von Angst liegen höher als die aus den Studien für die Standardisierung der beiden Instrumenten (Ntonias und Demertzis, 1983a; Liakos und Giannitsi, 1984). Die Autoren diskutieren die höheren Prävalenzraten und die gefundenen geschlechtsspezifische Unterschiede im Hinblick auf kulturspezifische Aspekte bezüglich der Kinderernährung und die hohe Bedeutung des Schulerfolges für griechische Kinder und ihre Familien. Im Hinblick auf die gefundenen Beziehungen zwischen Copingstrategien und weiteren untersuchten Variablen fassen die Autoren wie folgt zusammen: „Intercorrelations among anxiety, depression and coping scores yielded similar patterns for male and female students, with depression and anxiety positively related to maladaptive coping effort, and negatively related to adaptive coping effort. However, coping effort scores (both maladaptive and adaptive) were, in general, more closely related to anxiety and depression among females than among males. Regression analysis indicated that, in general, only maladaptive (and not adaptive) coping strategies predicted anxiety and depression, among both males and females. In the present research, the evidence regarding coping style is not clear. Among both male and female students, (a) the same (maladaptive) coping strategies were found to be associated with anxiety and depression, and (b) depression and anxiety were positively related to maladaptive coping effort, and negatively related to adaptive coping effort. However, coping effort scores (both maladaptive and adaptive) were, in general, more strongly related to anxiety and depression among females than among males.“ (Trilivas et al., 1998)

Ziel der Arbeit von A. Christopoulou (1996) war die Untersuchung der psychischen Gesundheit von 334 Jugendlichen im Alter von 15 und 16 Jahren aus Athen unter Verwendung des „Youth Self-Report Questionnaire“ von Achenbach (1991b).

Die Mädchen dieser Stichprobe erwähnen signifikant mehr psychologische Probleme als die Jungen. Keine signifikante Geschlechtsunterschiede fanden sich in den Subskalen, die das Übertreten von Geboten und die Aggressivität der Kinder messen sollen. Im Hinblick auf die im DSM-IV oder im ICD-10 für emotionale Störungen aufgeführten Symptomlisten könnten einige Ergebnisse dieser hier zitierten Untersuchung als im Widerspruch dazu stehend aufgefasst werden. Die Mädchen berichten gegenüber den Jungen vermehrt über psychische Schwierigkeiten, dabei vor allem über emotionale Probleme. Dennoch beteiligen sie sich, ihren Angaben zufolge, in einem höheren Masse als die Jungen an schulischen, familiären und sozialen Aktivitäten. Die Zahl der von den Kindern berichteten psychologischen Probleme ist für die Kinder der ersten Lyzeumsklasse niedriger als für die Kinder der zweiten Klasse. Außer dem Geschlecht und der Schulstufe ist Erwerbstätigkeit neben dem Schulbesuch noch ein im Zusammenhang mit dem Auftreten psychischer Problemen stehender Faktor. Die Jugendlichen, die in irgendeiner Form erwerbstätig sind, berichteten häufiger über externalisierte und internalisierte Probleme als die Gleichaltrigen. Im Gegensatz zu Befunden aus früheren Untersuchungen (Tsiantis et al, 1994; Motti-Stefanidi et al., 1993, s.u.) steht die sozioökonomische Herkunft des Kindes und die Bildung seiner Eltern in keiner statistisch signifikanten Beziehung mit dessen psychischer Befindlichkeit. Zwar stehen Bildungsniveau und Beruf der Eltern in Zusammenhang mit der Teilnahme der Jugendlichen an den erfassten

unterschiedlichen Aktivitäten, zeigen aber keine Zusammenhänge mit dem Auftreten emotionaler Probleme. Die Autorin betont, dass die erfassten psychopathologischen Phänomene mit einer Reihe soziodemographischer Variablen zusammenhängen, dass diese Zusammenhänge bisher aber, wegen ihrer Komplexität, nicht deutlich genug nachgewiesen werden konnten. Sie schlägt vor, die Zusammenhänge zwischen den von ihr benannten Variablen und dem Auftreten psychopathologischer Symptome speziell zu untersuchen.

Zu den Forschungsarbeiten, die ihre Informationen direkt bei den Kindern selbst gesammelt haben, gehört auch die Arbeit von Fichter, Xepapadakos, Quadflieg, Georgopoulou und Fthenakis (2004). Hauptziel dieser Studie war ein Vergleich der psychischen Gesundheit von griechischen Jugendlichen in Veria (Griechenland) und von griechischen Jugendlichen in München unter Rückgriff auf eine identische Untersuchung, die vor 18 Jahren in beiden Städten unter Verwendung der gleichen Instrumente durchgeführt worden ist. Für die Einschätzung der jeweils vorliegenden psychopathologischen Belastungen wurde eine nach einer Faktorenanalyse auf 28-Items festgelegte Version des General Health Questionnaires (GHQ-28, Goldberg & Hillier, 1979) verwendet. Befragt wurden im Jahre 1980 insgesamt 815 Jugendliche im Alter zwischen 10 und 21 Jahren aus München und 2920 Jugendliche im gleichen Alter aus Veria. Im Jahre 1998 haben an der Folgeuntersuchung 881 Jugendliche aus München und 2920 aus Veria teilgenommen.

Aus der Untersuchung im Jahre 1980 hatte sich ergeben, dass die für die griechischen Jugendlichen erhobenen Daten eine stärkere psychopathologische Belastung darstellende Erhöhung sowohl in den Scores der Gesamtskala als auch in den Subskalen „somatic complaints“, „anxiety“, „social dysfunctioning“ und „depression“ des GHQ-28 anzeigten als die Werte der gleichaltrigen Münchner.

Die im Jahr 1998 erhobenen Daten zeigten im Vergleich dazu einen starken Anstieg der psychopathologischen Symptome in beiden Orten an. Der ehemals signifikante Unterschied in der Ausprägung psychopathologischer Belastungen zwischen Jugendlichen in Veria und denen in München bestand nicht mehr. Eine Ausnahme davon ergab sich in der Subskala „Depression“, wo die griechischen Jugendlichen aus Veria wiederum signifikant höhere Gesamtpunkte erreichten.

Zu beiden Messzeitpunkten, 1980 und 1998, zeigten sich geschlechtsspezifische Unterschiede. Die Mädchen wiesen mehr Symptome als die Jungen auf. Von den Mädchen der Gesamtstichprobe wird dann auch ein Großteil der Steigerung im Auftreten depressiver Symptome im Zeitraum von 1980 bis 1998 getragen.

Die Autoren der Studie stellen die in beiden räumlich weit auseinander liegenden Orten beobachtete Zunahme psychopathologischer Belastungen bei griechischen und deutschen Jugendlichen als das Hauptergebnis ihrer Untersuchung dar. „The increase in psychopathology in both locations is alarming and deserves further exploration“ (Fichter et al., 2004, S. 27).

Zu den eindrücklichen Ergebnissen dieser Studie ließe sich anmerken, dass nach Veränderungen in der Lebenssituation von griechischen und deutschen Jugendlichen zwischen den Messzeitpunkten 1980 und 1998 gesucht werden könnte. Im Jahr 1980 befand sich Griechenland in der Umbruchphase zwischen dem Ende der Diktatur und der Realisierung einer demokratischen Staatsordnung. Die Zukunftsperspektiven für in der griechischen Provinz lebende Jugendliche waren sehr unterschiedlich und von Ungewissheit geprägt. Für München darf man 1980 als ein Jahr politischer und wirtschaftlicher Stabilität bezeichnen. Zum zweiten Messzeitpunkt, 1998, sind sowohl die griechischen als auch die deutschen Jugendlichen gemeinsam Bürger in der EU und beide Gruppen jeweils Bürger in einem Staat mit wachsenden ökonomischen Schwierigkeiten. Insbesondere die auch für Jugendliche krisenhafte Situation am Arbeitsmarkt belastete sowohl die Lebensperspektiven der griechischen als auch die der deutschen Jugendlichen.

Nach Zusammenhängen zwischen psychologischen Problemen und Schulleistungen suchten Papatheofilou und Mitarbeiter (1988). Unter Verwendung des Rutter Questionnaire befragten sie die Eltern und Lehrer von 603 Kindern zwischen 6 und 8 Jahren. Die Lehrer und Eltern berichteten über psychische Auffälligkeiten der Kinder. Danach wurden die Jungen gegenüber den Mädchen als aggressiver und hyperaktiver wahrgenommen. Den Mädchen hingegen wurden mehr Ängste und Phobien zugeschrieben.

Insgesamt 44% der Eltern dieser Untersuchung berichteten, daß ihre Kinder unter ernststen psychischen Problemen zu leiden hatten. Die zu einem Vergleich heranziehbareren entsprechenden Prozentsätze aus den Ergebnissen einer auf der Isle of Wight und in Inner London durchgeführten Studie (Rutter et al., 1976) erbrachten Werte von 6% für die Isle of Wight und 12% für Inner London. Die Studie von Rutter et al. benutzte die gleichen Instruments und die gleichen cut-off Werte wie sie in der griechischen Studie Verwendung fanden.

Anders als die von den griechischen Eltern gegebenen Einschätzungen stehen die Urteile der griechischen Lehrer über griechische Kinder in relativer Übereinstimmung zu denen englischer Lehrer über deren Schülerinnen und Schüler. Nach Meinung der befragten griechischen Lehrer sollten nur 20% der griechischen Kinder der psychisch hoch gefährdeten Gruppe zugerechnet werden. Ähnliche Prozentsätze ergeben sich aus der oben erwähnten englischen Untersuchung von Rutter und Mitarbeitern.

Bibou, Kiosoglou und Stogiannidou (2001) haben die Lehrer und Eltern von 238 Grundschulkindern aus Thessaloniki mit dem Einsatz des Fragebogens „Strengths and Difficulties Questionnaire“ (SDQ, Goodman, 1994) nach Fähigkeiten und Schwierigkeiten ihrer Kinder befragt. Untersucht wurde von Bibou und Mitarbeitern inwieweit Lehrer- und Eltern in ihren Urteilen über diese Kinder übereinstimmten. Erwartet werden sollte das Aufkommen von Ergebnissen, die denen aus international bekannt gewordenen Studien zu der untersuchten Thematik gleichen. Vermutet wurde insbesondere das Auftreten signifikanter Unterschiede zwischen unterschiedlichen Gruppen von Fremdbeurteilern, wie sie in dieser Untersuchung durch die Gruppen der Eltern und der Lehrer repräsentiert waren. Die griechischen Eltern berichten dann auch signifikant häufiger als die Lehrer über das Auftreten psychischer Schwierigkeiten bei den Kindern der Stichprobe. Vor allem wird auf emotionale Störungen verwiesen. Die Lehrer wiederum sehen die Hauptprobleme der Kinder in deren Hyperaktivität und einer eingeschränkten pädagogischen Leitbarkeit. Die größte Übereinstimmung zwischen Eltern und Lehrern ergab sich für die Subskala „Hyperaktivität / Aufmerksamkeitsprobleme“. Die geringste Übereinstimmung zwischen Eltern und Lehrern bestand für die Skala „Emotionale Probleme“. Die Autoren verweisen für die Lehrkräfte auf eine gegenüber den Urteilen der Eltern höhere Reliabilität. Unabhängig davon, ob die Lehrerurteile valide sind, stellen sie sich als stabiler und stringenter dar. Das Elternurteil steht in dieser Untersuchung auch in Zusammenhang mit deren psychischer Gesundheit. Auf der anderen Seite finden sich Zusammenhänge zwischen den Einschätzungen der Lehrer und den Schulleistungen der Kinder. Nach dem Urteil von 92,8% der in dieser Arbeit befragten Lehrkräfte treten emotionale Störungen unter den Kindern ihrer Klasse nicht auf.

Dieses Ergebnis steht in Übereinstimmung mit einem Befund aus einer Untersuchung von Poulou und Norwich (2000), die griechische Lehrkräfte zu deren Einschätzungen über die Häufigkeit des Auftretens und die Stärke der Ausprägung von Verhaltensproblemen und emotionalen Auffälligkeiten bei den Kindern ihrer Klassen befragten. Aus dieser Arbeit ergab sich, dass die Lehrkräfte einerseits emotionale Probleme als eine gravierende Schwierigkeit für Schulkinder einschätzen, andererseits aber das Auftreten solcher Störungen als sehr selten annehmen (Poulou & Norwich, 2000).

Eine weitere, die Angaben von Eltern über ihre Kinder nutzende Studie ist die von Motti-Stefanidi, Tsiantis und Richardson (1993). Sie haben die Eltern von 466 in Athen lebenden Kindern im Alter von 6-11 Jahren mit der Absicht befragt, auf diese Weise einen Überblick über emotionale Probleme, Verhaltensauffälligkeiten und soziale Fähigkeiten der Kinder dieser Stichprobe gewinnen zu können. Die Datenerhebung wurde unter Verwendung des Child Behaviour Checklist (CBCL; Achenbach & Edelbrock, 1983) durchgeführt und dies dann auch mit der Absicht, diese Skala als Screeningsinstrument für Forschungen im griechischen Raum zu standardisieren. Die Autoren heben als wichtiges Ergebnis ihrer Arbeit den hohen Anteil von Eltern hervor, die ihre Kinder als psychisch belastet ansehen. Fast 45% der Eltern in der untersuchten Stichprobe schreiben den eigenen Kindern schwierige Verhaltensprobleme zu. Dieser Prozentsatz ist höher als in anderen Ländern (siehe dazu weiter unten). Mehr als 50% der Eltern in dieser Stichprobe charakterisieren ihre Kinder über Auffälligkeiten wie streitsüchtig, große Aufmerksamkeit einfordernd, eifersüchtig, zwanghaft, nervös und schüchtern. Mehr als 50% der Jungen wurden als aggressiv und hyperaktiv, mehr als 50% der Mädchen als phobisch vor Tieren, Situationen oder Orten bezeichnet. Ältere Kinder werden von den Eltern relativ häufig als übergewichtig eingestuft.

Die Angaben der Eltern variieren signifikant mit dem sozio-ökonomischen Status der Familien. Eltern aus unteren sozioökonomischen Schichten erwähnen zahlreichere Verhaltensprobleme und darunter insbesondere Probleme im sozialen Nahraum sowie Schulprobleme im Zusammenhang mit niedrigen Schulleistungen. Demgegenüber finden sich weitaus weniger Einlassungen auf spezielle soziale Kompetenzen, von denen im Vergleich dazu Eltern aus Familien in mittleren und oberen sozialen Schichten vermehrt berichten.

Eine der häufiger in den oben beschriebenen Arbeiten verfolgte Zielsetzung war der Vergleich der griechischen Ergebnissen mit entsprechenden Informationen aus den USA (McDonald, Tsiantis, Achenbach, Motti-Stefanidi & Richardson, 1995) und aus Deutschland (Tsiantis, Motti-Stefanidi, Richardson, Schmeck & Poustka, 1994).

Über den Vergleich der Ergebnisse aus der griechischen Untersuchung von Motti-Stefanidi, Tsiantis und Richardson (1993) mit denen aus einer Stichprobe von 790 Kindern aus Deutschland ergaben sich beachtenswerte Unterschiede. Die Werte der griechischen Kinder auf der Gesamtskala, aber auch die auf den Skalen für externalisierte und internalisierte Probleme und die auf den Kompetenzskalen im CBCL lagen, im Sinne einer stärkeren Problembelastung, höher als die aus Deutschland berichteten Angaben. Während unter den griechischen Eltern der untersuchten Stichprobe bis zu 45% den Kindern psychische Probleme zuschreiben, liegen die entsprechenden Prozentsätze aus der deutschen Untersuchung deutlich tiefer. Die Durchschnittswerte betragen 20.7% für die Jungen und 16.5% für die Mädchen. Eine Ausnahme davon ergibt sich über den Vergleich der von deutschen und griechischen Eltern bewerteten Schulleistungen ihrer eigenen Kinder. Die griechischen Eltern schreiben ihren Kindern in höherem Maße schulische Leistungen zu, als dies die deutschen Eltern für die deutschen Kinder tun. (Tsiantis et al., 1994).

Nach statistischen Analysen auf Itemebene hat sich ergeben, dass in 74 von 120 Items signifikante Unterschiede zwischen den zwei Ländern nachweisbar sind. Die deutschen Eltern geben nur bei 4 Items (Aufmerksamkeitsprobleme, Bauchschmerzen, Streiten, Enuresis) höhere Auftretenshäufigkeiten als die griechischen Eltern an. In allen anderen Fragen verhelfen die Eltern der griechischen Kinder denen zu signifikant höheren Punktwerten. Beispielsweise weinen, den Eltern zufolge, etwa 40% der griechischen Jungen und etwa 44% der griechischen Mädchen häufig. Für die deutschen Kinder verweisen deren Eltern auf 15% vermehrt weinende Jungen und auf 16% Mädchen. Nach den gleichen Quellen sind 36% der griechischen Jungen und 39% der griechischen Mädchen häufig besorgt, während für deutsche Jungen 17% und für Mädchen 14%

angegeben werden. Etwa 42% der Jungen und 53% der Mädchen aus Griechenland leiden, wie deren Eltern meinen, unter Phobien, im Gegensatz zu 19% und 24% unter den Knaben und Mädchen der deutschen Stichprobe.

Ähnliche Unterschiede zeigen sich nach einem Vergleich zwischen parallelisierten demographisch Stichproben von 356 Kindern im Alter von 6 bis 11 Jahren aus Griechenland und aus Amerika (McDonald et al., 1995). Die griechischen Eltern vergeben für ihre Kinder signifikant höhere Werte auf den Subskalen „Angst/Depression“, „Aufmerksamkeitsstörungen“, „Internalisation“, Externalisation“, „delinquentes Verhalten“ und „aggressives Verhalten“ als die amerikanischen Eltern an. Im Hinblick auf die Kompetenzskalen stufen die Eltern aus den USA ihre Kinder im Hinblick auf soziale Kompetenz und Aktivitäten in Bezug auf viele Tätigkeiten höher ein als die griechischen Eltern ihre Kinder. Das Gegenteil ergibt sich wiederum für die Skala, die sich auf die eingeschätzte Schulleistungen der Kinder bezieht. Die griechischen Eltern erkennen ihren Kindern höheren Schulerfolg zu als die amerikanischen.

Die von früheren Untersuchungen (Motti-Stefanidi et al., 1993; Mc Donald et al., 1995; Tsiantis et al., 1994; Papatheofilou et al., 1988) gefundenen höheren Prävalenzraten für psychische Probleme bei griechischen Kindern waren Ansatzpunkt für eine weitere Arbeit von Motti-Stefanidi, Besevegis und Giannitsas (1996). Die Autoren haben mit ihrer Arbeit herauszufinden versucht, auf welche Weise die Lehrkräfte in Griechenland das „kompetente, psychologisch gesunde Kind“ im schulischen Raum definieren. Das jeweils persönlich vertretene Glaubenssystem über das, was ein schulisch kompetentes griechisches Kind ausmacht, also die jeweiligen persönlichen Theorien darüber, wurden bei 58 Lehrern und Lehrerinnen in griechischen Grundschulen in Athen zu erfassen versucht. Die beteiligten Lehrer und Lehrerinnen hatten zunächst sowohl die freie Beschreibung eines kompetenten Junges als auch die eines ebensolchen Mädchens und beides jeweils aus ihren eigenen Klassen zu geben. Danach waren die Fragebögen „California Child Q-set“ (CCQ; Block, 1978) und Achenbachs „Teacher’s Report Form“ (TRF; Achenbach, 1991a) für eben die zuvor beschriebenen Kinder auszufüllen.

Die Befunde aus dieser Untersuchung sprechen dafür, dass die persönlichen Theorien der befragten Lehrkräfte über kompetente Schulkinder in hohem Maße homogen und sowohl vom Geschlecht der Lehrkräfte als auch von dem der Kinder unabhängig sind. Dieses Ergebnis steht in Übereinstimmung mit relevanten Forschungen in westlichen Ländern.

Die kompetente Kinder wurden, wie es sich aus den freien Beschreibungen ihrer Lehrer ergibt, als sehr sozial beschrieben. Sie verfügten auch über gute Beziehungen zu den Erwachsenen ihrer Umgebung. Diese Kinder sind nach der Meinung der Lehrkräfte leistungsstark, fleißig, engagiert in schulischen Aktivitäten und konzentriert.

Ähnliche Einstellungen ergeben sich nach Analysen der Informationen aus dem CCQ. Aus dem Profil der kompetenten Kinder ergibt sich deren hohe Leistungsorientierung, ihre guten sozialen Beziehungen sowohl zu Gleichaltrigen als auch zu Erwachsenen, ihr Wissensdurst und ihr Interesse an der Welt, ihr Sinn für Humor und die Bereitschaft zur Kooperation. Die weniger kompetenten Kinder sind nach der Meinung ihrer Lehrer durch aggressives Verhalten gekennzeichnet, durch geringes Selbstwertgefühl und weiterhin durch soziale Probleme in vielen interpersonalen Beziehungen.

Aus dem Vergleich der Informationen aus dem CCQ mit denen aus dem TRF hat sich ergeben, dass die Gruppe der kompetenten Kinder signifikant höhere Werte in der Bewertung der Schulleistungen und niedrigere Werte in den Subskalen der Dimensionen Aufmerksamkeitsstörung“ und „Interpersonale Probleme“ als die Gruppe der weniger kompetenten Kinder aufweist. Die zwei Gruppen unterscheiden sich voneinander leicht auf der Subskala des Angstsyndroms und deutlich auf der Subskala „Depression“.

Eine Studie, die sich ebenfalls mit der Standardisierung der "Child Behaviour Checklist" (Achenbach, 1983) und des TRF (Achenbach, 1991a) befasste, ist die von Roussos und Mitarbeiter (1999). Obwohl der CBCL schon in zwei Studien mit griechischen Kindern verwendet worden war (Motti-Stefanidi et al., 1993; Tsiantis et al., 1994) hatten Roussos und Mitarbeiter eine weitere Untersuchung zur Standardisierung für notwendig erachtet. Die Autoren bringen zwei Argumente vor, die ihre Wiederholung begründen. Einerseits haben Motti-Stefanidi und Tsiantis in ihren Studien eine frühere Version des CBCL verwendet. Die basiert gegenüber der nachfolgenden auf abweichenden Definitionen für die zu messenden Konzepte. Weiterhin sind die Stichproben der vorangegangenen Untersuchungen nicht hinreichend repräsentativ für die griechische Bevölkerung, da sie nur mit Kindern aus Athen arbeiteten. „This excluded the possibility of studying the effects of urbanization and, therefore, intracultural differences. This is of particular interest in Greece since in certain respects there are vast differences in the way of life between cities and rural areas" (Roussos et al., 1999, S. 170).

Die Autoren haben mit ihrer Studie die Eltern und Lehrer von 1200 griechischen Kindern im Alter zwischen 6 und 12 Jahren aus Groß-, Mittel- und Kleinstädten durch den Einsatz des CBCL und des TRF angesprochen. Damit sammelten sie die gewünschten Informationen über die psychische Gesundheit der mit der Stichprobe repräsentierten Kinder.

Auch diese Studie bestätigt die Findungen der vorausgegangenen insoweit, als auch die Eltern und die Lehrkräfte der neuen Stichprobe den Jungen mehr Verhaltensprobleme zuschreiben als den Mädchen. Die Unterschiede betreffen vor allem sogenannte externalisierte Probleme. Die Lehrer beschreiben darüber hinaus die Jungen im Vergleich zu den Mädchen als weniger anpassungsfähig gegenüber schulischen Anforderungen. Die von den Eltern gegebenen Antworten stehen in keiner Beziehung zu den einzelnen Wohnorten, aber hängen signifikant mit dem sozioökonomischen Status der Familien zusammen. Wie schon in der Untersuchung von Motti-Stefanidi (1993) geben Eltern aus Familien mit niedrigerem sozioökonomischen Status ihren Kindern tendenziell höhere Scores auf der Problemskala und niedrigere Scores in der Skala sozialer Kompetenzen. Nach der Meinung der befragten Lehrkräfte haben die Kinder aus Kleinstädten weniger Verhaltensprobleme und auch größere soziale Fertigkeiten als Kinder aus den Mittel- und Großstädten. In Bezug auf die sozioökonomische Herkunft des Kindes stimmen die Informationen der Lehrkräfte mit denen aus den Elternhäusern überein.

Aus der auf Itemebene geführten Analyse der gesammelten Informationen ergab sich allerdings, dass Eltern und Lehrkräfte das Gewicht ihrer Aufmerksamkeit auf unterschiedliche Verhaltensweisen richten. Die Eltern weisen vermehrt auf die emotionalen und sozialen Aspekte des Verhaltens ihrer Kinder hin. Die Lehrer hingegen beachten bei Jungen vornehmlich deren schulische Zielstrebigkeit und bei den Mädchen deren Lerneifer. Die Urteile von Eltern und Lehrern korrelieren allgemein nur sehr niedrig miteinander. Zwei Ausnahmen davon stellen zum einen die Subskala der Aggressivität für die Jungen und zum anderen die der Aufmerksamkeitsprobleme für beide Geschlechter dar.

Über den Vergleich der Ergebnisse aus der dargestellten Studie mit denen aus einer Arbeit von McDonald et al., (1995) wird die Ähnlichkeit der Befunde erkennbar. Die Eltern aus Griechenland beschreiben ihre Kinder als ängstlich, depressiv und aggressiv. Infolgedessen haben die griechische Kinder höhere Werte als die gleichaltrigen amerikanischen und dies sowohl auf den Subskalen „Internalisierte Probleme“ und „Externalisierte Probleme“ als auch in der Gesamtskala der Auffälligkeiten. Die griechischen Kinder werden als problembelasteter eingeordnet. Die befragten griechische Lehrer und Eltern stimmen in der Wahrnehmung depressiver Symptomen und Aufmerksamkeitsprobleme überein. Die Auffassungen differieren im Hinblick auf Aggressivität. Die Autoren heben als besonders bemerkenswert hervor, dass die griechische

Lehrer emotionalen Problemen bei ihren Kinder mehr Aufmerksamkeit schenken als die Lehrkräfte aus der amerikanischen Stichprobe. Sie kommentieren diesen Befund mit einer Frage: „These results are intriguing. Offspring of a more traditional culture than that of the USA appear as being more aggressive and delinquent than American children. They are also pictured as having more emotional problems. Is this a reflection of real life stresses, sensitivity on the part of parents and teachers to their children’s problems, or is it purely an artefact generated by cultural factors affecting responses to the instrument?“ (Roussos et al., 1999, S. 170)

Trotz aller Unterschiede zwischen den hier beschriebenen Untersuchungen, dies sowohl im Hinblick auf den jeweiligen Forschungsgegenstand als auch in Bezug auf die Methodik, sind mehrere Ergebnisse als Belege oder Hinweise auf das Auftreten depressiver Belastungen bei griechischen Kindern und Jugendlichen zu verstehen. Die mit ihren Arbeiten zitierten Autoren nehmen das Phänomen des Auftretens emotionaler Störungen aus sehr unterschiedlichen Perspektiven wahr. Nur ein Teil der Arbeiten sucht an den Kindern und Jugendlichen nach Informationen über emotionale Störungen. Ein anderer Teil befasst sich mit Berichten, die von deren Interaktionspartnern gegeben worden sind. Dabei werden auch im Verhältnis zur Auftretensproblematik randständig erscheinende Probleme betrachtet. McDonald diskutiert beispielsweise die in Vergleich zu amerikanischen Stichproben erhöhten Angaben zum Auftreten aus Fremdeinschätzungen im Zusammenhang mit kulturell unterschiedlich determinierter Berichterstattung seitens griechischer oder amerikanischer Eltern. Auch das kulturell unterschiedliche Verständnis von Nachfragen durch Organisationen wird als potentiell determinierend betrachtet. Den Kenntnisstand über das Auftreten depressiver Störungen berühren diese Untersuchungen nur mittelbar.

Dennoch tragen alle diese Untersuchungen in der ihnen jeweils eigenen Art zur Ausdifferenzierung der Perspektiven auf Depressionen bei griechischen Kindern und Jugendlichen bei. Notwendig ist die Unterscheidung zwischen diagnostisch gesicherten Informationen über das Auftreten emotionaler Störungen einerseits und den auf andere Informationsquellen zurück gehenden Aufschlüssen. Ein Großteil der zitierten Untersuchungen berührt weniger das Auftreten der Störungen als persönliche Theorien von Lehrern und Eltern darüber. Das sind relevante Forschungsfragen, aber sie geben nur mittelbar Möglichkeiten, Wissen über das Auftreten von Depressionen zu vermehren. Immerhin werden so aber auch wichtige methodische Aspekte der depressionsspezifischen Forschung zur Diskussion gebracht, dies allerdings fast ausschließlich unter dem Aspekt interkultureller Vergleiche. „Formal data regarding child mental health referral rates in Greece are not available. However, a WHO report has indicated that Greek parents are reluctant to acknowledge their children’s emotional or mental health services for these problems (Tsiantis et al, 1982). Instead, they are more likely to seek help from friends, relatives, or paediatricians. If mental health and special education services are, indeed, less often sought by Greek than American parents, then the Greek sample may have included more children who would have been excluded from the American sample, which could have contributed to the higher scores found for Greek children“ (McDonald et al, 1995, S. 10).

Einigen Anlass zum Nachdenken über die mit der Feststellung des Auftretens von depressiven Störungen verbundenen methodischen Probleme gibt auch die Arbeit von Fichter, Xepapadakis, Quadflieg, Georgopoulou und Fthenakis (2004). Die von ihnen gefundene Zunahme psychischer Belastungen während einer Periode von 18 Jahren wirft zahlreiche Nachfragen auf. Eine davon könnte sein, ob Selbstberichte über psychische Zustände, auch wenn als identische betrachtete Messinstrumente verwendet werden, voraussetzungslos vergleichbar sind. Als vergleichbar sollten sie zweifellos gelten. Zu erörtern und zu untersuchen bliebe dann, welchen Einfluss etwa beobachtbare Veränderungen im Sprachgebrauch, in der öffentlichen Meinung und im

Selbstverständnis von Jugendlichen auf das jeweils gezeigte Antwortverhalten bei den Messungen ausüben. Die Bereitschaft von Jugendlichen, sich über ihre emotionalen Probleme zu äußern, könnte sich in relativ kurzen Perioden geändert haben. Diese These stützen Fichte et al. mit dem Hinweisen auf eine vergleichende Studie von Madianos und Mitarbeitern (1999) über die Einstellungen gegenüber psychischen Erkrankungen unter Einwohnern der griechischen Hauptstadt in den Jahren 1980 bis 1994. Über den Vergleich der Einstellungen zwischen den zwei Messzeitpunkten hat sich gezeigt, dass 1994 die Teilnehmer positivere und liberalere Einstellungen gegenüber psychisch erkrankten Menschen äußerten als das zuvor zu beobachten war. Sie erwiesen sich als toleranter gegenüber dem Auftreten psychischer Abweichungen. Außerdem waren 1994 die Befragten im Vergleich zu den aus der Stichprobe im Jahre 1980 besser über die Ursachen psychischer Störungen informiert. Fraglich bleibt dann wiederum, ob die verglichenen Stichproben als in allen relevanten Aspekten parallelisiert gelten durften. Die Möglichkeit des Auftretens unterschiedlicher sozio-kultureller Einflüsse auf Fremd- und Selbstwahrnehmung psychischer Belastungen wurden von vielen Autoren und so auch von McDonald (1995), Tsiantis (1994) und Motti-Stefanidi (1993) diskutiert.

Tsiantis beispielsweise geht davon aus, dass die Gesellschaft jeweils mitbestimmt, welches Erziehungsverhalten und welches Verhalten von Kindern oder Jugendlichen als „normal“ oder „pathologisch“ zu gelten hat. Nach Harkness und Super (1990) wären so interkulturelle Unterschiede im prozentual bestimmten Auftreten von Psychopathologien zu erklären.

In solchen Zusammenhängen erwähnen griechische Psychologen in der Diskussion ihrer Forschungsergebnisse oft eine Reihe von Veröffentlichungen über die Kindererziehung in Griechenland. Tsiantis (1982) erwähnt beispielsweise in einem Report der WHO über die psychische Gesundheit und psychosoziale Entwicklung von Kindern in Griechenland, dass die griechischen Kinder von ihren Eltern übermäßig geliebt werden, was aber mit der von den Eltern erhobenen Forderung nach Erfüllung ihrer Erwartungen durch das Kind verbunden sei. Empiriker dürften sich allerdings fragen, wie solche Aussagen empirisch geprüft werden können, was als freundlich formulierter Zweifel an der empirischen Fundierung solcher Aussagen zu verstehen wäre. Parker und Lipscombe (1981) gehen davon aus, dass die Überfürsorglichkeit griechischer Eltern schüchterne und zurückhaltende Persönlichkeiten aufbaue. Doumanis (1983) und Vasiliou & Vasiliou (1982; 1984) betonen in ihren Arbeiten die Auswirkungen geschlechtsspezifischer Erziehungspraktiken. Griechische Eltern ermutigten ihre Jungen unabhängig und anspruchsvoll zu sein und dabei ihren Problemen stoisch zu begegnen. Von den Mädchen werde erwartet, dass sie warmherzig agierten und ihre interpersonellen Beziehungen entwickelten. Die zitierten Publikationen folgen weitgehend den zu ihrer Zeit zu beobachtenden internationalen Trends in mehreren Bereichen der Psychologie und vermitteln die nach Griechenland.

Auf eine Zunahme an psychischen Störungen in der griechischen Gesellschaft verweisen allerdings auch epidemiologische Studien an unterschiedlichen Altersgruppen. Madianos und Stefanis (1992) haben eine statistisch signifikante Zunahme zwischen zwei Messpunkten im Jahre 1978 und 1984 an depressiven Störungen in allen geographischen Gebieten Griechenlands und erhöhte Prävalenzraten für Major Depression bei beiden Geschlechtern festgestellt. Die Untersuchungen beziehen sich aber nicht auf Schulkinder. Aus einer Reihe von Untersuchungen ergibt sich auch eine Zunahme von Problemen in Bereichen, die mit psychischen Belastungen verbunden sind. Kokkevi und Mitarbeiter (2000a) erwähnen beispielsweise, dass der Alkoholkonsum zugenommen hat, vor allem unter Frauen und Ierodiakonou (1998) spricht von einer Zunahme an Suizidversuchen. Ähnliches gilt für den Drogenkonsum bei Jugendlichen beider Geschlechter (Kokkevi et al., 2000b). Madianos und Stefanis (1992) sehen die Zunahme an psychischen Erkrankungen durch den gesellschaftlichen Wandel verursacht, der nicht nur politisch verursacht sei, sondern auch auf eine zunehmende Verstädterung und Industrialisierung

zurückzuführen sei. „The increase in psychopathology observed over time has to be seen also considering the contextual changes which took place in Greece and which had an influence on the population in general and on adolescents particularly” (Fichter et al., 2004).

Die im gegebenen Zusammenhang notwendige, so auch durchgeführte und hier nun abzuschließende Umschau in der aktuellen Literatur über psychologische Forschung in Griechenland war, in Übereinstimmung mit den Zielsetzungen der hier geführten Erörterung, auf Studien über emotionale Störungen bei Kindern und Jugendlichen in Griechenland gerichtet. Sie dient vorrangig dem Ziel, die nachfolgend anzusetzende, in ihrer Durchführung und ihren Ergebnissen darzustellende empirische Untersuchung zum Auftreten von depressiven Symptomen bei griechischen Schulkindern am gegebenen griechischen Forschungszusammenhang orientieren zu können.

Der um 2004 vorzufindende Forschungsstand über Depressionen bei Schulkindern weist zahlreiche Erkenntnislücken auf, die durch weitere empirische Untersuchungen zum Auftreten depressiver Störungen verringert werden könnten.

Die Untersuchung sollte dazu beitragen, den bestehenden Informationsstand über das Auftreten depressiver Störungen bei griechischen Kindern im Schulalter zu erhöhen. Ein solches Vorhaben stellte sich als sinnvoll dar, wenn das gegenwärtig verfügbare Wissen als ergänzungsbedürftig angesehen werden müsste. Es erübrigte sich, wenn verlässlich bekannt sein sollte, wie verbreitet das Auftreten depressiver Störungen unter griechischen Schulkindern ist.

Unter der Voraussetzung, dass die gezogene Literaturstichprobe eine hinreichende Entscheidungsbasis darstellt, sind dann die nachstehenden drei Feststellungen zu treffen:

1. die unmittelbar oder mittelbar mit dem Auftreten depressiver Störungen bei Kindern und Jugendlichen in Griechenland bis 2004 befassten psychologischen Studien belegen das Auftreten depressiver Störungen bei griechischen Kindern und Jugendlichen,
2. landesweite und im Hinblick auf sozioökonomische Basisdaten repräsentative Untersuchungen zur depressiven Belastung von Kindern und Jugendlichen in Griechenland fehlen bislang,
3. daraus ergibt sich die zum gegebenen Zeitpunkt an der Forschungslage ausgerichtete Begründung für die Durchführung einer solchen Studie.

Diesen Feststellungen versuchte die in den weiteren Ausführungen dargestellte Untersuchung zu folgen.

Kapitel 3

Bestimmung des Objekts einer Untersuchung zur Auftretenshäufigkeit depressiver Störungen bei griechischen Schulkindern

Bereits in den einleitenden Erklärungen zu den Erörterungen über das hier behandelte Untersuchungsvorhaben war das allgemeine Forschungsinteresse dargestellt worden und auch, dass in dieser Arbeit dem von Bortz und Döring (2002, S. 113 uf.) Verständnis der Funktion populationsbeschreibender Untersuchungen für epidemiologische Forschung gefolgt werden wird. Das ist hier, dem nun bis hier erreichten Stand der Erörterungen entsprechend, erneut zu bekräftigen. Nach Bortz und Döring sind populationsbeschreibende Untersuchungen für die epidemiologische Untersuchung von besonderer Bedeutung. Solche Untersuchungen verschaffen unter anderem Aufschluss über Prävalenz und Inzidenz. „Im allgemeinen Verständnis beziffert die Prävalenz die Verbreitung eines bestimmten Notstandes, einer Störung oder eines Phänomens und die Inzidenz das Neuauftreten des fraglichen Sachverhaltes.“ (ebd.) Indem eben dies auch die erklärte Absicht der beabsichtigten Untersuchung sein soll, ist vor der Aufnahme der eigentlichen Untersuchungsarbeit zweifelsfrei zu bestimmen, welcher Notstand gemeint ist und was dann folglich der Untersuchungsgegenstand sein wird.

Das scheint insoweit bereits hinreichend bestimmt zu sein, als das Auftreten depressiver Störungen bei griechischen Kindern zum Ziel gesetzt worden ist. Erfasst werden soll die Prävalenz der depressiven Störungen bei griechischen Kindern. Dass ein Auftreten depressiver Störungen ein gesellschaftlicher Notstand im Sinne von Bortz und Döring ist, wird an dieser Stelle axiomatisch unterstellt, selbst, wenn das in zahlreichen Gesellschaften der Gegenwart nicht so verstanden werden sollte, weil andere Notstände, wie etwa Aids, Hunger oder kriegerische Auseinandersetzungen in Teilen Afrika als ungleich härtere Plagen erlebt werden. Notstände relativieren einander.

Das Ziel, die Prävalenz depressiver Störungen bei griechischen Kindern im Schulalter erfassen zu wollen, nötigt dazu, diesen gemeinten Gegenstand zu operationalisieren. Ohne eine solche Operationalisierung könnten depressive Störungen nicht erfasst werden. Von ihrer Existenz wäre empirisch nicht mehr bekannt als ein sie benennender Begriff, der ihnen eine sprachliche Repräsentanz verschafft, aber keine empirisch fassbare Existenz. Vom Grundsatz her ist dies keine andere Situation als die, in der sich beispielsweise Skinner während seiner Tierversuche mit der Skinner-Box befand. Dass Menschen auch Tauben Hunger zubilligen, schafft für das, was Skinner bei entsprechenden Konditionierungsversuchen als Verstärker einsetzen wollte, eine sprachliche Repräsentation, aber eben keine weitergehende empirische Realität. Erst über eine Operationalisierung wird eine solche Realität bekannt. Indem Hunger mit Abwesenheit von Futter in Verbindung gebracht wird und dafür der Begriff der Deprivation Verwendung findet, scheint auch eine Möglichkeit zur Operationalisierung erkennbar. „Die 24stündige Abwesenheit von Futter“ könnte als Operationalisierung gewählt werden. Das operationalisierte Ereignis lässt sich empirisch verlässlich erfassen. Das Beispiel gibt auch Gelegenheit, einen bei Operationalisierungen häufig auftretenden und zumeist unerwünschten Nebeneffekt zu veranschaulichen. Wenn ursprünglich nach einer Operationalisierung von Hunger gesucht worden sein sollte, dann darf die exemplarisch benutzte Operationalisierung auch als Beispiel für eine durch diesen Vorgang verursachte Konzeptverschiebung betrachtet werden. Die mehrstündige Abwesenheit von Futter korreliert möglicherweise mit Hunger oder Verlangen nach Nahrung, ist dafür aber auch nicht identisch mit Hunger.

Der für den Ansatz einer empirischen Untersuchung unumgängliche Prozess der Operationalisierung des Gegenstandes der Untersuchung wird zweckmäßigerweise durch das Aufsuchen des vorherrschenden Verständnisses für das Untersuchungsobjekt begonnen.

Um das hier dann irgendwann nachfolgend aufgabengerecht und vor allem auch begründet leisten zu können, erscheint ein Eingehen auf Definition und Klassifikation depressiver Störungen unumgänglich, obschon auch absehbar scheint, dass dies keine hier übernehmbare Operationalisierung bieten wird.

3.1 Depressive Störungen bei Kindern und Jugendlichen als historischer Gegenstand

Ein die Entwicklung des Verständnisses depressiver Störungen bei Kindern suchender Rückblick bedarf unumgänglich der Nennung eines Suchbegriffes. Ein früheres Verständnis kann in entsprechenden Texten enthalten, genannt, beschrieben und gewürdigt worden sein, auffindbar wird es, wenn Begriffe bekannt sind, denen es zugehört. Der Begriff der Melancholie scheint, weil hinreichend tradiert, ein geeigneter Suchbegriff zu sein. Das Erwachsene betreffende Störungsbild der Melancholie ist seit antiken Zeiten bekannt. Über vergleichbare Störungen bei Kindern und Jugendlichen wird demgegenüber erst seit zwei Jahrhunderten als Gegenstand der Psychiatrie geschrieben. Das muss nicht verwundern, Krankheiten oder Störungsbilder sind Konstrukte, die in ihrer Existenz und Bedeutung dem sozialen und gesellschaftlichen Wandel unterworfen bleiben. Charcot schrieb 1886: „Krankheit ändert sich nicht. Wir ändern uns, lernen mehr über die Krankheit und nehmen sie anders wahr.“ (aus Mastaler, 2001, S. 5). Nach Charcot wäre der Krankheit eine überdauernde Existenz zuzuschreiben, die aber keineswegs zwingend zum Erkennen ihrer Existenz durch Menschen führen müsste. Das gilt für Melancholie oder depressive Störungen bei Kindern auch. Des Weiteren erschweren depressive Störungen ihr Erkennen, weil die störungsspezifischen Symptome, so etwa bei Trauer, temporär auch in Zusammenhang mit unterschiedlichen Lebensereignissen außerhalb einer psychiatrischen Symptomatik auftreten können. Das bei der Diagnose depressiver Störungen als Hauptkriterium beachtete Vorkommen niedergedrückter Stimmung und der ebenso kritisch bewertete Aktivitätsverlust lassen sich außerhalb von bestehenden Störungen auch als temporäre reaktive Äußerungen einer Persönlichkeit auf speziellen, meist belastenden Lebensumständen wahrnehmen. Insoweit wäre solchen Phänomenen auch keine psychiatrische Bedeutung zuzuschreiben. Die Grenzen zwischen klinischen und nicht-klinischen depressiven Zuständen sind schwer zu bestimmen und noch schwerer zu erfassen. Zudem werden die Kernsymptome von mehreren Nebensymptomen Besonderheiten begleitet, die auch in anderen Störungsbildern einen kennzeichnenden Symptomwert haben. Die im Vergleich zur Ausbildung der Psychiatrie für Erwachsene relativ späte Entwicklung der Pädopsychiatrie als einem eigenen Bereich ist eine der konstitutiven Bedingungen für die historische Verzögerung im Auftreten von Beschreibungen der Melancholie bei Kindern und Jugendlichen. Noch im 19. Jahrhundert wurden emotionale Probleme und Verhaltensauffälligkeiten von Kindern vornehmlich als Reaktionen auf Erziehungs- und Schulschwierigkeiten verstanden (Nissen, 2002). Erst zu Beginn des 20. Jahrhunderts setzte die systematische wissenschaftliche Beschäftigung mit psychischen Problemen bei Kindern ein (Groen und Petermann, 2002). Möglicherweise hat diese Feststellung nur begrenztem Erklärungswert, denn schon aus der Mitte des 19. Jahrhunderts sind Auffassungen wie die von Friedreich bekannt (Trott, 2002), nach der alle an Erwachsenen beobachtbaren psychischen

Störungen auch bei Kindern auftreten können. In der ersten Phase der geschichtlichen Entwicklung des Konzeptes von Depression im Kindesalter wurde dementsprechend auch das seit Jahrhunderten benutzte Konzept der Melancholie als Gemütskrankung nun auch auf Schulkinder angewendet. Die in ihrer Zeit bekannten und geschätzten Psychiater Schüle, Emminghaus und Maudsley vertraten entsprechende Krankheitsbilder (Rossmann, 1991). Groen und Petermann (2002) erwähnen, dass die Entstehung der These von den belastenden Lebensereignissen als Ursache emotionaler Probleme bei Kindern auf die wissenschaftliche Diskussion in jenen Jahren zurückgeht. Insbesondere frühe Verlusterlebnisse wurden als belastende Lebensereignisse angesehen. Viel zitiert werden die Äußerungen des in seiner Zeit sehr bedeutenden Berliner Psychiaters Griesinger (1817-1868), der sich Mitte des 19. Jahrhunderts auch um die ärztliche Versorgung depressiver Kinder bemüht hatte und Behandlungsvorschläge entwickelte: „Auch melancholische Erkrankungen können, wenngleich sehr viel seltener, in all ihren Variationen auch im Kindesalter auftreten.“ (aus Trott, 2002, S. 129). Die von Griesinger vertretenen Auffassungen verbreiteten sich und fanden Beachtung. Berichtet wird von der Existenz pädagogischer Einrichtungen in der Umgebung Berlins, die sich auch für die Aufnahme und Betreuung melancholischer Kinder und Jugendlicher anboten.

Die gegen Ende des 19. Jahrhunderts einsetzende zweite historische Phase des Umgangs mit depressiven Belastungen bei Kindern wurde von dem wachsenden Einfluss der Psychoanalyse bestimmt. Begünstigt wurde das auch durch das Desinteresse der sich inzwischen als eigene Wissenschaftsdisziplin etablierten akademischen Psychologie an emotionalen Problemen. Die in positivistischer Weise um Wissenschaftlichkeit bemühte Psychologie fand keine ihren eigenen Ansprüchen genügenden Methoden zur Erfassung von Emotionen. Solche Methoden waren etwa im Bereich der Wahrnehmungspsychologie entdeckt worden. Die klinisch orientierte Psychoanalyse sah für den Umgang mit Emotionen und emotionalen Störungen methodisch keine unüberwindlichen Probleme. Der Psychoanalyse wurde schließlich, allein schon wegen ihres Engagements, höhere Kompetenz im Verständnis psychischer Störungen zuerkannt als der damaligen akademischen Psychologie. Für den Umgang mit depressiven Störungen bei Kindern führte das zu vermehrter Beachtung der psychoanalytischen Lehrmeinungen. Die Neurosenlehre der Psychoanalyse bewegte nicht nur die medizinische Fachwelt. Sigmund Freud (1856-1939) begründete die These von der durch das Über-Ich regulierten Entwicklung der Depression. Da nun psychoanalytisch aber andererseits das Erreichen der vollen Funktionstüchtigkeit des Über-Ichs erst für höhere Altersstufen in der Entwicklung junger Menschen angenommen wurde, ergab sich daraus zwingend die Unmöglichkeit der Entstehung von Depressionen im Kindesalter. Dieses Dogma führte zu dem für die erste Hälfte des 20. Jahrhunderts kennzeichnenden Umgang mit Störungsphänomenen im Kindesalter. Als innerpsychischer Anlass zur Entwicklung einer depressiven Störung und damit auch als Auslöser des speziellen pathogenisierenden Prozesses wurde in Übereinstimmung mit Freud der reale oder symbolische Verlust eines geliebten Objektes, vor allem einer geliebten Person, angesehen. Eine danach aufkommende Aggressivität wird, wenn sie durch den Verlust gegen das geliebte Objekt gerichtet worden sein sollte, vom Über-Ich nicht toleriert. Es sanktioniert den aggressiven Verstoß gegen das Liebesgebot durch Schuldzuschreibungen an die durch Verlusterleben zum Aggressor gewordene Persönlichkeit. Eine über das Schulterleben ausgebildete Selbstaggression ist der psychopathologische Kern der depressiven Störung. Sie ist letztlich eine selbstauferlegte Busse für das eigenmächtige und aggressive Übertreten des auch bei erlebtem Verlust gültig bleibenden Schutzgebotes für eine geliebte Person. Die aggressiven Gefühle gegen die geliebte Person kehren als Selbstaggressionen

zu dem Aggressor zurück und bestrafen ihn. Da Kinder sich noch in einem psychosexuellen Stadium befinden, in dem das Über-Ich funktionsuntüchtig ist, entwickelt sich eine für Aggressionen gegen verloren geglaubte geliebte Personen vorhergesagte depressive Reaktion nicht. Als Vertreter dieser Position werden in der einschlägigen Literatur Mahler (1961), Rie (1966) und Rochlin (1959) genannt.

Interessant ist in diesem Zusammenhang eine von Rossmann gegebene Anmerkung zu den Konsequenzen aus den psychoanalytischen Erklärungen: „Die infantilen Formen wurden dabei keineswegs als weniger ernst oder schwerwiegend betrachtet, sondern als ein spezifischer und mit den Erwachsenenformen nicht vergleichbarer Ausdruck der universellen Konflikte im Zusammenhang mit Separation und Objektverlust.“ (Rossmann, 1991, S. 15).

Auch für die in der Psychoanalyse vertretenen Sichtweisen zur Depression ist im Übrigen dem immer wieder angebrachten Hinweis auf die Heterogenität der psychoanalytischen Lehrmeinungen zu folgen. Abweichungen von der klassischen These über die Ursachen der Depressionen finden sich beispielsweise in den Schriften von Melanie Klein (1960), von Malmquist (1971) und von Sandler und Joffe (1965) (vgl. dazu Trott, 2002; Hoffmann, 1991).

Die Überzeugungen, die psychoanalytischen Debatten über Depressionen und insbesondere über depressive Belastungen von Kindern für die ersten Jahrzehnte des 20. Jahrhunderts ermöglichte, relativierte die dann weltweite Beachtung erreichende empirische Forschung von R. Spitz und K. Wolf (1946). Von ihr meinte Hoffmann (1991), sie sei ein Meilenstein in der Forschungsgeschichte. Diese Einschätzung ist in mehrfacher Hinsicht berechtigt. Mit der von R. Spitz durchgeführten Untersuchung wurde das Primat eines Dogmas durch die Rückkehr zur empirisch geleiteten Orientierung in der Forschung zur Depression bei Kindern ersetzt. Die Arbeiten von R. Spitz relativierten darüber hinaus das psychiatrische und psychologische Denken in der Psychopathologie insoweit, als sie eindringlich die starke Begrenztheit des Erklärungswertes empirisch ungesicherter Theorien exemplarisch verdeutlichten.

Die durch hohe Sterblichkeit unter hospitalisierten Kindern veranlasste Untersuchung von R. Spitz führte zur Einführung des von ihm gebildeten Konzeptes der „anaklitischen Depression“. Die damit beschriebene theoretische Position basierte auf den Beobachtungen des Verhaltens hospitalisierter Kleinkinder, die nach dauerhafter Trennung von den engsten Bezugspersonen, ihrer Mutter, überdauernde Reaktionen von Traurigkeit zeigten, mit Weinen, Apathie, Schlaflosigkeit, Gewichtsverlust und Zurückgezogenheit reagierten. Die progressive Entwicklung des Störungsbildes führt zu schweren Entwicklungsschäden bis hin zu Todesfällen. Die trotz der Altersdivergenzen beeindruckende Ähnlichkeiten der Symptomatik dieser Kleinkinder mit vergleichbaren Störungen von Erwachsenen bestärkte Spitz in der Überzeugung, dass die von ihm beobachteten pathogenen Phänomene belege dafür sind, dass Kinder unter depressionsähnlichen Zuständen leiden können (Groen und Petermann, 2002; Hoffmann, 1991).

Eine dritte Phase in der seit dem 19. Jahrhunderte ablaufenden Neuorientierung im Verständnis emotionaler Störungen bei Kindern und Jugendlichen ging von einer um 1960 in den USA einsetzenden Entwicklung aus und verbreitete sich schnell weltweit. Diese Entwicklung wurde durch den in jenen Jahren in den USA gegenüber der Psychoanalyse deutlich zunehmendem Einfluss der behavioralen Verhaltenstherapie ausgelöst und getragen. Eine in Europa schon im 19. Jahrhundert vertretene Auffassung setzt sich, nun aber von empirischen Befunden gestützt, erneut durch. Danach können die bei Erwachsenen zu beobachteten psychischen Störungen grundsätzlich

auch bei Kindern und Jugendlichen auftreten. Dementsprechend ist das Auftreten depressiver Störungen, so wie es von Erwachsenen her bekannt ist, grundsätzlich auch bei Kindern und Jugendlichen möglich. Eine in den Anfängen dieser Phase favorisierte Konzept ist das der „maskierten Depression“, auch als Konzept „depressiver Äquivalenten“ besprochen (Cytryn and McKnew, 1972; Glaser, 1968). Die depressive Symptomatik bei Kindern würde, den Annahmen zufolge, durch andere Verhaltensweisen überlagert werden und sei somit verdeckt. Aus dieser Deckung heraus wirkt „die maskierte Depression“ als Verursacher vielfältiger Belastungen. Sie wurde so als verantwortlich für Verhaltensstörungen gesehen, aber ebenso auch für das Auftreten von somatischen Beschwerden, Angstzustände, delinquentem Verhalten oder Schulschwierigkeiten. (Essau, 2002; Reicher, 1998; Rossmann, 1991). Die behauptete pathogenisierende Omnipotenz der „maskierten Depression“ nötigte schließlich zur Aufgabe des Konzeptes, weil eine Ursache, die alles erklärt, nichts mehr erklären kann. Die maskierte Depression ließ sich nicht mehr differentialdiagnostische von anderen Störungsbildern unterscheiden.

Die Diskussion über die Existenz depressiver Zustände bei Kindern und Jugendlichen veränderte auch in Europa die Perspektiven. War zu Beginn der dritten Phase noch umstritten, ob Depressionen bei Kindern in beachtenswertem Umfang und wie bei Erwachsenen auftreten, so diskutierte man auf dem 4. Kongress der Europäischen Pädopsychiatern (im August 1971) bereits andere Thesen, nämlich die, dass „depressive states account for an important and relatively large share of mental disorders in children and adolescents“ (Annell, 1971, aus Hoffmann, 1991, S. 18).

Die massive Kritik an dem Konzept der verdeckten Depression (Carlson & Cantwell, 1980b; Cytryn, McKnew & Bunney, 1980; Kovacs & Beck, 1977) verhalf der These von der grundsätzlichen Übereinstimmung zwischen Depressionen bei Kindern und Jugendlichen einerseits und Erwachsenen andererseits zum Durchbruch. Den bei Kindern und Jugendlichen auftretenden Depressionen wurde klinische Bedeutung zuerkannt. Diese Position spiegelt sich in der Tatsache wider, dass in der dritten Fassung des DSM (1980, APA) bereits die gleichen Kriterien sowohl für die Diagnose depressiver Störungen bei Erwachsenen als auch bei Kindern und Jugendlichen verwendet werden. Nicht erst seit diesem Jahr nimmt die Zahl der Publikationen über Depressionen bei Kindern und Jugendlichen ständig zu (siehe dazu auch Rossmann, 1991).

Mit dem Erscheinen des DSM-IV wird einerseits die altersübergreifende Gleichsetzung der Hauptkriterien für depressive Störungen gefestigt, andererseits bahnen sich auch spezifische Unterscheidungen zwischen den für Erwachsene oder für Kinder und Jugendliche geltenden Kriterien an. Beispielsweise führt die Störung der Regulation des Körpergewichtes bei Erwachsenen zu anderen Konsequenzen für das Körpergewicht als entsprechende Regulationsstörungen bei Kindern. Die Veränderungen setzen bei Erwachsenen an einem mehr oder weniger stabilisierten Körpergewicht an. Bei Kindern interagieren Störungen der Gewichtsregulation mit einem generell auf Massenzuwachs eingerichteten Organismus. Die durch das Wachstum bei Kindern gegenüber Erwachsenen veränderte alterstypische Normalität kann bei mehr oder weniger identischen Störungen, wie sie Gewichtsanomalien gegeben wären, zu unterschiedlichen Symptomvarianten führen. So wäre bei Kindern beispielsweise „...das Ausbleiben der im Rahmen der altersgemäßen Entwicklung zu erwartenden Gewichtszunahme auffallend...“ (Essau und Petermann, 2002). Auch für die Diagnose einer dysthymen Störung führt DSM-IV eine Änderung ein, die in Beziehung zu den besonderen Entwicklungsumständen bei Kindern und Jugendlichen steht. Das für die Diagnose einer dysthymen Störung bei Erwachsenen geltende zeitliche Kriterium, nach dem Symptome mindestens für zwei Jahre

aufgetreten sein müssten, wird für Kinder und Jugendliche auf ein Jahr herabgesetzt (Essau und Petermann, 2002).

Im Vergleich der etwa um das Jahr 2000 erschienenen Publikationen zum Auftreten von Depressionen bei Kindern und Jugendlichen mit denen früherer Jahre ist das inzwischen erreichte Ausmaß des übereinstimmenden, fast homogenen Bezuges der Darstellungen auf DSM-IV und ICD-10 unverkennbar. Das ergibt sich auch im internationalen Vergleich. Die international wirksame Orientierungsfunktion dieser Manuale ist überaus deutlich nachweisbar. Die internationale Zusammenarbeit, um die sich insbesondere die WHO, die Gesundheitsorganisation der Vereinten Nationen, bemüht hatte, zeigt, dies auch dank der weltweiten Unterstützung nationaler Stellen, nachhaltige Wirkung. Bei diesem Stand in der Entwicklung der wissenschaftlichen Diskussion zum Auftreten von Depressionen bei Kindern und Jugendlichen empfiehlt sich auch, die für den Ansatz einer epidemiologischen Forschungsarbeit unerlässliche Operationalisierung an den in DSM-IV und ICD-10 aufgeführten diagnostischen Kriterien zu orientieren. Deren augenscheinlich wirksame ordnende Funktion könnte allerdings auch zu der Annahme verleiten, dass die dem Untergang mit Depressionen unterliegenden wissenschaftlichen Probleme durch das Erreichen von internationalen Übereinkünften zu lösen wären oder sogar schon gelöst worden seien.

3.2 Definition und Klassifikation depressiver Störungen bei Kindern und Jugendlichen

Die mit DSM-IV und ICD-10 als aktuellen Handbüchern repräsentierte Entwicklung einer wissenschaftlichen Neuorientierung zum Störungsbild der Depressionen bei Kindern und Jugendlichen symbolisiert das in vieler Hinsicht überaus hilfreiche Erreichen aktuell wirksamer internationaler Übereinkünfte. Diese Übereinkünfte umfassen implizit oder explizit die Zusicherung für eine nahe Zukunft die vereinbarte Klassifikationen und Symptomlisten zu verwenden und zu prüfen. Die bisherige Praxis des Umganges mit DSM und ICD sollte im übrigen insoweit fortgesetzt werden, als jede aktuelle Ausgabe mit dem Tage ihres Erscheinens zur Überprüfung der Notwendigkeit einer nächsten Revision ansteht. Die Einführung gemeinsam benutzter Klassifikationssysteme und Symptomlisten ist eines von vielen Hilfsmitteln für die Fortsetzung der wissenschaftlichen Arbeit in allen Bereichen des Umgangs mit Depressionen bei Kindern und Jugendlichen. Zahlreiche Probleme sind ungelöst, weitere werden entdeckt und alle stehen, falls das für den jeweiligen Fall nicht doch schon erreicht worden sein sollte, zur Bewältigung an.

Kendell schrieb 1976, dass die in der internationalen pädopsychiatrischen Literatur dargestellte Versuche, die unterschiedliche Bedeutungsebene des Depressionsbegriffes zu bestimmen, genauso problematisch sind, wie die entsprechende im Erwachsenenbereich. Trotz aller Entwicklungen seitdem, ergeben sich noch heute aus den unterschiedlichen Studien über Depressionen bei Kindern und Jugendlichen inkonsistente Ergebnisse, die die aktuelle, aus den internationalen nosologischen Systemen gegebene Definition von depressiven Störungen bei Kindern in Frage stellen (Hoffmann, 1991; Reicher, 1998). In den folgenden Erörterungen wird darauf eingegangen werden.

Ein unbewältigtes Grundproblem ist die mit jeder Nutzung eines jedem Klassifikationssystems entstehende Nötigung zur individuellen Interpretation der klassifikatorisch verwendeten Begriffe. Eine Nötigung zur individuellen Interpretation entsteht, weil jeder, der das System anwenden will, für sich bestimmen muss, welche Bedeutung dem jeweils verwendeten Begriff zu unterlegen ist, beispielsweise dem „einer niedergedrückten Stimmung“, wobei „niedergedrückte Stimmung“ bereits eine deutschsprachige Variante der „depressiven Stimmung“ ist. Die Möglichkeit besteht anzunehmen, in Deutschland würden alle mit der Diagnose von Depressionen befassten Persönlichkeiten in der Zuordnung von berichteten oder beobachteten Verhaltensphänomenen zu diesem Begriff übereinstimmen. Ebenso könnten aber auch begründete Zweifel an dem Erklärungswert solcher Annahme vertreten werden.

In der um Depressionen bei Kindern und Jugendlichen bemühten psychologischen Literatur des ausgehenden 20. Jahrhunderts konzentrierten sich Erörterungen der Bedeutungsproblematik auf die Interpretation des Begriffes der Depression. So sah Pearce (1976) beispielsweise drei Bedeutungen des Begriffes „Depression“. Damit sei erstens ein durchaus alltäglich möglicher und somit gewissermaßen auch „normaler“ Abfall gemeint, der aber auch als Ausfall, als Absturz oder schlicht als negativ gerichtete Abnahme der Stimmungslage zu verstehen sei, falls nicht doch und spezieller das Ausbleiben einer im situativen Zusammenhang erwartbaren emotionalen Reaktion der Gegenstand gewesen sein könnte. Der mit depressiver Stimmungslage notgedrungen befasste Diagnostiker hätte zu entscheiden, ob in einem von ihm betrachteten aktuellen Fall, die depressive Stimmung so, wie hier beschrieben, verstanden werden müsste. Trotz der Verwendung eines als eindeutig zu kennzeichnenden Begriffes ist der Ermessensspielraum des Diagnostikers relativ weit. Im interpersonalen Vergleich von Diagnosen könnte sich die Streuung der Ermessensspielräume vervielfachen. Die diagnostischen Unschärfen verringern sich auch nicht, wenn im diagnostischen Interview der Klient gefragt werden sollte, ob der von ihm wahrgenommene depressive Zustand eher als „normal“ einzustufen sei. Depressive Menschen könnten eine solche Frage empirisch korrekt bejahen, weil die Häufung des Auftretens vergleichbarer Zustände eine entsprechende Normalität herstellt. Pearce (s.o.) verwies zweitens auf eine vom Alltäglichen und dessen Normalität abzuhebende Bedeutungsebene der „Abnormität der Stimmungslage“. Die wahrgenommene Abnormität der Stimmungslage stellt sich allein schon her, indem sie selten auftritt. Pearce ordnet sie daraufhin einem speziellen Syndrom zu. Die abnorme Stimmungslage führt zu Beeinträchtigungen. Drittens kann die depressive Stimmungslage einer entsprechenden Krankheit zugeordnet sein, in der die depressive Stimmung sich qualitativ von der „normalen“ Stimmung unterscheidet, deutliche Ursachen erkennen lässt und Vorhersagen gestattet. Anzumerken bleibt, dass jede der verwendeten Bestimmungen dieses Ordnungssystem die Bedeutungsproblematik nicht aufhebt, sondern im Gegenteil an jeder begrifflichen Position ganz neue und eigene Bedeutungsprobleme einführt.

Zu erwarten wäre, dass auch jedes weitere, mit der Zielsetzung einer Auflösung der Bedeutungsproblematik vorgetragene kategoriale System seine eigene Bedeutungsproblematik mit sich bringt.

Gittelman-Klein (1976) beispielsweise unterschied zwischen der alltäglichen und der zweideutigen wissenschaftlichen Verwendung des Wortes. Im Gegensatz zu dem umgangssprachlichen Verständnis von Depression als „...referring to a subjective mood state exclusively....referring to unhappiness or misery“ hat das selbe Wort in der wissenschaftlichen Sprache zwei andere Bedeutungen. Diese sind: “on the one hand it refers to a state, trait or symptom, which may occur in varying severity as a secondary complication in a diversity of mental or physical disorders.... On the other hand, depression is a pathological clinical entity that

regularly implies a reduction in hedonic level, loss of interest, and a diminution in one's sense of competence, all in the context of a dysphoric state or psychic pain. This state is accompanied by a number of other symptoms that have been given differing emphases by various investigators." (Gittelman-Klein, 1977, zitiert nach Angold, 1988, S. 602).

Von System zu System wechseln die Begriffe. Da jeder Begriff einen Bedeutungshof hat, kann das Bedeutungsproblem nicht durch das Auswechseln von Begriffen aufgelöst werden.

Graham (1980) schlägt in Bezug auf psychische Phänomene allgemein vor, zwischen einer Krankheit, einer Störung und einem Syndrom zu unterscheiden. Für depressive Zustände schließt Graham den Begriff der Krankheit aus, da es weder bei Erwachsenen noch bei Kindern möglich ist, physiologische und biochemische Prozesse als ausschließliche Ursache für die Entstehung von Depressionen zu identifizieren. In einer von Laux (2002) herausgegebenen und als wissenschaftliche Bestandsaufnahme um das Jahr 2000 gedachten Schrift über Depressionen findet sich eine identische Position vertreten. In Bezug auf Depressionen bei Kindern darf nach Graham auch nicht von Störungen gesprochen werden. „Dazu wäre es notwendig zu zeigen, dass Kinder mit depressiven Störungen aufgrund sozialer, biochemischer, genetischer und therapeutischer Faktoren reliabel und valide von Kindern mit anderen Störungen unterschieden werden können“ (Hoffmann, 1991, S. 31, dort zitiert nach Graham, 1980, S. 288), das wären Voraussetzungen, die nach Graham sehr wahrscheinlich nicht bei allen depressiven Kindern als erfüllt anzusehen wären. Graham schlägt vor, für Depressionen bei Kindern nur den Begriff des Syndroms, als das überzufällig gleichzeitige Auftreten von Einzelsymptomen, zu verwenden. Auf dieser Ebene setzt Graham zwischen die Begriffe der depressiven Reaktion und depressiven Haltung gegeneinander. Eine depressive Reaktion folgert aus belastenden Ereignissen und hört nach einer „Lösung“ der problematischen Situation auf. Demgegenüber wäre die depressive Haltung vor allem als ein kommunikatives Signal mit Appellfunktion zu interpretieren. Es verlangt Zuwendung oder Hilfe.

Auch Graham lost erwartungsgemäß die Bedeutungsproblematik nicht auf. Im Übrigen steht er in Dissens zu DSM-IV und ICD-10. Für den nach vertretbaren Operationalisierungen suchenden Empiriker könnte sich damit die Bezugnahme auf Graham verbieten. Aber insbesondere für den Empiriker unterscheiden sich die Positionen der Handbücher auf der einen und die von Graham auf der anderen Seite vor allem durch ihren sozialen und gesellschaftlichen Status. Die eine Position ist beispielsweise durch die WHO gestützt, Graham aber nicht. Weder das eine noch das andere ist ein Nachweis für den jeweils mit der einen oder der anderen Position erreichten speziellen Erklärungswert für Depressionen bei Kindern und Jugendlichen.

Auch Angold (1988) setzt unterschiedliche Weisen der Bestimmung von Depressionen gegeneinander. Er verweist auf sieben unterscheidbare Fassungen: (1) „...a description of the low end of the ordinary fluctuations of normal mood...“, (2) „...a description of the unhappiness, sadness or psychic pain felt as a response to some unpleasant situation or event...“, (3) „...a trait-referring to a continuing, relatively stable, anhedonic state characterising an individual, as in the concept of depressive personality disorder“, (4) „...an individual symptom – referring to a pathological low or sad mood“, (5) „...a syndrome...“, (6) „...a disorder...“, (7) „...a disease...“ und (8) „...the cause of handicap...“ (Angold, 1988, S. 602-603).

Wer nach Prävalenzen für das Auftreten depressiver Störungen bei Kindern und Jugendlichen zu suchen gedenkt, findet sich mit seinen Bemühungen um den operationalisierten Ansatz seiner

Untersuchung mit den Prävalenzen des Auftretens konkurrierender Bestimmungen des gemeinten Objektes konfrontiert. Die Übereinstimmung im Untersuchungsdesign beschränkt sich häufig nur auf die in der Verwendung von Oberbegriffen. So scheint der Begriff „Diagnose der depressiven Störung“ auf hohe Gemeinsamkeiten zu verweisen. Ob die in der Durchführung der Diagnose auch bestanden haben, ist ungewiss oder sogar fraglich. Der Begriff Diagnose steht für eine Kategorie oder eine Klasse von Tätigkeiten, die in der diese Klasse konstituierenden Bestimmung miteinander übereinstimmen, deshalb gehören die einzelnen Mitglieder dieser Klasse an. Im übrigen aber unterscheiden sich Mitglieder dieser Klasse voneinander und zum Teil auch erheblich. Die Unterschiede betreffen potentiell sowohl die speziellen Zielsetzungen als auch die zum Zielerreichen auszuführenden Arbeitsschritte und weiterhin vieles mehr. Die Gleichsetzung des Arbeitsablaufes wird über die gleichsinnige oder synonyme Operationalisierung des Handlungsablaufes zu erreichen versucht. Die Bestimmung des oder der Objekte, die Untersuchungsgegenstand sein sollen ist, da Tätigkeiten von der Zielsetzung her zu entwickeln sind, die grundlegende primäre Operationalisierung. Falls empirische Untersuchungen nach unterschiedlichen Objekten suchen sollten, könnten die Ergebnisse aus beiden Arbeiten nicht unmittelbar als Informationen über ein identisches Objekt betrachtet werden. Die operationale Übereinstimmung von Untersuchungsobjekten ist zwar noch keine hinreichende, aber immer eine unersetzliche Bedingung für die Sicherung der unmittelbaren Vergleichbarkeit der Ergebnisse aus empirischen Untersuchungen.

Nun ergibt sich aber, dass, in welchen Zusammenhängen auch immer betrachtet, der Verdacht zu begründen ist, die operationale Bestimmung des Objektes einer auf depressive Störungen gerichteten populationsbeschreibenden Untersuchung werde durch nicht tolerierbare persönliche Freiräume bei den zu treffenden Entscheidungen belastet. Das sei wegen des Fehlens inhaltlich valider Entscheidungskriterien für die zu vollziehenden Operationalisierungen gegenwärtig nicht abweisbar. Beispielsweise wäre ein Symptom unter Verwendung der Begriffe „Stimmung“ oder auch „Gefühlslage“ oder in ähnlicher Weise für empirische Arbeiten operational nicht zureichend ausdifferenziert. Entschieden werden müsste jeweils, wie denn der in der Symptombeschreibung auftretende Gegenstand empirisch erfasst werden soll.

Wenn schon die Bestimmung dessen, was, in Übereinstimmung mit dem ausgewählten Symptom, in welcher Weise zu messen ist, in erhebliche Probleme führt, vervielfachen sich die, wenn bereits in der Ausfüllung der Störungsbilder und in der Erstellung der Symptomlisten deutliche Widersprüche auftreten sollten. Indizien dafür sind gegeben. So sind die durch die führenden Handbücher, DSM und ICD, zur Depression bei Kindern und Jugendlichen vertretenen Positionen empirisch keineswegs schlüssiger belegt als die von Graham (s.o.) verteidigten Auffassungen.

Unterschiedliche Grade der Übereinstimmung werden erfassbar, wenn die Vergleiche unter Verwendung von Ordnungskategorien wie denen der nosologischen Ebene, der Syndrom- und der Symptomebene verwendet werden. (Carlson & Cantwell, 1980a; Compass & Hammen, 1996; Kazdin, 1981). Der Grad der Übereinstimmung zwischen unterschiedlichen Positionen nimmt in absteigender Richtung der Aufzählung zu. Zwischen Graham und den Ausführungen in DSM-IV und ICD-10 besteht auf Symptomebene der höchste Grad von Gemeinsamkeiten. Zu welcher Konsequenz eine solche Feststellung im Hinblick auf die Absicht führt, verantwortbare Operationalisierungen für den Ansatz einer um Populationsbeschreibung bemühten Untersuchung nach Prävalenzen suchenden Untersuchung zu finden, ist zunächst noch nicht zwingend

abzusehen. Eine denkbare Konsequenz könnte sein, sich ausschließlich an Symptomlisten zu orientieren.

Um darauf eingehen zu können, ist die Wiedergabe der Begriffe erforderlich, mit denen in den beiden Handbüchern das Auftreten von Depressionen bei Kindern klassifikatorisch und symptomatologisch zu repräsentieren versucht wird.

Vor den notwendig werdenden weiteren Erörterungen erscheint nützlich, die Situation zu bestimmen, in der die alles in allem doch recht aufwendige, weil Ressourcen bindende, Suche nach Operationalisierungen, die als populationsbeschreibende und zugleich depressionsspezifische Kriterien zu erkennen sind, aufgegeben werden darf. Eine inhaltlich korrekte, praktisch aber wenig hilfreiche, weil die schon in der Frage verbalisierten Suchkriterien wiederholende Antwort könnte angeben, die Suche sei zu beenden, wenn die den Kriterien entsprechende Operationalisierung gefunden worden sei. Wo solche Operationalisierungen auffindbar sein könnten, ergäbe sich aus der Antwort nicht. Wenn daraufhin der Verweis auf die in diagnostischen Handbüchern zu finden Symptomlisten verfolgt werden sollte, wären die schon bekannten Probleme erneut zu finden. Die immer wieder neu auftretende Schwierigkeit kann am Beispiel verdeutlicht werden. In einer Symptomliste könnte etwa die nachstehende Formulierung zu finden sein: "...Absinken der Aktivitäten unter das zuvor bestehende (...beobachtete, gegebene, existierende...) persönliche Niveau...". Für den Ansatz einer darauf zu beziehenden empirischen Untersuchung wäre dann noch immer über zahlreiche Modalitäten der Aufnahme von Daten zu entscheiden. Die Bestimmungen, nach denen solche Entscheidungen vollzogen werden, liegen, da sie nicht im Symptom definiert sind, im Belieben der diese Aufgabe übernehmenden Persönlichkeiten. Sie hätten jeweils für ihre Untersuchung zu entscheiden, wie die Validität der Messungen und deren weitere Messqualitäten gesichert werden könnten und sollten. Zu solchen Entscheidungen gehörte unter anderem die Bestimmung der zu verwendenden Art der Quantifizierung von Aktivitäten. Wären dafür die auftretenden Aktivitäten abzuzählen und in einer Summe zu vereinigen, oder müsste deren Dauer jeweils gemessen werden, was dann aber auch zu Entscheidungen darüber nötigte, ob Äquivalenzregelungen statthaft sein könnten. Vielleicht wäre eine Wochenstunde Radfahren durch zwei Stunden Aktivität bei Computerspielen oder vier Stunden als Zuschauer bei Sportsendungen aufzurechnen. Möglicherweise dürfte aber das Anschauen von Fernsehfilmen nicht als Aktivität sondern vielmehr als Passivität gesehen werden. Auch der Hinweis, man möge sich doch der diagnostischen Interviews bedienen, brächte keine Problemlösungen, sondern lediglich Verschiebungen der persönlichen Verantwortlichkeit. Nun hätte der 14jährige Junge oder die 17jährige junge Frau zu entscheiden, was die Interviewer und Interviewerinnen nicht mehr entscheiden müssen. Der Hinweis auf die Notwendigkeit der Erstellung von Diagnosen ist auch keine Problemlösung, sondern lediglich die Delegation der Bestimmungsaufgabe an einen Diagnostiker oder an dessen Kollegin.

Der Verdacht, die Aufgabe einer empirisch verlässlichen Erfassung des Bestehens depressiver Störungen bei Kindern und Jugendlichen sei bislang keineswegs zufriedenstellend gelöst, darf geäußert werden. Das wiederum hebt nicht die Nötigung auf, für die intendierte populationsbeschreibende Untersuchung den Gegenstand operational zu bestimmen, nach dem mit der Datenaufnahme gesucht werden soll.

Eine weitere Annäherung an eine für den Ansatz der empirischen Untersuchung nicht ersetzbare operationale Bestimmung des Untersuchungsgegenstandes sollte am ehesten durch die eingehendere Befassung mit dem Störungsbild der Depression erreichbar sein.

Ein solches Störungsbild umfasst mehr als die Klassifikation seiner diagnostisch einsetzbaren Inhalte, wird aber in wesentlichen Aussagen durch das ihm unterliegende Klassifikationssystem repräsentiert.

Eine Klassifikation wird notwendig, wenn insgesamt mehr als eine relevante Klasse von Objekten auftritt. Das ist sowohl für Krankheiten als auch für Störungen der Fall. Klassifikatorische Unterscheidungen ordnen die Varianten auftretender Depressionen, indem sie Merkmale oder Ausprägungsgrade von Merkmalen gruppieren, benennen und inhaltlich gegeneinander abgrenzen. Zum Entwurf klassifikatorischer Systeme werden folglich Unterscheidungsmerkmale gebraucht, die entscheiden helfen in welche der klassifikatorischen Gruppierungen ein bekanntwerdendes Objekt einzuordnen ist. Die Hierarchie der Klassen von Krankheiten oder Störungen schafft ein Ordnungssystem, das als diagnostisches Leitsystem Verwendung findet. Mit der diagnostischen Arbeit wird, wenn Störungen vermutet werden, die Störungsklasse aufzufinden und nachzuweisen versucht, die im aktuellen Fall das Befinden eines Klienten stört. Ein diagnostisches Leitsystem als Übersicht potentiell auftretender Klassen von Störungen ist für die Durchführung der diagnostischen Arbeit eine unerlässliche, aber noch keine hinreichende Bedingung.

Wenn diagnostisch nach Klassen von Störungen oder Krankheiten differenziert werden soll, ist vorab die Frage zu beantworten, woran die Objekte der einen oder anderen Klasse voneinander unterschieden werden können. In den medizinischen Sprachgebrauch eingegangen ist die Verwendung des Begriffes Symptom. Im Altgriechischen wurden mit diesem Wort sowohl der Zufall als auch eine vorübergehende Eigentümlichkeit bezeichnet. Für Störungen wie für Erkrankungen gilt, dass sie als vorübergehend aufgefasst werden und an eigentümlichen Merkmalen zu erkennen sind. Dann sind auch ihre Eigenheiten vorübergehende Auffälligkeiten, eben Symptome. Indem sie nicht mehr nachweisbar wären, dürfte das als Indiz dafür gewertet werden, dass auch die jeweilige Störung oder Krankheit, deren Auftreten sich mit diesen Eigenheiten repräsentierte, bereits abwesend ist und sich damit als „vorübergehend“ erwiesen hat und überwunden ist.

Die hier so beharrlich, weil von erheblicher Bedeutung, erneut aufzugreifende Problematik ist die der Wahrnehmbarkeit der diagnostisch verwendeten Merkmale.

Wenn dafür Kriterien eingesetzt werden sollten, lassen sich unterschiedliche Begriffe verwenden. Eine kommunikable Form der Bestimmung von Qualitätsmerkmalen ist, dass die im testtheoretischen Verständnis objektiv, reliabel und valide sein sollten.

Obschon diese Bestimmung methodisch vertretbar sein sollte, führte sie bei ihrer Anwendung, beispielsweise in der diagnostischen Praxis, zu erheblichen Problemen. Die Angaben von Patientinnen oder Patienten sind im testtheoretischen Sinne sehr wahrscheinlich nur selten objektiv, reliabel und valide.

Auf dieses Dilemma ist aus methodisch relevanten Gründen hinzuweisen. Um Diagnostik dennoch möglich zu machen, ist der Diagnostiker, will er helfen, auf die Interpretation der mitgeteilten Angaben angewiesen. Die von Diagnostikern dabei erreichten unterschiedlichen

Erfolgsquoten führten zu der Begrifflichkeit von der diagnostischen Kunst, der Diagnose als Kunst.

Aber auch alle die, die in empirischen Studien die in den untersuchten Stichproben auftretenden Probanden und Probandinnen fragen, ob ihnen so oder so sei, haben mit eben diesen Problemen umzugehen, der potentiellen diagnostischen Unzuverlässigkeit der in Untersuchungen aufgenommenen Merkmale.

Einen wirksamen Schutz bieten eindeutig bestimmte und erfassbare Merkmale. „Ein Knochenbruch liegt vor, wenn im Röntgenbild dies oder jenes sichtbar ist.“ Einen solchen Stand der Bestimmung relevanter Merkmale hat die Diagnostik emotionaler Störungen bislang nicht und auch nicht mit diagnostischen Interviews erreicht und wohl auch nicht erreichen können. Und somit entsteht das Problem der Bestimmung des Objekts einer Untersuchung zur Auftretenshäufigkeit depressiver Störungen bei griechischen Kindern und muss, obschon auf oberbegrifflicher Ebene das Objekt unmissverständlich benannt ist, auch auf der Ebene der zu erfassenden Merkmale zu bestimmen versucht werden, um die Studie durchführen zu können.

Die Problematik der operanten Bestimmung von Erkennungsmerkmalen wird in der Beschäftigung mit den aktuellen Diagnosesystemen wieder gefunden und dort an vorhersagbarer Stelle. Die Bestimmung von Untersuchungsobjekten ist auf oberbegrifflichere Ebene relativ eindeutig zu vollziehen. Die Bestimmungsprobleme treten dort auf, wo auf die Frage zu antworten ist, an welchen Merkmalen die jeweils genannten Symptome erkennbar sein sollen.

Nicht ersetzbar ist jeweils die Antwort auf die Frage, woran das Auftreten oder die Abwesenheit eines Symptoms möglichst zweifelsfrei erkennbar wird. Wenn „niedergedrückte Stimmung“ ein Symptom sein sollte, muss der Empiriker auch bei der Durchführung einer populationsbeschreibenden Studie erklären, woran diese Stimmungslage erkannt werden kann und soll.

Dem Zwang zur Bestimmung erfassbarer Kriterien als Suchobjekte, das, was hier operante Bestimmung genannt wird, ist auch durch hierarchische Kategorisierung oder Klassifizierung nicht zu entkommen. Andererseits müssen die Suchkriterien als untere Ebene des diagnostischen Klassifikationssystems homogen mit dessen anderen Ebenen sein.

Die nachstehende Tabelle gibt einen Überblick über die in DSM-IV und ICD-10 zur Kennzeichnung der für Untereinheiten von Depressionen bei Kindern und Jugendlichen verwendeten Klassen.

Tabelle 3.1: Klassifikation depressiver Störungen nach DSM-IV und ICD-10

DSM-IV	ICD-10
<u>Affektive Störungen</u>	<u>Affektive Störungen</u>
<u>Depressive Störungen:</u>	F31 Bipolare affektive Störungen
296.xx Major Depression	F32 depressive Episoden
300.4 Dysthyme Störung	F33 Rezidivierende depressive Störungen
311 Nicht näher bezeichnete Depressive Störung	F34 Anhaltende affektive Störung
	F34.0 Zylothymia
	F34.1 Dysthymia
<u>Bipolare Störungen:</u>	F38 Sonstige affektive Störungen
296.xx Bipolar I Störung	F38.10 Rezidivierende kurze depressive Störung
296.89 Bipolar II Störung	F39 Nicht näher bezeichnete affektive Störung
301.13 Zylothyme Störung	
296.80 Nicht näher bezeichnete Bipolare Störung	
	<u>Neurotische, Belastungs- und somatoforme Störungen</u>
	F41.2 Angst und depressive Störungen gemischt
<u>Anpassungsstörungen:</u>	F43.2 Anpassungsstörungen
309.0 Mit depressiver Verstimmung	F43.20 Kurze depressive Reaktion
309.28 Mit Angst und Depressiver Verstimmung	F43.21 Längere depressive Reaktion
309.4 Mit Emotionalen Störungen und Störungen des Sozialverhaltens	F43.22 Angst und depressive Reaktion, gemischt
	<u>Verhaltens- und emotionale Störungen mit Beginn in der Kindheit und Jugend</u>
	F92.0 Störung des Sozialverhaltens mit depressiver Störung

Unter Kenntnisnahme der im DSM-IV und im ICD-10 aufgeführten Klassen unterschiedlicher affektiver Störungen werden im Hinblick auf die notwendigerweise zu vollziehenden weiteren Arbeitsschritte, die nachstehenden Folgerungen gezogen:

1. die unvermittelte Nutzung der in DSM-IV und ICD-10 benutzten Klassen affektiver Störungen zur Bestimmung von Merkmalen, die in der anzusetzenden empirischen Studie Verwendung finden können, führte nicht zu methodisch vertretbaren Ergebnissen und wäre, zumindest im Hinblick auf DSM-IV, auch nicht kompatibel mit dem Aufbau und den Intentionen des Systems.
2. Eine die eingeführten Klassen nutzende Operationalisierung wäre aber zu vollziehen, wenn sie wie folgt formuliert wird: „Untersuchungsgegenstand der nach Indizien oder Indikatoren für das Auftreten affektiver Störungen bei Kindern und Jugendlichen forschenden Untersuchung sind die zum Erhebungszeitpunkt bei allen Diagnostikern in Griechenland jeweils bekannten eigenen Diagnosen, wenn sie nach ICD-10 erarbeitet worden sind.“
3. Eine Variante dieser Operationalisierung wird durch das Einsetzen von DSM-IV an Stelle von ICD-10 gebildet.
4. Diese Operationalisierungen werden in den hier nachfolgend zu führen Erörterungen als „Operationalisierung über Diagnosen“ bezeichnet.

Wenn in Fortführung der Suche nach weiteren unter Kenntnisnahme von DSM-IV und ICD-10 realisierbaren Operationalisierungen die Symptomlisten aus beiden Handbüchern in die Überlegungen einbezogen werden, können dann auch die sich dann ergebenden Möglichkeiten bedacht werden. Diese Listen wären in den für jede einzelne Störungsklasse geltenden Varianten

zu beachten. Auf diese Weise ergäbe sich eine für DSM-IV und eine für ICD-10 repräsentative Gesamtliste. Wenn die jeweiligen Gesamtmengen an Symptomen operationalisiert werden sollten, entstünden mehrere Schwierigkeiten:

1. jede der beiden Gesamtmengen führte zu einer Vielzahl von Operationalisierungen, die in ihrer Gesamtzahl innerhalb einer empirischen Untersuchung kaum handhabbar sein dürften,
2. sollten die Operationalisierungen für jedes Symptom in einer Gesamtliste eingetragen werden, stünde eine Vielzahl identischer Operationalisierungen nebeneinander, die dann ihre diagnostische Funktion verloren hätten
3. würden demgegenüber die diagnostisch relevanten Symptomgruppen erhalten bleiben, entspräche dies dem jeweils in einer Einzeluntersuchung zu realisierenden diagnostischen Prozess.
4. Dann aber bietet sich im Hinblick auf eine Beachtung der Symptomlisten in DSM-IV und ICD-10 an, so wie schon zuvor, die Operationalisierung über Durchführung der Diagnosen nach DSM-IV und ICD-10 zu realisieren.

Auf der Suche nach einer empirisch nützlichen Operationalisierung des Untersuchungsgegenstandes für eine nach Hinweisen über das Auftreten von depressiven Störungen bei Kindern und Jugendlichen forschenden Untersuchung bot sich unter Rückgriff auf die im DSM-IV und ICD-10 genannten Klassen und Symptomlisten die Setzung von Diagnosen als geeignetes Untersuchungsobjekt an.

Das zuvor mehrfach genannte und erörterte Problem der operanten Bestimmung ist mit der Einsetzung von bekannt werdenden Diagnosen als Untersuchungsobjekt nicht aufgehoben.

Die Schwierigkeiten, auf die in diesen Ausführungen bei der Suche nach einem nutzbaren und zugleich nützlichen Untersuchungsobjekt für eine nach Hinweisen auf das Bestehen depressiver Störungen fahndenden empirischen Studie, wiederholt und eindringlich verwiesen wird, könnten als akademisch überhöhte Problemkonstruktion angesehen werden. Der unterliegende Sachverhalt ist allerdings eher von technisch oder sachlicher Nüchternheit: für ein gesuchtes Objekt kann nicht erklärt werden, durch welche eindeutig wahrnehmbaren Merkmale es sich von ähnlichen anderen unterscheiden lässt. Damit ist die potentielle Fehlerhaftigkeit dennoch durchgeführter Sucharbeit erkennbar. Für die Suche von Speisepilzen wären bei vergleichbarer Ausgangslage und giftigen Doppelgängern tödliche Irrtümer vorhersagbar.

Wenn nun aber, wegen der Schwierigkeiten bei der Bestimmung eindeutiger Merkmale – hier als Problem der operanten Bestimmung bezeichnet – weder über Diagnosen noch über andere Formen des Rückgriffs auf die Symptomlisten der Klassifikationssysteme den zu erhebenden Qualitätsforderungen nicht in hinreichendem Maße genügt werden kann, sind aus diesem Sachverhalt eindeutige Folgerungen ableitbar. Sollte die Behauptung einer bislang nicht hinreichenden operanten Präzisierung von Suchmerkmalen als Untersuchungsobjekte empirischer Untersuchungen zum Auftreten depressiver Störungen nicht widerlegt werden, wäre daraus sowohl die potentielle Fehlerhaftigkeit aller Diagnosen von emotionalen Störungen als auch aller auf Symptomlisten zurückgreifenden empirischen Untersuchungen anzunehmen.

Eine dem entsprechende Aussage, die medizinische Diagnose der Depression sei zur Zeit nicht möglich, findet sich in dem von Laux (2002) herausgegebenem Buch über das zur Jahrtausendwende nutzbare Wissen über Depressionen.

Für die hier betriebene Suche nach möglichst eindeutiger Bestimmung von validen und wahrnehmbaren Merkmalen depressiver Störungen als Untersuchungsobjekte sollten die Ansprüche gesenkt werden. Wenn wahrnehmbaren eindeutigen Merkmale definiert werden könnten, würden auch keine empirisch eindeutigen Suchergebnisse zu erwarten sein.

Falls dieser Sachverhalt angehalten sollte, müsste versucht werden, Suchstrategien zu entwickeln, die dennoch zu Sucherfolgen führten. Ziel der Suche ist jeweils, alle von depressiven Störungen betroffenen Schulkindern zu erfassen. Wenn sich das nun als nicht realisierbar erweisen sollte, müssten Suchziele bestimmt werden, die sich dem eigentlich verfolgten Ziel der lückenlos korrekten Identifizierung der betroffenen Schulkindern mit akzeptablen Aufklärungsraten näherten. Akzeptabel wäre – zwangsläufig – immer der höchste erreichbare Prozentsatz.

Zur Begründung dieser Setzung werden zwei Aussagen getroffen, von denen angenommen wird, dass sie uneingeschränkt zutreffen. Mit der ersten Aussage wird unterstellt, bis zum Zeitpunkt an dem diese Feststellung getroffen wird, habe weltweit nicht eine Untersuchung realisiert werden können, die, nach DSM-IV oder ICD-10, empirisch korrekt die zweifelsfreie Validität der jeweils festgestellten Auftretenshäufigkeiten depressiver Störungen bei Schulkindern nachweisen können. Falls daraufhin auf Untersuchungen mit diagnostischen Interviews verwiesen werden sollte, wird dem hier entgegen gehalten, dass allein schon die Frage nach der bisherigen Dauer des Auftretens der Symptome, an Schulkindern und Angehörige ihres sozialen Umfeldes gerichtet, Zweifel an der Validität solcher Aussagen verlange. Die so erhobenen Daten sind nirgendwo empirisch überprüft worden. Wenn solche Einwände bei einem so wesentlichen Kriterium wie das der Dauer übergangen werden, ist das zu kritisieren. Falls die bisherige Dauer des Auftretens der Symptome nicht dem Störungsbild entspräche, läge die vermutete depressive Störung auch nicht wirklich vor. Die zweite Aussage unterstellt, weltweit sei die Praxis der empirischen Untersuchungen dem nach der ersten Aussage behaupteten Sachverhalt bis jetzt, weil nicht anders möglich, angepasst und werde mit Verfahren geführt, deren unabwiesbare Unschärfe in der Treffsicherheit ihrer Feststellungen zu der Einführung der Bezeichnung „Screening-Verfahren“ führte.

Wenn die methodischen Bedingungen für die Realisierung einer zweifelsfreien Identifizierung emotionaler Störungen bei Schulkindern so schwierig zu handhaben sein sollten, wie hier behauptet wird, dürften auch alle Bemühungen, eine solche Untersuchung durch die Auswahl zweifelsfrei wahrnehmbarer Merkmale für das Auftreten der als Indikatoren für das Bestehen depressiver Störungen bei Schulkindern heranziehbarer Merkmale letztlich vergeblich sein. Dann wäre die Suche nach Vorgehensweisen, die eine Optimierung des Suchergebnisses erwarten ließen, eine unter den gegebenen Umständen empfehlenswert zu bewertende Strategie.

Diese wie zuvor beschriebenen auftretenden und hier zu erörternden Probleme im validen Erkennen der in den einschlägigen Störungsbildern beschriebenen Symptome für emotionale Störungen sind in einer anderen Perspektive als zwangsläufig auftretend, als erwartbar und in ihrer kausalen Abhängigkeit nicht aufhebbar zu erkennen und zu beschreiben.

Die Systematisierung dieser Bedingungen hat zwei diagnostischer Perspektiven einzuführen. Die Suche nach diagnostischen Merkmalen setzt zumindest einen Menschen, ein Individuum, eine Persönlichkeit voraus, die sich diagnostisch betätigt. Sie kann prinzipiell in zwei Blickrichtungen suchen. Die zu suchenden Objekte befinden sich entweder außerhalb des eigenen Körpers oder innerhalb. In beiden Fällen sind die Wahrnehmungen immer Ereignisse, die der Tätigkeit des

eigenen Nervensystems beruhen. Aber die Gesamtmenge dieser aufnehmbaren Informationen unterscheidet sich in solche, die prinzipiell auch durch andere Personen wahrnehmbar wären und die anderen, die nur von dem wahrnehmenden Menschen als psychische Ereignisse fassbar sind. Die eine Kategorie wird gern als die der objektiven oder besser objektivierbaren Informationen bezeichnet, während die andere Kategorie die der subjektiven Informationen umfasst.

Bis vor etwa zwei Jahrhunderten hätte diese Unterscheidung der diagnostisch nutzbaren Informationen möglicherweise ausgereicht. Dann wären einige der Bedingungen für diagnostische Tätigkeit in allgemeiner Form etwa wie folgt zu beschreiben gewesen. Diagnosen sollen diagnostische Objekte erkennen. Das geschieht, indem die jeweils von ihrer Umgebung unterschieden werden. Dazu werden diagnostisch treffsichere Merkmale benötigt. Entsprechende Sachverhalte bestehen überall, wo, wie in der Diagnose, bestimmte Objekte erkannt werden müssen, so etwa auch in der Biologie, Zoologie, Physik, Chemie, der Technik, aber auch im alltäglichen Geschehen. Die zur Identifizierung von Objekten nutzbaren Merkmale lassen sich nach deren wahrnehmungsspezifischen Eigenheiten kategorisieren. Eine nutzbare Kategorie wäre die der, im Prinzip, wahrnehmbaren Merkmale. Denen stünden dann die nicht wahrnehmbaren Merkmale gegenüber. Diese Kategorien wären mit den zuvor genannten zu kreuzen.

Tabelle 3.2a: Vier für diagnostische Tätigkeit unterscheidbare Kategorien

	wahrnehmbar	nicht wahrnehmbar
Objektivierbar	a	b
Subjektiv	c	d

Die Buchstaben a, b, c und d stehen für vier potentiell erreichbare unterschiedliche Diagnosequalitäten. Die durch Feld a symbolisierten Bedingungen gestatten das Erreichen eindeutiger Diagnosen. In Feld b und d sind Bedingungen symbolisiert, die keine Diagnose gestatten. Das mit c bezeichnete Feld symbolisiert Bedingungen, die zu subjektiver Gewissheit führen könnten.

Das dargestellte Schema elementarer Diagnosebedingungen reicht selbst für Untersuchungen aus, in denen Eigenheiten diagnostiziert werden sollen, die für Menschen unsichtbare Merkmale einschließen. In der Verhaltensbiologie der ersten Hälfte des 20. Jahrhunderts wurden zahlreiche Untersuchungen darüber geführt, zu welchen Leistungen andere Lebewesen, beispielsweise Bienen, fähig wären. So konnte nachgewiesen werden, dass Bienen die Farbe Ultraviolett wahrnehmen können, wozu Menschen nicht fähig sind. Untersuchen konnten die es dennoch, indem sie wahrnehmbare Merkmale für das Auftreten von Ultraviolett definierten und eine Versuchsanordnung nutzten, in der das Aufsuchen eines vorhergesagten und durch die Farbe Ultraviolett markierten Ortes als Nachweis für die Wahrnehmung dieser Farbe gelten konnte.

Ein anderes, erweitertes Schema wird aber diesen und anderen Bedingungen besser gerecht. Die Entwicklung der Wissenschaften in den vergangenen dreihundert Jahren ist unter anderem auch dadurch bemerkenswert geworden, dass zum einen die Wahrnehmbarkeit von Objekten und Merkmalen durch technische Neuerungen, etwa durch Mikroskope und Elektrodenmikroskope, in

zuvor ungeahnter Weise erweitert werden konnte und fortgesetzt erweitert wird. Die sogenannten bildgebenden Verfahren erfassen auf technisch unterschiedliche Weise das im Prinzip unsichtbare Innere von Körpern und stellen Informationen zur Verfügung, die sichtbare Darstellungen der von ihnen aufgenommenen Informationen über das Innere von Objekten, so auch vom menschlichen Organismus geben.

Danach wären dann sichtbare Objekte und ihre Merkmale weiterhin eine Kategorie diagnostisch verwertbarer Informationen. Als neue zweite Kategorie wären alle durch technische Hilfsmittel wahrnehmbar werdenden Informationen hinzu. Dazu rechneten dann auch Informationen aus Radiosignalen oder anderen physikalischen Quellen. Die dritte Kategorie bliebe weiterhin die der weder direkt noch mittelbar wahrnehmbaren Objekte und Merkmale.

Tabelle 3.2b: Vier für diagnostische Tätigkeit unterscheidbare Kategorien

	wahrnehmbar		nicht wahrnehmbar
	direkt	mittelbar	
objektivierbar	A	A	B
subjektiv	C	A	D

Die Buchstaben a, b, c und d stehen für vier potentiell erreichbare unterschiedliche Diagnosequalitäten. Die durch Feld a symbolisierten Bedingungen gestatten das Erreichen eindeutiger Diagnosen. In Feld b und d sind Bedingungen symbolisiert, die keine Diagnose gestatten. Das mit c bezeichnete Feld symbolisiert Bedingungen, die zu subjektiver Gewissheit führen könnten.

Für die Felder, die mit a bezeichnet wurden, gilt, dass sie prinzipiell Bedingungen symbolisieren, unter denen valide Diagnosen oder valide Identifizierungen durch empirische Populationsbeschreibungen erreichbar sind, wenn die relevanten Merkmale gefunden und benutzt werden.

Für emotionale Störungen gilt, dass wegen der im Störungsbild enthaltenen emotionalen Symptome, über die gegenwärtig bestenfalls subjektive Gewissheit herstellbar ist, die nicht objektiviert werden kann, dass keine eindeutigen und validen Diagnosen möglich sind.

Das wäre nur möglich, wenn nach den Störungsbildern nicht mehr das zu objektivierende, nachweisbare Auftreten emotionaler Symptome verlangt wäre, sondern stattdessen bereits das Bekennen einer subjektiven Gewissheit oder Überzeugung vom Auftreten eines emotionalen Symptoms als hinreichend für dessen Einbeziehung in eine Diagnose wäre.

Eine dem entsprechende Operationalisierung wäre beispielsweise: „Gibt an, sicher zu sein, dass die niedergedrückte Stimmung im Verlauf der letzten 52 Wochen an vier Tagen jeder Woche länger als sechs Stunden andauerte.“

Empirisch sicher wäre jeweils, ob die so befragten Kinder oder Jugendlichen die entsprechende Frage bejahten oder verneinten. Auch wenn die Validität der Behauptungen angezweifelt werden sollte, wäre damit keineswegs darüber entschieden, welchen diagnostischen Wert eine solche Angabe hätte.

Augenscheinlich ist die Bestimmung einer Merkmalskombination, die in empirischen Untersuchungen zweifelsfrei zur treffsicheren Erfassung von Personen geeignet wäre, die depressiv gestört sind, mit erheblichen Schwierigkeiten verbunden.

Der damit erreichte Stand der Bemühungen, eine Merkmalskombination als Untersuchungsgegenstand für die anzusetzende Untersuchung zum Auftreten depressiver Störungen unter griechischen Schulkindern zu finden, wird mit den unmittelbar nachfolgenden Anmerkungen nicht überwunden, sondern lediglich veranschaulicht und ergänzt.

Die Suche nach bestimmten Objekten in einer jeweils definierten Stichprobe unterliegt von Fall zu Fall eventuell sehr unterschiedlichen Bedingungen, verläuft andererseits nach einem weitgehend übereinstimmenden Tätigkeitsschema. Wenn nach Auftretenshäufigkeiten für bestimmte Objekte geforscht werden soll, scheint die Frage nach dem dafür als geeignet auszuwählenden Untersuchungsobjekten auf hierarchisch oberbegrifflichen Ebenen allein schon über die allgemeine Zielsetzung beantwortet zu sein. Wer daran interessiert sein sollte, empirisch zuverlässig zu erfahren, welche Anzahl von Schiffen am 31. Mai 200x die Meeresstrasse von Gibraltar durchfährt, hätte mit seiner Untersuchungsstrategie zu sichern, dass alle die Meerenge passierenden Schiffe gezählt werden. Die zu bestimmenden Untersuchungsobjekte sind Schiffe und als solche prinzipiell durch objektivierbare Beobachtungen zu erfassen. Die Objekte genügten den Bedingungen, die in den weiter oben verwendeten Schemata durch das Feld a symbolisiert sein sollten. Die einsetzenden Nebel und in der Nacht würden die Schiffe für die Beobachter aller Wahrscheinlichkeit nach nicht mehr zweifelsfrei durch Augenschein und dies auch nicht unter Zuhilfenahme von Nachtsichtgläsern erkennbar bleiben. Die für diesen Fall einsetzbaren Radargeräte brächten aber wahrnehmbare Symbolisierungen der Objekte auf die Radarschirme. Noch immer entsprächen die Umstände den in den weiter oben als Feld a symbolisierten Bedingungen. Wenn nun allerdings versucht werden sollte, die emotionale Befindlichkeit des jeweiligen Kapitäns beim Passieren der Meeresenge zu erforschen, versagten alle zur objektivierenden Erfassung der Merkmale emotionaler Befindlichkeiten einsetzbaren Verfahren. Nicht die emotionale Befindlichkeit selbst, sondern lediglich die Auskunft des Befragten oder die Auskünfte seiner Umgebung über die mutmaßliche emotionale Befindlichkeit wären zu erlangen. Das gesuchte Objekt ist, wie in den oben verwendeten Schemata das Feld c symbolisiert, nicht zu erfassen, bestenfalls nur Auskünfte über das Objekt der emotionalen Befindlichkeit.

Ein viel geringerer verbaler Aufwand zur Strukturierung der bei der Diagnose und anderen Erkennensleistungen vorliegenden Bedingungen wird notwendig, wenn das jeweils behandelte Störungsbild als hypothetisches Konstrukt aufgefasst wird. Die sind nach übereinstimmendem Verständnis niemals selbst zu beobachten. Von ihrem Auftreten wird ausgegangen, wenn alle die Korrelate nachweisbar auftreten, über die das Konstrukt bestimmt wird. Das wäre für das Störungsbild >major depression< die Liste der sie kennzeichnenden Symptome. Dabei tritt eine Verdopplung des Konstruktstatus insofern ein, als in der Liste der mit dem Konstrukt korrelierenden Symptome erneut hypothetische Konstrukte auftreten. Die >niedergedrückte Stimmungslage<, oder die dafür in Entsprechung verwendeten Worte, repräsentieren selbst hypothetische Konstrukte. Auch die werden über die ihnen eigenen Korrelate nachgewiesen. Eine korrekter Nachweis für das Auftreten des Störungsbildes wäre demnach über eine zumindest über zwei Ebenen gehen die Hierarchie von Symptomlisten – vielleicht auch über die Konstruktion eines entsprechend verzweigten Hoffmanschen Entscheidungsbaum - möglich und weiterhin auch erforderlich.

Aus den dargestellten und unter Berücksichtigung unterschiedlicher Perspektiven geführten Erörterungen von Möglichkeiten zur angemessenen Bestimmung einer als Untersuchungsobjekt einsetzbaren Merkmalskombination für die Suche nach griechischen Schulkindern, die von depressiven Störungen betroffen sind, ergeben sich die der gesuchten ergeben sich die nachstehenden Sachverhalte:

1. die gesuchte optimierte Merkmalskombination wäre durch eine auf DSM-IV oder ICD-10, oder beide, bezogene Liste der für das benutzte Störungsbild geltenden wahrnehmbaren Merkmale seiner Korrelate
2. Die Anwendung dieses formal korrekten Ansatzes für jede auf die Identifizierung depressiver Störungen gerichtete Tätigkeit stößt auf methodische Probleme.
3. Die Liste der das hypothetische Konstrukt des Störungsbildes definierenden Korrelate enthält emotionale Symptome die bis jetzt nicht objektiviert erfassbar sind, sondern nur durch subjektive Bekundungen über ihre Existenz bekannt gemacht werden.
4. Die große Zahl der bei dem Versuch der Identifizierung depressiver Störungen abzuarbeitenden Merkmale belastet, zumindest noch gegenwärtig, die Durchführung in mehrfacher und dabei auch in ökonomischer Hinsicht.

Die genannten Schwierigkeiten mindern, solange sie nicht überwunden werden können, das mit Identifizierungsversuchen erreichbare Maß an Validität gegenüber dem angestrebten Maximum.

Für die Bestimmung eines durch wahrnehmbare Merkmalskombinationen anwendbar zu machenden Untersuchungsobjektes, das bei der Identifizierung depressiver Störungen in einer Stichprobe griechischer Schulkinder optimal einsetzbar wäre, geben die mitgeteilte Sachverhalte die Beschränkungen an, unter denen die Entwicklung des Untersuchungsansatzes zu versuchen ist.

Allem Anschein nach ist die durch den Einsatz von in ihrer Validität nicht kontrollierten Diagnosen, von diagnostischen Interviews und Screening-Instrumenten gekennzeichnete Praxis eine Entsprechung der methodisch herrschenden Beschränkungen. Dieser Sachverhalt lässt zwei Perspektiven öffnen. In der einen erscheinen Versuche, die genannten Beschränkungen zu durchbrechen oder aufzuheben angeraten zu sein. In der anderen Perspektive wäre unter Kenntnisnahme der herrschenden Beschränkungen zu versuchen, die Ausprägung dieser Beschränkungen mit jedem neuen Ansatz zur Identifizierung depressiver Störungen nach Möglichkeit zu minimieren.

Das ist dann auch für den Ansatz der hier geplanten Studie zu versuchen. Die Bestimmung des Untersuchungsgegenstandes hätte demnach auch zu versuchen, mit einem operant bestimmten Untersuchungsgegenstand eine ökonomische Durchführung der empirischen Untersuchung bei möglichst hoher Validität möglich werden zu lassen.

Da in den längst schon international betriebenen Studien in unterschiedlichen Sprachen über ein als identisch aufgefasstes Objekt geforscht wird, eben den verschiedenen depressiven Störungsbildern, sollte auch der Frage nachgegangen werden, ob die von den Störungsbildern gemeinten Konstrukte selbst oder aber auch ihre sie definierenden Korrelate sprachunabhängig sind. Das empirisch zu klären hat einige Bedeutung für die immer wieder aufzunehmende Suche

nach dem, was sie symbolisieren. Dies jedenfalls so lange, wie die Kommunikation über sie zwischen potentiell betroffenen Menschen und jenen, die nach depressiven Störungen suchen im wesentlichen sprachlich determiniert ist.

3.3 Diagnostik affektiver Störungen

Die für die Bestimmung des Untersuchungsobjektes, der dann nachfolgend in ihrem Ansatz zu beschreibenden Untersuchung mit griechischen Schulkindern, als handlungsleitend zu nutzende theoretische Position ist in dem vorangehenden Abschnitt dieser Erörterungen dargestellt worden und soll so beibehalten werden. Die in diesen Ausführungen als aktuell gegeben betrachteten Bedingungen für alle mit dem Aufsuchen von depressiven Störungen in Stichproben oder Populationen befassten Tätigkeiten umfassen auch die Diagnosen depressiver Zustände, denn sie folgen auch der genannten Zielsetzung. Insofern bedürfte es an dieser Stelle keiner weiteren Ausführungen zur Diagnose depressiver Störungen. Wenn dazu nun doch Anmerkungen gegeben werden, geschieht dies aus zwei Gründen. Zum einen ist in den bisherigen Darstellungen die Diagnostik depressiver Störungen als eigene Tätigkeit nur randständig behandelt worden. Zum anderen erscheint eben dies möglicherweise als Verstoß gegen die Wahrnehmung solcher Diagnosen als eigentliches Kriterium für den Nachweis des Bestehens depressiver Störungen.

In der hier zu vertretenden Sichtweise ist die hierarchisch obere Kategorie durch den Terminus >Tätigkeiten mit dem Ziel des Erkennens depressiver Störungen< besetzt.

Die Diagnostik affektiver Störungen wird hier nachfolgend betrachtet, um entscheiden zu können, ob das Objekt der auf die Feststellung von Prävalenzen affektiver Störungen zu richtenden empirischen Untersuchung die Diagnose solcher Störungen sein soll.

Auszugehen ist zum einen von der sachlogischen Beziehung in der Zielsetzung Prävalenzen für ein bestimmtes Objekt empirisch verlässlich festzustellen. Wenn das Ziel die Feststellung von Prävalenzen für ein spezielles Objekt ist, sollte das Objekt der Untersuchung eben dieses Objekt, nämlich die affektiven Störungen bei Kindern und Jugendlichen sein. Die Beachtung dieser sachlogischen Beziehung wird durch einen weiteren Sachverhalt verhindert. Das gemeinte Objekt ist nicht unmittelbar erkennbar. Der valideste Nachweis seiner Existenz wird über Diagnosen erbracht. Daraus ergibt sich, das für das Ziel, höchst präzise Prävalenzraten zu erfassen, keine Realisierungsmöglichkeit besteht. Wenn die Erstellung von Diagnosen, also von Berichten über das Auftreten affektiver Störungen, die Arbeitsweise ist, mit der ein maximaler Grad an Validität erreicht werden kann, vermag eine auf die Erfassung von Prävalenzen gerichtete Untersuchung auch keine höhere Validität als die entsprechenden Diagnosen zu erreichen.

Um das abschätzen zu können, wird im Folgenden auf die Diagnosen affektiver Störungen einzugehen sein.

Über diagnostische Arbeit wird versucht, in bestehenden oder vermuteten Störungen oder Krankheiten das Auftreten spezieller Störungsbilder zu entdecken und nachzuweisen. Damit ist Diagnostik abhängig von den in Störungsbildern oder Krankheitsbildern genannten und beschriebenen Störungen oder Krankheiten. Für diese Ebene der Befassung mit Störungen oder Krankheiten wird der Terminus nosologische Ebene verwendet. Wie Störungen oder Krankheiten entdeckt, bestimmt und beschrieben werden, bedarf einer eigenen wissenschaftlichen Analyse und Darstellung. Immerhin wird über den Begriff der nosologischen Ebene erkennbar, dass die dort

aufzufindenden nosologischen Einheiten als im weitesten Sinne intellektuelle Konstruktionen zu verstehen sind, deren Realitätsbezug durch wissenschaftliche Arbeit zu sichern versucht wird. Wenn wissenschaftlich alles das sein kann, was Wissen schafft, reicht der Terminus vom wissenschaftlichen Arbeiten auch zur Repräsentation solcher historischer Zeiten aus, in denen nach gegenwärtigen Maßstäben noch vorwissenschaftlich gearbeitet wurde. Für den Bereich der affektiven Störungen könnte also davon ausgegangen werden, dass bereits Hippokrates und andere griechische Ärzte und Philosophen, so dann später auch der zuvor zitierte Plutarch, die Melancholie nosologisch eingeführt und beschrieben haben. Die nosologischen Einheiten gehen auf die im weiteren oder engeren Sinne wissenschaftliche Tätigkeit von Menschen zurück, Welchen Realitätsbezug die einzelne und also jeweils zu betrachtende nosologische Einheit hat, ist nicht unmittelbar von deren Existenz auf der nosologischen Ebene abzuleiten, sondern nur von den sie betreffenden empirischen Nachweisen.

Da diagnostische Arbeit auf nosologische Einheiten bezogen ist, hängt die wissenschaftliche Qualität diagnostischer Arbeit primär von der Qualität der Darstellungen auf nosologischer Ebene ab und erst sekundär von weiteren Bedingungen. Aber auch die müssen betrachtet werden, wenn die Möglichkeiten und Leistungen diagnostischer Dienste eingeschätzt werden sollen. Das gilt so auch für die diagnostische Arbeit zu affektiven Störungen.

Für die Bestimmung affektiver Störungen als nosologische Einheiten ist zu erkennen, dass die dabei einzubeziehenden psychischen und physiologischen Phänomene und dann auch deren Bewertung und Einteilung als normal oder pathologisch in unmittelbarer und zyklischer Beziehung zu den verfügbaren Möglichkeiten für ihre Erfassung stehen kann. Die jeweiligen unterschiedlichen theoretischen Annahmen über die Natur psychopathologischer Zustände führen zu unterschiedlichen Erfassungsmethoden. Der systematische Vergleich ihrer Ergebnisse kann dann wiederum einer Überprüfung der ihnen selbst unterliegenden hypothetischen Konstrukte dienen. Störungs- und Krankheitsbilder sind in dem Masse in ihrer Stabilität bedroht, wie empirische Nachweise für ihre Existenz fehlen. Graham (1980) und auch andere Autoren (siehe Laux, 2002) sprechen davon, dass der Diagnostik affektiver Störungen so viele Grundlagen fehlten, dass eine medizinisch korrekte diagnostische Arbeit nicht möglich sei.

Verhulst und Koot betonen in einer Erörterung der Bedingungen für die diagnostische Identifizierung psychischer Störungen die Relativität der benutzten Basisannahmen: „In psychiatry there has been much argument about how to define a case. Although there are a number of different and competing views on what constitutes a case, few would deny that the concept of psychiatric illness is nothing but a man-made abstraction, justified by its convenience and liable at any time to be adjusted. Illness may be conceptualized in a number of ways, such as a normal response to stress, as the statistically determined deviation from the normal, as the result of abnormality in physical functioning or just in the basis of “therapeutic concern”, that is whether individuals themselves (and, in the case of children, their parents or teachers) or the doctor are convinced that they need clinical attention..... The value of different approaches to the definition of caseness should be tested empirically” (Verhulst and Koot, 1992, S. 36).

Die Diagnostik depressiver Störungen im Kindes- und Jugendalter ist zum Ende des 20. Jahrhunderts ein eigener Forschungs- und Wissensbereich geworden. Das spiegelt sowohl das unmittelbare als auch das mittelbare wissenschaftliche Interesse an den dort behandelten Problemen wider. Unmittelbar interessiert sind die mit Diagnostik befassten Spezialisten.

Mittelbares Interesse besteht, weil affektive Störungen bei Kindern und Jugendlichen zunehmend als Risikofaktoren für die Gesundheit von Kindern und Jugendlichen verstanden werden.

Indem innerhalb der Diagnostik als Forschungs- und Wissensbereich die Entwicklung objektiver, reliabler und valider Methoden als eine Hauptaufgabe gesehen wird, setzen sich die aus der diagnostischen Praxis kommenden Impulse vermehrt durch. Sie selbst gehen vermutlich auf vermehrte Anforderungen von diagnostischer Arbeit zurück. Mit dem wachsenden Bekanntheitsgrad der affektiven Störungen bei Kindern und Jugendlichen in der Öffentlichkeit wird auch die Einbeziehung dieses Störungsbildes in gesundheitspolitische Maßnahmen angeregt, was zwangsläufig zur Nachfrage nach diagnostischen Leistungen führt. Der Diagnostik als Forschungs- und Wissensbereich drohte wissenschaftlicher Schaden, wenn sie in Wahrnehmung der Entwicklung sekundärer Bedingungen für diagnostisches Arbeiten die Kontrolle ihrer primären Abhängigkeiten von den nosologischen Einheiten des Bereiches affektiver Störungen vernachlässigte.

Erneut zu verweisen ist darauf, dass die hohe Akzeptanz, die DSM-IV und ICD-10 inzwischen erreichten, die Publikation kritischer Einlassungen gegenüber den dort vertretenen nosologischen Positionen nicht gerade fördern wird. Beispielsweise nannte Steinhausen in der Ausgabe seines Lehrbuches über psychische Störungen bei Kindern und Jugendlichen (1993) mehrere Einwände gegen die mit den neuen Versionen der Manuale verbreiteten Setzungen. In einer Folgeausgabe von 1996 fehlen diese Einwände bereits. Vielleicht ist Steinhausens frühere Kritik durch inzwischen bekannt gewordene Forschungsergebnisse gegenstandslos geworden, Vielleicht wurde sie aber auch aus anderen Gründen zurückgezogen. Eine Reduktion der kritischen Erörterung der aktuell vertretenen nosologischen Positionen und eine damit einhergehende Zurücknahme darauf bezogener Forschung dürfte nicht im wissenschaftlichen Interesse liegen. Indem der Erklärungswert der nosologischen Setzungen aus DSM-IV und ICD-10 als unbezweifelbar angesehen werden sollte, bestünde die Gefahr, dass alle auf der Basis von DSM-IV und ICD-10 und über sie validierten Messinstrumente mit ihren Ergebnissen immer wieder neu und fortlaufend den Erklärungswert ihrer eigenen nosologischen Basis bekräftigten. Als exemplarische kritische Anmerkung zur gegenwärtig vertretenen Gleichsetzung von Symptomen affektiver Störungen von Erwachsenen, Kindern und Jugendlichen wird gefragt, ob denn die aus der Entwicklung von Kindern resultierenden alterstypischen kognitiven Divergenzen zum kognitiven Status von Erwachsenen bereits hinreichend berücksichtigt worden sind. Symptomatologische Divergenzen könnten unter anderem auch wegen der offensichtlich starken Unterschiede im Zeiterleben und insbesondere in der Wahrnehmung von Vergangenheit, Gegenwart und Zukunft gesehen werden. Der „Zeitraum“, in dem Kinder sich orten, ist ein anderer als der von Erwachsenen in den unterschiedlicher Altersgruppen. Für die Existenz nosologisch bemerkenswerter Unterschieden in der Symptomatologie der affektiven Störungen bei Erwachsenen einerseits und Kindern und Jugendlichen andererseits sprechen auch empirische Befunde. Kovacs et al. (1994) berichteten über Fälle von Dysthymie im Kindesalter und gaben als Hauptsymptome traurige Stimmung, Reizbarkeit und Ärger an. Die anderen als typisch für depressive Störungen aufgefassten Symptome fehlten fast völlig. Akiskal (1990) schlug eine Unterteilung der Dysthymie in 4 Subtypen vor (1. chronisch-sekundäre Dysthymie; 2. unvollständig remittierte, residuale major depression; 3. subaffektive Dysthymie; 4. Charakterspektrumsstörung). Als wesentlich für die Evaluation der nosologischen Einheiten, der Kategorien und der Symptomlisten ist deren empirische Absicherung zu sehen. Im Widerspruch

dazu werden Diskussionen über nosologische Probleme nicht selten nur theoretisch oder unter verallgemeinerndem Hinweis auf klinische Erfahrungen begründet. Das sollte als nicht hinreichend und immer als Anlass zu empirischen Nachforschungen gesehen werden. Definitionen können nicht richtig oder falsch sein, sondern lediglich unterschiedlich. Ihre wissenschaftliche Qualität erreichen sie über den nachgewiesenen wissenschaftlichen Nutzen. Der kann unterschiedlich definiert sein, etwa als Steigerung des jeweiligen Erklärungswertes oder auch als Zunahme von Interventionserfolgen. Der jeweils zu erbringende empirische Nachweis der Annahmen ist nicht ersetzbar.

Die hier geführte Erörterung tritt dafür ein, speziell innerhalb des Wissenschaftsbereiches der Diagnostik affektiver Störungen bei Kindern und Jugendlichen, der wissenschaftlichen Beachtung der Entwicklungen auf der nosologischen Ebene deutlichen Vorrang einzuräumen. Das nützt allen Teilbereichen der Diagnostik.

Die Effizienz der diagnostischen Tätigkeit sichert dann aber auch die Grundlagen aller präventiven und therapeutischen Maßnahmen gegen depressive Störungen, die Bewertung solcher Interventionen und die Beschäftigung mit zahlreichen theoretischen Forschungsfragen. Zu den letzteren gehört auch die aus der epidemiologischen Forschung gestellte Frage nach der Häufigkeit des Auftretens depressiver Symptome, Syndrome und Störungen bei bestimmten Populationen.

Insoweit als dann nosologische Einheiten aktuell als nutzbares Leitsystem der diagnostischen Arbeit akzeptiert sein sollten, tritt die Methodik zwangsläufig in den Vordergrund der Beachtung. Gelegentlich scheint das Verständnis von Methodik auf die Betrachtung der Messinstrumente reduziert zu werden. Demgegenüber wäre einzuwenden, dass Methodik im weiteren Sinne die gesamte Tätigkeit des Diagnostikers umfasst. Zunächst unterscheidet sich die Tätigkeit des Diagnostizierens als Prozess in der psychologischen Perspektive nicht von anderen Tätigkeiten. Mit W. Hacker könnte der Prozessbeginn einer solchen Tätigkeit mit der Übernahme der Aufgabe und deren persönlicher Interpretation durch den Diagnostiker gesehen werden. In diesem Sinne ist die aktuell geleistete diagnostische Arbeit von der Lehrmeinung über Diagnostik zu unterscheiden. Die, etwa in Lehrbüchern dargestellte Diagnostik, ist das, was von damit befassten Spezialisten über den Prozess des Diagnostizierens in Bezug auf die jeweils aufgenommene Zielsetzung, gelehrt wird. Die im Land geleistete diagnostische Arbeit unterscheidet sich davon ebenso, wie etwa das Verhalten von Autofahrern von der geltenden Verkehrsordnung oder die Unterrichtspraxis von Lehrenden in öffentlichen Schulen von Lehrbüchern der Didaktik. Das letztgenannte Beispiel wäre eine besonders offene Relation, weil pädagogisch wie schulrechtlich den Lehrkräften in Deutschland ein gewisses Maß an pädagogischer oder didaktischer Freiheit zugestanden ist. Für Autofahrer in Deutschland dürfte solche persönliche Freiheit gegenüber der Verkehrsordnung nur in extremen Notfällen zugestanden werden. Der Diagnostiker hat sich an den aktuellen Stand der Wissenschaft zu halten, die er anwendet. Für die hier weiterzubringenden Erörterungen ist die Unterscheidung von Diagnostik als Lehre und diagnostischer Tätigkeit als persönliche Leistung eines Diagnostikers wesentlich. Der Diagnostiker kann und wird in vielen Fällen auch zugleich jemand sein, der allein oder zusammen mit anderen die jeweils geführte Intervention, zumeist die Behandlung, plant, durchführt und evaluiert.

Für die hier weiter zu führende Erörterung ist dann auch noch eine zweite Feststellung wesentlich. Die für Kinder und Jugendliche zu Diagnose affektiver Störungen entwickelte Diagnostik ist historisch der Entwicklung einer entsprechenden Diagnostik für Erwachsenen nachgeordnet. Das gilt nicht nur im Hinblick auf die Verfahrensweisen und Messinstrumente, sondern vor allem auch in Bezug auf die alltägliche diagnostische Praxis. Die Diagnostik affektiver Störungen für Erwachsene hat gegenüber der Aufnahme einer alltäglichen diagnostischen Praxis außerhalb der Forschungsstätten und Universitätskliniken mindestens einen Vorlauf von einem Jahrhundert. Als Indikatoren für den gemeinten Vorlauf könnte die in Arztpraxen vorherrschende Überzeugung von der Möglichkeit des Auftretens einer Depression bei Erwachsenen einerseits und bei Kindern und Jugendlichen andererseits gelten. Ein solcher Wahrscheinlichkeitswert ist durchaus ein für die Tätigkeit des Diagnostizierens wesentliches Kriterium. Der diagnostizierende Arzt kann aus unterschiedlichen Gründen bei einem Patienten nicht alle nur irgendwie denkbare Diagnosen durchführen. Die Tätigkeit des Diagnostizierens wird von Annahmen über die Wahrscheinlichkeit des Auftretens beispielsweise von affektiven Störungen geleitet. Falls ein Arzt vermuten müsste, das ein spezielles Krankheitsbild so selten auftritt, dass in einem aktuellen Fall mit dem Auftreten dieser Erkrankung nicht wirklich zu rechnen sei, wird die diagnostische Tätigkeit eher durch andere Verdachtsmomente gerichtet werden und eben nicht durch eine eher abwegige Vermutung. In diesem Sinne ist der Verdacht, ein Erwachsener könnte an affektiven Störungen leiden, seit gut einem Jahrhundert durchaus eine diagnostische Option. Eben dies lässt sich von der Vermutung eines Auftretens affektiver Störungen bei Kindern und Jugendlichen nicht behaupten.

Aber auch eine dritte Feststellung ist für die hier weiter zu führende Erörterung noch wesentlich. Da die Entwicklung von Psychiatrie und Psychologie in Griechenland der in den sogenannten westlichen Industriestaaten zeitlich verzögert folgt, ist begründet zu behaupten, dass die eben hier zuvor getroffenen zwei Feststellungen in Griechenland verstärkt zutreffen. Dies gilt insbesondere im Hinblick auf den diagnostischen Standard in der griechischen Provinz. Diese Feststellung gewinnt verstärkte Bedeutung durch einen sozio-kulturellen Unterschied zwischen solchen Ländern, etwa wie USA, Frankreich, Großbritannien und Deutschland einerseits und Griechenland andererseits. Auf den ist in der entsprechenden Fachliteratur bereits mehrfach verwiesen worden. Traditionell sind griechische Familien daran gewöhnt, in allen Fällen gesundheitlicher Belastungen, insbesondere aber bei psychischen Problemen zunächst nach Hilfe im eigenen Familienverband zu suchen. In ländlichen Bezirken war dies für lange Zeitabschnitte in der bisherigen Entwicklung die einzig realistische Hoffnung auf Hilfe, weil Ärzte oft relativ weit entfernt und durch Gebirge oder See von den Hilfesuchenden getrennt praktizierten. Erst in jüngster Zeit und vor allen Dingen in den großen Städten änderten sich die tradierten Verhaltensweisen.

Vor dem Hintergrund dieser drei Feststellungen sind nun die folgenden Sachverhalte zur Kenntnis zu nehmen.

1. Bei Laux (2002) ist ein für Deutschland geltender und empirisch gestützter Hinweis auf die für Erwachsene geltenden Prävalenzraten affektiver Störungen zu finden. Danach ergab sich aus drei zwischen 1995 und 1998 durchgeführten epidemiologischen Studien (n=5131) erfüllten 16,4% lebenszeitbezogen die diagnostischen Kriterien für eine „major depression“ oder eine Dysthymie. (ebd. S. 1) Die Einjahresquerschnittsprävalenz lag nach Wittchen et al. (1999) bei knapp 15%. Eine europäische Untersuchung (Lepine et al., 1997) fand für schwere depressive Störungen eine 6-Monatsprävalenz von etwa 7%. Diese

und andere Untersuchungen belegen die Verbreitung depressiver Störungen unter Erwachsenen. Laux (ebd., S. 1 – 6) beziffert und belegt den durch das Auftreten depressiver Störungen entstehenden hohen volkswirtschaftlichen Schaden.

2. Für das jeweils bestehende Gesundheitssystem ist daraufhin zu wünschen, dass es den auftretenden depressiven Störungen als erstes eine entsprechend hohe diagnostische Effizienz entgegen zu setzen weiß. Ganz im Gegensatz dazu stehen die Untersuchungsergebnisse einer internationalen WHO-Studie (Ustün und Sartorius, 1995). Diese Studie fand unter den Patienten der ärztlichen Primärversorgung einen Prozentsatz von 10% wegen depressiver Störungen behandlungsbedürftiger Patienten. Weitere 6,5% aus der Gesamtheit aller Patienten litten an klinisch bedeutsamen Ausprägungen von major depression oder Dystymie. Aber nur 54% der nach ICD-10 bestehenden depressiven Störungen wurden von den Primärärzten erkannt. Dieses „diagnostische Defizit“ (Ustün und Sartorius) ist in seinen Entstehungsbedingungen möglicherweise nachvollziehbar, sein Ausmaß bleibt dennoch besorgniserregend. Weitere Untersuchungsergebnisse verweisen darauf, dass der dargestellte Sachverhalt augenscheinlich der gängigen Praxis entspricht. Nur jeder zweite Patient mit manifester Depression wird von dem aufgesuchten Allgemeinarzt als depressiv erkannt (Linden et al., 1996). Der mit jener Studie nachgewiesene Trend in den durch diagnostische Tätigkeit erreichten Erkennensleistungen wird mit einer Studie von Wittchen et al. (2001) erneut aufgefunden und somit bestätigt. Von etwa 15000 Patienten in Allgemeinarztpraxen litten fast 11% an Depressionen und gut die Hälfte von ihnen zum wiederholten Mal. Die Hausärzte erkannten die jeweils vorliegende depressive Belastung nur in der Hälfte dieser Fälle. In dem vor der Studie liegenden Jahr hatten die depressiven Patienten den Hausarzt fast doppelt so häufig aufgesucht wie die nicht-depressiven Patienten.
3. Aus den dargestellten Findungen ist zu ersehen, dass die diagnostische Tätigkeit der sogenannten Primärärzte, die in der Regel nicht als Spezialisten für psychische Störungen einzuordnen sind, das Auftreten depressive Störungen bei den eigenen Patienten in großem Ausmaß übersieht.
4. Wenn aber das „diagnostische Defizit“ (Ustün und Sartorius) so überaus groß ist, dass etwa jeder zweite Patient mit depressiven Störungen nicht als solcher erkannt wird, ist die diagnostische Tätigkeit im Hinblick auf das Erkennen von Depressionen als überaus unzuverlässig anzusehen.
5. Für Griechenland ist ein Auftreten von „diagnostischen Defiziten“ (Ustün und Sartorius) wegen der Besonderheiten der historischen Entwicklung und in Bezug auf den Landesdurchschnitt als sehr wahrscheinlich in seinem Ausmaß noch höher als hier zuvor dargestellt zu erwarten.
6. Für die diagnostische Tätigkeit in den Allgemeinpraxen muss wegen der historischen Entwicklungsverzögerung für die Einstellung zum Auftreten von depressiven Störungen bei Kindern und Jugendlichen mit mindestens gleich hohen „diagnostischen Defiziten“ gerechnet werden wie bei Erwachsenen. Sehr wahrscheinlich liegt dieses spezielle diagnostische Defizit aber eher noch höher als das bei Erwachsenen. Für die Abschätzung „diagnostischer Defizite“ im Erkennen von depressiven Störungen bei Kindern und Jugendlichen in Griechenland können keine belegbaren Gründe angeführt werden, nach denen das Defizit dort kleiner sein könnte. Das Gegenteil ist der Fall. Für Griechenland sollte von höheren „diagnostischen Defiziten“ ausgegangen werden.

7. Das dargestellte Ausmaß der empirisch belegten „diagnostischen Defizite“ (Ustün und Sartorius) spricht nicht für eine Nutzung allgemeinärztlicher Diagnosen zum Auftreten von Depressionen bei Kindern und Jugendlichen innerhalb des Ansatzes einer epidemiologischen Studie zur depressiven Belastung griechischer Schulkinder. Die Verwendung allgemeinärztlicher Diagnosen als Operationalisierung für das Auftreten von depressiven Störungen bei Kindern und Jugendlichen dürfte zu einer nicht hinnehmbaren Divergenz zwischen den in Griechenland auftretenden depressiven Störungen und deren diagnostischen Erkennen führen. Das zu erwartende hohe „diagnostische Defizit“ „diagnostische Defizit“ (Ustün und Sartorius) würde ein nicht erwünschtes hohes epidemiologisches Defizit nach sich ziehen. Ein solches epidemiologisches Defizit verhinderte die Annäherung an das Untersuchungsziel, der Vermittlung von zutreffenden Informationen über die Verbreitung depressiver Störungen bei Kindern und Jugendlichen in Griechenland.

Da die Möglichkeit einer Verwendung allgemeinärztlicher Diagnosen über Depressionen bei Schulkindern als operante Bestimmung für Depressionen zum gegebenen Zeitpunkt nicht mehr in Betracht zu ziehen ist, erübrigt sich in dem hier zu erörternden Zusammenhang auch das weitere Eingehen auf Diagnosen der Depression bei Schulkindern. Stattdessen wird nun im Folgenden der Frage nachgegangen, wie denn die bisher international durchgeführten epidemiologischen Untersuchungen zur depressiven Belastung von Kindern und Jugendlichen die Bestimmung ihres Untersuchungsgegenstandes vollzogen haben.

3.4 Gegenstand der auf die Erfassung von Prävalenzen für affektive Störungen bei Kindern und Jugendlichen gerichteten Untersuchungen

Wenn Untersuchungen das Auftreten von affektiven Störungen bei Kindern und Jugendlichen erfassen sollen, muss ihr Untersuchungsgegenstand operational bestimmt werden. Auf der Suche nach einer für die Erfassung der zu einem gegebenen Zeitpunkt erfassbaren Häufigkeitsrate für affektive Störungen geeigneten Objektbestimmung war die Möglichkeit, „das Vorliegen einer entsprechenden allgemeinärztlichen Diagnose“ als Operationalisierung des Untersuchungsgegenstandes einzusetzen, wegen eines mutmaßlich zu erwartenden hohen „diagnostischen Defizits (Ustün und Sartorius, 1995) verworfen worden. Da bereits zahlreiche epidemiologischen Untersuchungen mit der gleichen Zielsetzung, wenn auch nicht in Griechenland, durchgeführt worden sind, liegt nahe, daraufhin zu prüfen, welche Objektbestimmungen in diesen Untersuchungen verwendet wurden. Allein schon das Argument einer anzustrebenden Vergleichbarkeit für Studien mit übereinstimmenden Zielsetzungen spräche für eine Übernahme bereits eingesetzter Operationalisierungen für depressive Störungen bei Kindern und Jugendlichen.

Wenn die operante Bestimmung des Untersuchungsobjektes über allgemein ärztliche Diagnosen verworfen wird, ist für alle weiteren Bestimmungsversuche vorherzusehen, dass ungeachtet aller etwa bestehenden Schwierigkeiten versucht werden muss, eine tolerierbare Annäherung an die Gewissheit über das Vorliegen einer depressiven Störung zu erreichen. Unterstellt wird, dass die exakt durchgeführte Diagnose, dies als Tätigkeit verstanden, die valideste Form der Erarbeitung einer Aussage über das aktuell gegebene Bestehen oder Nichtbestehen einer affektiven Störung

sei. Ohne weiteres Eingehen auf etwa vertretene andere Auffassungen, wird der Prozess des diagnostischen Arbeitens als eine der vielen möglichen Anlässe zum empirischen Forschen verstanden. Diagnostische Tätigkeiten sind explizit empirisch orientierte Nachforschungen zum fraglichen Bestehen eines Sachverhaltes, wie etwa im Falle eines Verdachtes über das Bestehen einer affektiven Störung, die mit dem Ziel betrieben werden, Gewissheit zu erlangen. Als erwiesen oder auch als gewiss gilt das, was empirisch korrekt nachgewiesen werden konnte. Ob diagnostische Tätigkeit im aktuellen Fall dieser Setzung entspricht, ihr nahe kommt oder deutlich von ihr abweicht, ist eine nachgeordnete Problematik. Der Anspruch muss vertreten werden, dass über diagnostische Tätigkeiten empirisch korrekte Nachweise des zu diagnostizierenden Sachverhaltes erarbeitet werden sollen. Wenn also ein diagnostisches Manual bestimmt, dass ein Symptom der Dysthymia bei Kindern schon ein Jahr bestanden haben müsste, zielt der explizite empirisch orientierte Anspruch auf einen empirisch vertretbaren Nachweis für die entsprechende Dauer des Auftretens dieses Symptoms. Falls nun in einem diagnostischen Interview der 15jährige Proband gefragt werden sollte, wie lange denn nach seiner Erinnerung das fragliche Symptom schon aufgetreten sei, ist die so gewonnene Information, falls keine weiteren Erkundungen hinzukommen sollten, kein empirisch zufrieden stellender Nachweis. Wer ihn dennoch als solchen akzeptierte, ignorierte alle in diesem Zusammenhang denkbaren Fehlerquellen. Zumindest verlangt werden sollte der Nachweis, das 15jährige Probanden emotionale Zustände für den Ablauf eines Jahres zutreffend erinnerten und vielleicht auch mit welchen Fehlerquoten. Auch solche Nachweise, die wohl nicht vorliegen, wären von Stichproben auf Populationen vorgenommene Verallgemeinerungen, die selbst wiederum mit Fehlern behaftet sein werden.

Wenn von diagnostischer Tätigkeit eine explizit empirisch orientierte Arbeitsweise mit der Begründung verlangt wird, nur so seien valide diagnostische Urteile erreichbar, wären alle die weiterhin auch noch in Betracht gezogenen anderen Möglichkeiten zur Bestimmung des Gegenstandes einer epidemiologischen Untersuchung per definitionem als im Vergleich zu empirisch gesicherten Diagnosen affektiver Störungen als weniger valide und folglich anzusehen sind. Wie dann auch immer der Nachweis des Bestehens einer affektiven Störung versucht werden sollte, er wäre wegen des Fehlens der empirisch verlässlichen Diagnose als unvollständig, als defizitär einzuordnen. Daraus folgert, dass solche defizitären Bestimmungen, sobald durchgeführt, letztlich alle der Bestätigung durch das Einbringen von validen Diagnosen bedürften. Epidemiologischen Untersuchungen müssten demnach als unvollständig angesehen werden, wenn sie nicht in jedem von ihnen mitgezählten Einzelfall auf eine empirisch verlässlichen Diagnosen verweisen könnten. Für sie stünde die Bestätigung ihrer epidemiologischen Validität aus. Da im speziellen Fall der Diagnose affektiver Störungen bei Kindern und Jugendlichen in vielen Regionen Europas sowohl die empirische Verlässlichkeit der aufzugreifenden allgemeinärztlichen Diagnosen als auch ihre Repräsentativität in Frage stehen, scheint unumgänglich zu sein, eine nach DSM-IV oder ICD-10 empirisch unvollständige Bestimmung des epidemiologischen Gegenstand hinzunehmen. Dann wären aber auch alle bekannt gewordenen oder bekannt werdenden Prävalenzraten für das Auftreten affektiver Störungen als mehr oder weniger verlässliche Schätzungen zu interpretieren.

Die dargestellten Beziehungen gelten, soweit die Annahme, empirisch orientierte Diagnosen seien valider als andere Verfahren, auch zutrifft. Sollten demgegenüber andere Verfahren als valider

anzusehen sein, wären die beschriebenen Folgerungen zurück zu ziehen. Dafür besteht bislang aber kein Anlass.

3.5 Bestimmungen des Gegenstandes einer epidemiologischen Untersuchung ohne unmittelbare Bezugnahme auf empirisch verlässliche Diagnosen

Solange die dargestellte Position als folgerichtig angesehen werden darf, ist die Bestimmung des Gegenstandes der für Griechenland geplanten epidemiologischen Untersuchung unter Beachtung der nachstehenden Sachverhalte vorzunehmen:

1. Die Bestimmung des Gegenstandes der beabsichtigten epidemiologischen Untersuchung unter Rückgriff auf empirisch verlässliche allgemeinärztliche Diagnosen führt zu vorhersehbaren erheblichen Ungenauigkeiten der Informationsnahme über das Auftreten affektiver Störungen bei Schulkindern in Griechenland.
 - 1.1 Zum einen steht die empirische Verlässlichkeit allgemeinärztlicher Diagnosen affektiver Störungen bei Schulkindern in Frage.
 - 1.2 Zum anderen bestehen erhebliche Zweifel an der Repräsentativität der aktuell vorhandenen Diagnosen.
2. Die unerlässliche operationale Bestimmung des Gegenstandes der anzusetzenden epidemiologischen Studie ist demnach auf die Einsetzung einer gegenüber DSM-VI und ICD-10 empirisch defizitären Untersuchungsobjektes zurück geworfen.
 - 2.1 Empirisch defizitär wird der Gegenstand einer epidemiologischen Untersuchung hier genannt, wenn durch seine operationale Bestimmung die vom geltenden Störungsbild für den Nachweis des Bestehens der Störung geforderten Informationen nicht hinreichend umfänglich und empirisch verlässlich sind.
 - 2.2 Defizitäre Bestimmungen werden epidemiologisch nutzbar, wenn prospektiv verwertbare Beziehungen zwischen dem Auftreten ihres Gegenstandes und dem diagnostizierten Auftreten der Störung bekannt geworden sind.
3. Da epidemiologische Untersuchungen allgemein Informationen über das Auftreten ihres Gegenstandes sammeln, bleiben defizitäre Bestimmungen zum Auftreten affektiver Störungen bei Schulkindern auf die Nutzung bestimmter Informationsquellen angewiesen. Das sind:
 - 3.1 vom Kind abgenommene (interne)
 - 3.1.1 somatische oder
 - 3.1.2 psychische Informationen,
 - 3.2 über das Kind aus seiner
 - 3.2.1 engeren oder
 - 3.2.2 weiteren Umgebung aufgenommene Informationen.

In den nachfolgenden Tabellen finden sich Übersichten über empirische Untersuchungen, die entweder ausschließlich oder unter anderem auf das Erfassen von Prävalenzraten zum Auftreten affektiver Störungen bei Kindern und Jugendlichen gerichtet waren. Mehrere Tabellen sind aufgeführt, weil in der Regel jede nur einen ausschnitthaften, eben exemplarischen Überblick vermittelt, der, den jeweiligen Zielsetzungen folgend, von der auswählenden Persönlichkeit und deren Arbeitsperspektive abhängig ist. Indem unterschiedliche Tabellen verschiedener Herkunft wiedergegeben werden, finden sich auch unterschiedliche Perspektiven beachtet.

Übereinstimmend ist für alle der funktionale Zusammenhang innerhalb der hier geführten Erörterung.

Diese Tabellen enthalten mit den in ihnen aufgeführten Untersuchungen einerseits allgemeine Hinweise auf epidemiologische Arbeiten zur Erfassung von Prävalenzraten für das Auftreten depressiver Störungen bei Kindern und Jugendlichen. Damit bieten sie aber auch, was hier gesucht worden ist, Beispiele für die operationale Bestimmung des Untersuchungsgegenstandes in solchen epidemiologischen Untersuchungen. Mit einer solchen Bestimmung wird für jede epidemiologische Untersuchung eine erste empirische Qualität determiniert, die der Erfassung des Objekts der Untersuchung.

Die aufgeführten Tabellen sind dementsprechend vor allem unter dem Aspekt der Bestimmung des Gegenstandes der epidemiologischen Untersuchung zu betrachten.

Die Angaben in der nachstehenden Tabelle 3.3 verweisen auf benutzte Klinische Interviews. Nicht erkennbar ist, ob etwa weitere Messinstrumente Verwendung fanden. Auszugehen ist jedoch davon, dass mit der Verwendung eines Klinischen Interviews auch eine für die Untersuchung insgesamt leitende Objektbestimmung gegeben wird. Zwar tritt mit jedem weiteren und nachgeordneten anderen Instrument auch wieder ein weiteres Objekt, eben das der neuen Messung auf, aber nun besteht auch Gelegenheit nach Relationen zwischen den aus unterschiedlichen Quellen kommenden Messergebnissen zu suchen.

Die Bestimmung des Objektes epidemiologischer Untersuchungen über an DSM oder ICD gebundene klinische Interviews bietet die Möglichkeit zur Realisierung einer weitreichenden Entsprechung der diagnostischen Maßnahmen zu der Beschreibung des Störungsbildes in den Manualen. Das Vorliegen einer Möglichkeit ist nicht identisch mit der Realisierung einer repräsentativen diagnostischen Abbildes der in den Manualen gebotenen Beschreibung einer speziell gemeinten Störung. Parameter der Qualität einer diagnostischen Repräsentation des Störungsbildes sind über die Konzepte von Objektivität, Reliabilität und Validität erfassbar.

Die mit Tabelle 3.3 wiedergegebene Auflistung epidemiologischer Untersuchungen zum Auftreten depressiver Störungen bei Kindern und Jugendlichen nach Groen und Petermann (2003) ergänzt von C. Kampsiou. Betrachtet werden Untersuchungen, die ihr Forschungsobjekt über klinische Interviews bestimmt haben. Implizit belegen die Listen eine spezielle Beachtung klinischer Interviews, die vor allem wohl über deren Bindung an die internationalen Handbücher und den damit speziell gebotenen diagnostischen Möglichkeiten erklärt werden darf. Im Übrigen gelten die zur Bestimmung des Objektes epidemiologischer Untersuchungen durch klinische Interviews zuvor nach Tab. 3.2 getroffenen Feststellungen auch weiterhin.

Tabelle 3.3: Auflistung epidemiologischer Untersuchungen zum Auftreten depressiver Störungen bei Kindern und Jugendlichen.

Tabelle in Anlehnung an Groen und Petermann, 2002, S. 38, ergänzt von C. Kampisiou

Autoren Jahr; Land	N	Alter (Jahre)	Diagnose Instrument / Kriterien	Auftretenshäufigkeiten (%)								
				Major Depression				Dysthyme Störung				
				LZ	12M	6M	P	LZ	12M	6M	P	
Kashani et al. (1983, New Zealand)	189	9 (T1)	K-SADS-E	1,8								
Anderson et al. (1987, New Zealand)	792	11 (T2)	DISC				0,5					1,8
Kashani et al. (1987, USA)	150	14-16	DICA				4,7					
Bird et al. (1988, USA)	386	4-16	DISC				5,9					
Fleming et al. (1989, Kanada)	2852	6-16	SDI				5,9					
Lewinsohn et al. (1991; 1993; 1994; USA)	1709	14-18	K-SADS-E	20,4			2,9					
Lewinsohn, Hops et al. (1993; USA)	1710	14-18	K-SADS-E / DSM-III-R	18,5			2,6	3,2				0,5
Cooper & Goodyer (1993; GB)	1068	11-16	DISC / DSM-III-R		6,0		3,6					
Reinherz et al. (1993; USA)	385	18	DIS-III-R/ DSM-III-R	9,4								
Goodyer & Cooper (1993, England)	1068(w)	11-16	DISC		6,0		3,6					
Polaino-Lorente & Domènech (1993, Spanien)	1093	8-11	CDRS-R				1,8					6,4
Feehan et al. (1994; Neuseeland)	930	18	DIS-III-R/ DSM-III-R		13,3		3,4		3,2			
Costello et al. (1996, USA)	1015	9, 11, 13	CAPA				0,03					0,1
Shaffer et al. (1996; USA u.a.)	1285	9-17	DISC / DSM-III-R			4,9						
Verhulst et al. (1997; Niederlande)	853	13-18	DISC / DSM-III-R			2,8				2,5		
Canals et al. (1997; Spanien)	290	18	SCAN / DSM-III-R				2,4					5,8
Garisson, et al. (1997, USA)	359	11-16	K-SADS				7,0 (Z1) 5,4 (Z2)					2,6 (Z1) 4,3 (Z2)
Essau et al. (1998; Deutschland)	1035	12-17	M-CIDI / DSM-IV	14,0				5,6				
Steinhausen et al. (1998; Schweiz)	1964	7-16	DISC-P / DSM-III-R			0,7				0,2		
Wittchen et al. (1998; Deutschland)	3021	14-24	M-CIDI / DSM-IV	11,8	5,3			3,0	2,9			
Olsson & von Knorring (1999; Schweden)	2300	16/17	DICA-R-A/ DSM-III-R	11,4	5,8				2,0			
Romano et al. (2001; Kanada)	1201	14-17	DISC / DSM-III-R			4,5				2,6		

Zeichenerklärung: N = Stichprobengröße; A = Alter in Jahren; LZ = auf Lebenszeit bezogene Prävalenz; 12M = auf 12 Monate bezogene Prävalenz; 6M = auf 6 Monate bezogene Prävalenz; P = Punktprävalenz / aktuell

In der anschließend von Essau (2002, S. 32) übernommenen Tabelle finden sich Interviewschemas aufgelistet, die für die Arbeit mit Kindern und Jugendlichen entwickelt worden sind.

Tabelle 3.4: Diagnostische Interviewschemas für die Anwendung bei Kindern und Jugendlichen (Essau, 2002, S. 32)

Diagnostische Interview-schemata	Diagnostisches System	Strukturierungsgrad	Informant	Zeitlicher Rahmen	Interviewer	Schulungsdauer
Diagnostic Interview Schedule for Children (DISC; Costello et al., 1985)	DSM-III, DSM-III-R	Hochstrukturiert	Kind, Eltern, Lehrer	6 Monate	Laien-interviewer, Kliniker	3 – 4 Wochen
Diagnostic Interview for Children and Adolescents (DICA-R; Herjanic/Reich, 1982)	DSM-III-R	Hochstrukturiert	Kind, Eltern	Interviewzeitpunkt, Lebenszeit	Laien-interviewer, Kliniker	1 Woche 3 Wochen
Kiddie-Schedule for Affective disorders and Schizophrenia (K-SADS; Puig-Antich / Chambers, 1978)	DSM-III, DSM-III-R	Halbstrukturiert	Kind, Eltern	Lebenszeit	Kliniker	1 – 2 Wochen
Interview Schedule for Children (ISC; Kovacs, 1985)	DSM-III,	Halbstrukturiert	Kind, Eltern	Interviewzeitpunkt,	Kliniker	„Sehr intensiv“
Child Assessment Schedule (CAS; Hodges, 1982)	DSM-III, DSM-III-R	Halbstrukturiert	Kind, Eltern	12 Monate	Laien-interviewer, Kliniker	5 – 10 Tage
Child and Adolescent Psychiatric Assessment (CAPA; Angold et al., 1995)	DSM-III-R ICD-10	Halbstrukturiert	Kind, Eltern	3 Monate	Laien-interviewer, Kliniker	1 Monat

Die nach Essau wiedergegebene exemplarische Übersicht über diagnostische Interviewschemas zur Erfassung depressiver Störungen bei Kindern und Jugendlichen verändert die bis hier betrachteten Gegebenheiten bei der Bestimmung eines Objekts der epidemiologischen Untersuchung durch diagnostische Interviews nicht, ergänzt aber, auch wenn vorwiegend auf DSM-III bezogen, die bisher beachteten Informationen. So ist aus der Übersicht auch zu entnehmen, dass die empirische Basis der Informationsnahme je nach dem verwendeten Instrument durch die Befragung von Eltern oder Lehrern erweitert wird. Diese Erweiterung bringt eine spezielle Entscheidungsproblematik in den Urteilsprozess. Da nicht gewusst wird, wer von den Informanten dem wirklich vorliegenden Sachverhalt mit seinen Darstellungen und inwieweit nahe kommt, ist zu entscheiden, welches Gewicht den Angaben der unterschiedlichen Personen beigemessen werden soll. Das wäre selbst bei übereinstimmenden Angaben zu bedenken. Sie könnten wechselseitig das Bestehen eines Sachverhaltes bekräftigen, werden aber nicht zwangsläufig durch Übereinstimmung realitätsnäher. Solange keine objektiven Außenkriterien verwendbar sind, ist immer wieder neu zu entscheiden, welchen Angaben welcher empirisch verlässliche Informationswert zukommt. Gegenüber Selbstberichten von Kindern und

Jugendlichen erwiesen sich Eltern und Lehrer in bisherigen Untersuchungen als relativ schwach informiert (Nevermann, 1993, Casanovas, 2003).

Da neben den diagnostischen Interviews weitere Messinstrumente und damit auch weitere Möglichkeiten der Bestimmung des Objekts für epidemiologische Untersuchungen zur Wahl stehen, wird die Erörterung auf sie ausgedehnt. Die nachstehende Tabelle bezieht exemplarisch einige deutschsprachige Verfahren zur Aufnahme von Selbstberichten über depressive Symptome von Kindern und Jugendlichen durch deren Auflistung in die Erörterung mit ein. Die Übersicht wird nach Essau (2002, S. 35) zitiert.

Tabelle 3.5: Beispiele von Selbstbeurteilungs-Fragebögen zur Erfassung von Depressionen bei Jugendlichen (Essau, 2002, S. 35)

Instrumente/ Autoren	Altersgruppe	Anzahl Items	Spezielle Merkmale
„Depression Self-Rating Scale“ (Birleson 1981)	7-14 Jahre	18	Items umfassen affektive, kognitive und behaviorale Symptome von Depression
„Center for Epidemiological Studies Depression Scale“ (Radloff 1991)	6-17 Jahre	20	Items beziehen sich auf Freunde und Eltern
„Reynolds Adolescent Depression Scale“ (Reynolds 1987)	12-18 Jahre	30	DSM-III Symptome für Major Depression und Dysthyme Störung
„Children's Depression Inventory“ (Kovacs 1992)	7-17 Jahre	27	Items umfassen affektive, kognitive und behaviorale Symptome von Depression
„Mood and Feelings Questionnaire“ (Angold et al. 1987)	8-18 Jahre	32	Deckt die Symptombereiche im DSM-III-R und andere Items wie Einsamkeit und das Gefühl, ungeliebt zu sein, ab.
„Depressionstest für Kinder“ (Rossmann 1993)	8-13 Jahre	55	3 Skalen: Dysphorie/Selbst-wert, agitiertes Verhalten, Müdigkeit/autonome Reaktionen

Wenn bei diagnostischen Interviews zur Aufnahme von Informationen über depressive Symptome bei Kindern und Jugendlichen, soweit sie inhaltlich und strukturell explizit an DSM oder ICD gebunden sind, eben diese Eigenheit als empirisch und diagnostisch bedeutungsvoll behandelt werden könnte, entfällt dieses Argument für die in Tabelle 3.4 aufgeführten Verfahren weitgehend. Für die Erfassung der jeweiligen Nützlichkeit von Interviews oder Fragebögen zur Selbstbewertung gewinnt der genannte Sachverhalt nur mittelbar Bedeutung. Die für alle Verfahren gleichermaßen in Frage stehende Validität bei der Erfassung depressiver Störungen ist für alle Messinstrumente zu prüfen und zu belegen. Der methodisch zu beachtende Unterschied in der impliziten Bestimmung des Objekts einer auf depressive Störungen gerichteten epidemiologischen Untersuchung durch das verwendete Messinstrument sollte nicht übergangen werden. Mit einem in seiner Konstruktion eng an den entsprechenden Manualen ausgerichteten diagnostischen Interview wird immerhin die symptomnahe Repräsentation des Störungsbildes durch das diagnostische Instrument versucht. Die Selbsteinschätzungen anregenden Messinstrumente wie DTK oder CDI verzichten weitgehend auf den Versuch einer symptomnahen Rekonstruktion eines depressiven Störungsbildes in ihrem Fragengeflecht. Sie unterstellen alle, dass die von ihnen verwendeten Fragen die Funktion von Indikatoren übernehmen können. Darin folgen sie der tradierten Praxis vieler psychologischer Fragebögen, die über relativ wenige Fragen zu einer möglichst starken Evidenz von Entscheidungen über das bestehende Ausmaß von Angst, Aggression oder anderen Phänomenen zu kommen versuchen. Die definierten Indikatoren sind die Objekte der danach durchgeführten Untersuchung. Mit dem CDI wird in diesem Sinne nicht nach sprachlichen Abbildungen eines depressiven Störungsbildes

gesucht, wie das diagnostische Interviews versuchen, sondern nach Indikatoren für das Bestehen oder Nichtbestehen depressiver Störungen. Dementsprechend zwingend ist für diese Verfahren dann auch der Nachweis, dass die ihrer Konstruktion unterliegenden Annahmen sich durch den nachgewiesenen Sucherfolg als nutzbringend verifizieren lassen. Für die diagnostische Leistungsfähigkeit des DTK von Rossmann (siehe 2. Neuauflage des Handbuches zum DTK) ist immerhin eine im empirischen Versuch an 66 Kindern nachgewiesene Sensibilität gegenüber diagnostizierten depressiven Störungen von etwa 81% Erkennensleistung zur Kenntnis zu bringen. Gegenüber dem hier an anderer Stelle schon erörterten „diagnostischen Defizit“ von 46% und der Erkennensleistung von 54% (Ustün und Sartorius, 1995) erreichte der DTK, wenn andere Einwände gegen einen solchen Vergleich übergangen werden, einen beachtenswerten diagnostischen Zuwachs. Empirisch steht immer wieder die Beantwortung der Frage an, welche Erkennensleistungen mit welchem Instrument nachweislich erreicht werden können. Eine epidemiologische Untersuchung gewinnt ihre Reputation in einem beträchtlichen Masse ebenso aus den für sie heran zu ziehenden Erkennensleistungen. Das wird im allgemeinen über die für die verwendeten Instrumente nachweisbarer Erkennensleistungen geschehen. Die Verlässlichkeit der Annahmen sollte sich dann, was auf deren Bedeutung verweist, aus entsprechenden Replikationen der Studien ergeben.

Da an dieser Stelle der Erörterungen vor allem anderen die unterschiedlichen Möglichkeiten zur Operationalisierung des Untersuchungsgegenstandes in epidemiologischen Untersuchungen zum Auftreten von depressiven Störungen bei Kindern und Jugendlichen zu erfassen und zu beschreiben sind, wird nun eine Übersicht herangezogen, in der C. Essau (2002, S. 43) darstellt, welche Methoden der Erfassung von depressiven Störungen gegenwärtig zu beachten sind.

Tabelle 3.6: Methoden zur Erfassung von Depression (nach Essau, 2002, S. 43)

Methoden	Vorteile	Nachteile
Rating-Skalen	<ul style="list-style-type: none"> • Schnell durchzuführen • Kostengünstig hinsichtlich des Einsatzes von Fachkräften und Ausbildungsanforderungen 	<ul style="list-style-type: none"> • Für die meisten Skalen gibt es keine Cutoff-Scores • Schwierigkeiten, unbeobachtete Symptome zu berichten • Geringe Übereinstimmung der unterschiedlichen Informanten
Selbstbeurteilungsfragebögen	<ul style="list-style-type: none"> • Gute psychometrische Eigenschaften • Leicht und kostengünstig durchzuführen • Liefert direkte Information des Kindes 	<ul style="list-style-type: none"> • Information beschränkt auf die Perspektive des Kindes
Direkte Beobachtung	<ul style="list-style-type: none"> • Erfassung des aktuellen Verhaltens • Erkennung von Umweltkontingenzen 	<ul style="list-style-type: none"> • Möglicherweise teuer • Einhaltung des Beobachtungsprotokolls manchmal problematisch • Erfordert eigenes Kodierungssystem
Diagnostisches Interview	<ul style="list-style-type: none"> • Untersuchung der Reaktionen des Kindes • Direkte Verhaltensbeobachtung • Reduziert Informationsvarianz insbesondere bei hochstrukturierten diagnostischen Interviewsschemata 	<ul style="list-style-type: none"> • Kann zeitaufwändig in der Durchführung sein • Halbstrukturierte Interviews sollten von Klinikern durchgeführt werden • Bei großer Stichprobe möglicherweise kostenintensiv • Möglicherweise Informationsverlust aufgrund von Screening-Items einiger diagnostischer Interviews

Die Übersicht führt Rating-Skalen, Selbstbeobachtungs-Fragebögen, direkte Beobachtung und diagnostische Interviews als gegenwärtig wählbare Verfahren zur Erfassung depressiver Störungen bei Kindern und Jugendlichen an. Mit jedem dieser Verfahren, falls es allein angewendet werden sollte, wird zwangsläufig eine spezifische Operationalisierung des Untersuchungsgegenstandes herbeigeführt. Rating-Skalen, Selbstbeobachtungs-Fragebögen und direkte Beobachtung haben als Gemeinsamkeit, dass sie jeweils nach Indizien für das Bestehen depressiver Störungen suchen lassen. Damit tritt dann ein, was zuvor erörtert worden ist, für diese Verfahren muss nachgewiesen werden, dass sie die in epidemiologischen Untersuchungen geforderten Erkennensleistungen erbringen. Dieser Nachweis wird in der Regel nicht in der epidemiologischen Untersuchung selbst zu erbringen sein, er müsste wohl mittelbar geführt werden, indem auf Erkennensleistungen verwiesen wird, die in Untersuchungen erbracht worden sind, die einen spezifischen Nachweis möglich machen. In solchen Untersuchungen müsste in der Regel bereits bekannt sein, welche depressiv belasteten und unbelasteten Kindern und Jugendlichen die Stichprobe bilden. Das zu bewertende Messverfahren hat in solchen Stichproben Gelegenheit, seine immer wieder neu in Frage stehende Vertrauenswürdigkeit unter Beweis zu stellen. Ebenso geeignet wäre ein Untersuchungsansatz, indem behauptete Erkennensleistungen nachträglich geprüft und verifiziert oder falsifiziert werden könnten.

Beachtenswert ist in diesem Zusammenhang, dass diagnostische Interviews, indem sie sich unmittelbar auf diagnostische Manuale beziehen, auf entsprechende methodische Vorzüge verweisen können. Allein damit ist aber keineswegs nachgewiesen, dass sie im Vergleich zu anderen Methoden die insgesamt besseren Messqualitäten bieten. Eine größere Zahl der um das Jahr 2000 durchgeführten epidemiologischer Untersuchungen nutzte, trotz umfänglicher und somit aufwendig zu bearbeitender Stichproben diagnostische Interviews. Das könnte als Anzeichen dafür gewertet werden, dass diesen Verfahren eben doch höhere Messqualitäten zuerkannt worden sind. Diese Untersuchungen nutzen die Möglichkeiten einer objektivierenden Bestimmung ihres Untersuchungsgegenstandes allerdings auch in unterschiedlicher Weise.

Ein die diagnostischen Möglichkeiten der in Tab. 5.5 aufgeführten Verfahrensweisen abwägender Vergleich berechtigt, sobald die potentielle empirische Verlässlichkeit als Kriterium eingeführt wird, zu einer dichotomen kategorialen Reduktion. Keine der gegenwärtig einsetzbaren diagnostischen Vorgehensweisen kann einer Kategorie maximaler diagnostischer Erkennensleistung zugeordnet werden, der eines unmittelbaren Nachweises einer depressiven Störung. Unmittelbar werden hier alle Nachweise genannt, durch die eine Störung selbst durch empirisch verlässliche Abhebung von einem ansonsten ungestörten Funktionszusammenhang identifiziert wird. Ein solches diagnostisches Verfahren ist gegenwärtig nicht bekannt. Alle einsetzbaren diagnostischen Methoden gehören in die Kategorie mittelbar identifizierender Diagnosen. Mittelbar werden sie hier genannt, weil nur vermittelnde Argumente deren Identifizierungserfolg mehr oder weniger wahrscheinlich werden lassen. Der kategoriale Unterschied ist der zwischen Gewissheit über die Validität der diagnostischen Aussage und gradueller Wahrscheinlichkeiten für das Vorliegen von Validität. Dann unterscheiden sich die einsetzbaren diagnostischen Vorgehensweisen und Mittel in dem Beitrag, den sie zur Steigerung der Wahrscheinlichkeit des Vorliegens valider Aussagen geben und geben können.

Diese Kategorie sammelt alle Verfahren, die hier zuvor als defizitär gegenüber den nach DSM-IV oder ICD-10 zu leistenden empirischen Nachweisen kategorisiert worden sind. Als defizitär waren hier zuvor alle Verfahrensweisen bezeichnet worden, die keine empirisch eindeutigen Nachweise für das Auftreten der entsprechenden Symptome bieten. Dem unterliegt ein explizit empirisch akzentuiertes Verständnis von Diagnose, wie es im klinischen Alltag nur bedingt vertreten werden kann. Die für Betroffene notwendige Hilfeleistung sollte und kann nicht aufgeschoben werden,

bis diagnostische Verfahren vorliegen, die puristischen empirischen Kriterien genügen werden. Die Qualität des Kliniklers erweist sich im Gegenteil daran, wie erfolgreich er trotz offensichtlich bestehender methodischer Behinderungen zu arbeiten vermag. Der Angabe eines Patienten, er habe sich den rechten Arm gebrochen, kann man in der ärztlichen Praxis seit dem 20. Jahrhundert so gut wie fast überall durch diagnostische Maßnahmen nachgehen, die solche Annahmen, je nach Situation, möglichst eindeutig zu verifizieren oder zu widerlegen vermögen. So könnte eine Röntgenaufnahme die angestrebte empirisch verlässliche Aufklärung verschaffen. Die zur Erfassung der Symptome affektiver Störungen eingesetzten Messinstrumente beschränken sich bis jetzt aber ausnahmslos auf die Aussagen der potentiell Betroffenen und kommen in Gefahr, deren Angaben als Nachweis für den behaupteten Sachverhalt zu nehmen. Aber jede dieser Angaben, die des potentiell Betroffenen wie auch die von Personen aus seinem Umfeld, sind lediglich Nachweise für sich selbst, für die gegebene Aussage. Möglicherweise ist die methodische Situation bei der Diagnose affektiver Störungen eine sehr spezielle empirische Problematik. Dem entspricht die aus einer etwas anderen Perspektive von Laux, (2002) gegebene Hinweis, dass eine eindeutige medizinische Diagnose der Depression bis heute nicht möglich sei. Die psychologische Diagnostik allerdings hat mit ihrer wissenschaftlichen Entwicklung im 19. und 20. Jahrhundert und auch danach Methoden geschaffen, die den Grad der Ungewissheit, der Selbstauskünften anhaftet, auf ein Maß zu reduzieren vermögen, das empirisch begründete Entscheidungen unter Einbeziehung des verbleibenden Risikos zulässt. Einer daran zu orientierenden diagnostischen Praxis steht möglicherweise der dann zwangsläufig zu leistende personelle und damit auch ökonomische Aufwand entgegen. Selbst in der alltäglichen medizinischen Praxis scheint auch gegenwärtig noch nicht unüblich zu sein, bei nicht lebensbedrohlich eingeschätzten Fällen von einer empirisch eindeutig gesicherte Diagnose abzusehen. Stattdessen wird geprüft, ob Besserung nicht auch über Vorgehensweisen erreicht werden kann, die weniger Aufwendungen als empirisch eindeutige Diagnosen erfordern. Diese Aussage ist allerdings lediglich eine auf Zufallsbeobachtungen beruhende und im Übrigen dann durchaus ungeprüfte Mutmaßung, eben eine Hypothese.

Die in der hier zuvor nach C. Essau wiedergegebenen Tabelle der derzeit verfügbaren diagnostischen Verfahren setzt einige der Bedingungen, unter denen die nicht übersehbare Operationalisierung des Objektes einer epidemiologischen Untersuchung zum Auftreten depressiver Störungen bei Kindern und Jugendlichen vollzogen wird. Eine solche Operationalisierung ist unumgänglich. Wer sie ablehnt, vielleicht weil der Begriff selbst eine gewisse Nähe zum klassischen Behaviorismus anzeigt, kann zwar den Begriff, nicht aber den Vorgang vermeiden. Für epidemiologische Untersuchungen wird, wie bei jeder Arbeit bestimmt, welche Tätigkeiten auszuführen sind.

Ein für die Einschätzung der epidemiologischen Qualität des an der Bestimmung des Untersuchungsobjektes mitwirkenden Messverfahrens nutzbares Kriterium ist dessen nachgewiesene Erkennensleistung. Das Instrument soll im Interesse der Bewältigung einer epidemiologischen Aufgabe helfen, die Untersuchung durchführen und Entscheidungen über das Auftreten der gesuchten Störung fällen zu können. Von Screening-Verfahren wird eine ökonomisch realisierbare Trennung der durch die jeweils gemeinte Störung unbelasteten Personen einer Stichprobe von den wahrscheinlich belasteten Personen erwartet. Screen als transitives Verb bezeichnet Prüfungen, die zum Ziel haben, die Anwesenheit oder Abwesenheit einer Störung oder allgemein einer Qualität zu erfassen. In diesem Sinne ist beispielsweise der ESDM von Domènech u.a. (1985) als ein spanisches Screening-Instrument zur Erfassung depressiver Störungen bei Schulkindern durch deren Lehrkräfte bezeichnet worden. Da Lehrkräfte im Allgemeinen nicht als Diagnostiker für affektive Störungen tätig werden, ist der ESDM offensichtlich von Domènech

und Mitarbeitern als ein Hilfsmittel zur Identifizierung solcher Schulkinder gedacht worden, für die das Erstellen einer spezifischen Diagnose als notwendig erachtet wird. Screening wäre in diesem Sinne eine Prüfung, deren Ergebnis jeweils einer spezifischen Nachprüfung unterzogen werden sollte. Domènech und Mitarbeiter (1985) schätzen dann aber weiterhin das hier als Beispiel ausgewählte Verfahren ESDM als potentiell diagnostisches Hilfsmittel ein. Danach wäre das Aufsteigen eines Screening-Verfahrens zum diagnostischen Hilfsinstrument durchaus möglich, wenn dessen diagnostische Nützlichkeit als erwiesen gelten darf. Von diagnostischen Interviews wird allerdings angenommen, dass sie die unmittelbare Ableitung von Diagnosen nach DSM-IV oder ICD-10 gestatten. Das erscheint zwangsläufig möglich, weil sie inhaltlich ausdrücklich auf diese Handbücher bezogen sind (siehe beispielsweise auch Essau, 2002, S. 30). Die Möglichkeit der Ableitung einer Diagnose ist für sich genommen aber nicht identisch mit einem Nachweis über deren diagnostische Qualität. Wenn schon Messinstrumente unabweisbar die Bestimmung des Objektes einer epidemiologischen Untersuchung wesentlich beeinflussen, dann sollten die Instrumente mit hohen diagnostischen Qualitäten sein. Das folgert unmittelbar aus den epidemiologischen Zielsetzungen.

Somit ist dann die nachgewiesene diagnostische Qualität eines Messinstrumentes ein für die Qualität der epidemiologischen Untersuchung bedeutsames Argument.

Für jedes einschlägige Verfahren steht somit dessen Validität, aber ebenso dessen Objektivität, Reliabilität wie aber weiterhin auch andere Verfahrensmerkmale epidemiologisch in Frage und bedürfen folgerichtig des speziellen Nachweises.

Eine kritische Würdigung der klinischen Interviews findet sich beispielsweise neben anderen bei Angold et al. (1996). So wird dort beispielsweise auch mitgeteilt, dass in einer untersuchten Stichprobe 31% der interviewten Jugendlichen in ihren Zeitangaben über die bisherige Dauer ihrer depressiven Verstimmung von Befragung zu Befragung unterschiedliche Angaben anboten. Die Autoren weisen damit explizit auf bestehende Reliabilitätsprobleme hin, die sich dann auch auf die erreichbare Validität auswirken und spezielle Validitätsprobleme schaffen. Wenn gegenüber den in ihren Zeitangaben über ihre Symptome instabilen 31% von Jugendlichen einer Stichprobe auf die 69% verwiesen wird, deren Angaben stabil blieben, ist damit nicht nachzuweisen, dass die Angaben der Mehrheit diagnostisch aussagekräftiger gewesen wären. Sie sind lediglich stabiler. Ebenso umgehen Hinweise auf strukturelle Schwächen einzelner Interviews die eigentliche Validitätsproblematik. Nachzuweisen bleibt letztlich, wie valide die jeweiligen Verfahren sind. Groen und Petermann (2002, S.36) gehen ausdrücklich auf die methodischen Qualitäten der bis 2002 vorliegenden epidemiologischen Untersuchungen zur Prävalenz affektiver Störungen ein. Sie relativieren die jeweils erreichte Aussagekraft durch die Auflistung mehrerer methodischer Probleme (Groen und Petermann, 2002, S.36).

„Allgemein sind bei der Bewertung und dem Vergleich der heute vorliegenden, teilweise doch recht unterschiedlichen Ergebnisse einzelner Studien die unterschiedlichen methodischen Vorgehensweisen zu berücksichtigen (vgl. Groen, Scheithauer, Essau & Petermann, 1997; Kessler et al., 2001; Roberts, Attkisson & Rosenblatt, 1998; Romano et al., 2001; Verhulst, 1995). So zeigen sich zum Beispiel methodische Inkonsistenzen bei

- der Verwendung von Erhebungsverfahren,
- der Berücksichtigung von diagnostischen Kriterien zur Fallbestimmung,
- der Größe und der Zusammenstellung der untersuchten Stichproben,
- den zugrundeliegenden Erfassungszeiträumen (z.B. Punkt-, Sechs-Monats-, Ein-Jahres- oder Lebenszeitprävalenz) und
- der Heranziehung unterschiedlicher Informationsquellen (z.B. Selbsteinschätzung der Kinder und Jugendlichen, Fremdurteile von Eltern und Lehrern).“

Die hier in diesem Kapitel geführte und nun zu einem vorläufigen Abschluss zu bringende Erörterung von Problemen der Bestimmung des Untersuchungsobjektes für epidemiologische Untersuchungen zum Auftreten depressiver Störungen bei Kindern und Jugendlichen legt nahe, die folgenden Feststellungen zu treffen:

1. Die Bestimmung des Untersuchungsobjektes einer epidemiologischen Untersuchung zum Auftreten depressiver Störungen bei Kindern und Jugendlichen ist eine unabweisbare Notwendigkeit.
2. Mit der operationalen Bestimmung des Untersuchungsobjektes wird festgelegt, wonach epidemiologisch gesucht wird.
3. Die operationale Bestimmung des Untersuchungsgegenstandes bestimmt die nach der Realisierung der Untersuchung gegebenen Aussagemöglichkeiten mit.
4. Das Erreichen einer zielgerechten Funktionalität der operationalen Bestimmung des Untersuchungsgegenstandes einer epidemiologischen Untersuchung wird durch mehrere Bedingungen begünstigt.
5. Eine unerlässliche Bedingung ist die Einsetzung eines zur Identifizierung des gesuchten Störungsbildes hinreichend valides Messinstrument.
6. Eine sehr wahrscheinlich nutzbringende Bedingung ergibt sich aus dem Einsatz mehrerer und aufeinander bezogener Messinstrumente.
7. Die verwendeten Messinstrumente bestimmen im Zusammenwirken mit dem Untersuchungsplan den Gegenstand der epidemiologischen Untersuchung.
8. Die optimale Auswahl der Messinstrumente und des Untersuchungsplanes entsteht durch deren Übereinstimmung mit der aktuell gegebenen Realisierbarkeit. Messinstrumente und Untersuchungsplan können ihre Qualitäten nur zur Geltung bringen, wenn sie mit den seitens der Durchführenden einsetzbaren personellen und materiellen Ressourcen realisierbar sind.
9. Messinstrumente und Untersuchungsplan sollten darüber hinaus den Bedingungen entsprechen, die für die Durchführung der Untersuchung den Lebensumständen der untersuchten Stichprobe entsprechen.

Wenn eingangs mit der Überschrift zu dem hier abzuschließenden Kapitel die Bestimmung des Gegenstandes epidemiologischer Untersuchungen zum Auftreten depressiver Störungen bei Kindern und Jugendlichen ohne empirisch verlässliche Diagnosen als P ein besonderes Problem erschien, so ist das hier zu korrigieren. Eine solche Bestimmung ist im Hinblick auf Untersuchungen zum Auftreten depressiver Störungen bei Kindern und Jugendlichen nicht das besondere, sondern das allgemeine und in der Forschungsarbeit wie auch in der klinischen Praxis recht alltägliche Problem.

3.6 Die Bestimmung des Untersuchungsgegenstandes für die durchzuführende epidemiologische Untersuchung mit griechischen Schulkindern

Die notwendige Bestimmung des Untersuchungsgegenstandes der an einer Stichprobe griechischer Schulkinder durchzuführenden epidemiologischen Untersuchung zum Auftreten depressiver Störungen folgt den mitgeteilten neun Prinzipien. Als ein spezielles Problem epidemiologischer Untersuchungen wurde die operationale Bestimmung des Untersuchungsgegenstandes aufgegriffen, weil in der gegenwärtigen Literatur zu den bisher durchgeführten Untersuchungen die Bestimmung des Gegenstandes nicht als Problem gesehen wird. Zwar wird (siehe auch Groen und Petermann, 2002) auf fehlende Vergleichbarkeit solcher Untersuchungen hingewiesen, die hat eine erste grundlegende Ursache darin, dass die

Operationalisierung des jeweiligen Untersuchungsgegenstandes variiert und dennoch von Untersuchungen zum Auftreten von Depressionen ausgegangen wird. Keineswegs notwendig ist, die Bestimmung des Untersuchungsgegenstands zu vereinheitlichen. Notwendig ist aber, im Vergleich der verschiedenen Untersuchungen die Unterschiedlichkeit der Operationalisierungen des Gegenstandes zu nennen und zu berücksichtigen. Diese Unterschiedlichkeit ist möglicherweise nicht nur als Defizit zu betrachten, möglicherweise gestattet sie auch Einsichten, die ohne diese Verschiedenartigkeit nicht erfassbar wären.

Soweit die operationale Bestimmung des Untersuchungsgegenstandes der hier zu betrachtenden Untersuchung die Auswahl der Messinstrumente betrifft, sind zunächst die Punkte 8 und 9 in der vorangegangenen Auflistung zu beachten. Da die zu ziehende Stichprobe über alle Landesteile Griechenlands verteilt sein soll, erscheint der Rückgriff auf allgemeinärztliche Diagnosen unangemessen. Erwartbar ist, dass eine Großzahl depressiv belasteter Schulkinder ärztlich unter diesem diagnostischen Aspekt bisher nicht bekannt geworden ist. Da weiterhin die einsetzbaren Ressourcen extrem gering sind, ist, auch wenn das in Betracht gezogen wäre, vom Einsatz diagnostischer Interviews abzusehen.

Darüber ergibt sich die Notwendigkeit, auf Fragebogen zurück zu greifen. Selbstbeurteilungen ermöglichende griechische Bogen standen zum Beginn der Planung nicht zur Auswahl. Die griechische Übersetzung eines in anderen Ländern schon erprobten Selbstbeurteilungsverfahrens bot sich an. Gewählt wurden der CDI von Kovacs (1985, 2004) und der DTK von Rossmann (1993, 2005). Beide sind erprobt und überzeugen in ihren Messqualitäten. Der CDI ist im Übrigen ein seit Jahren international erprobtes Verfahren. Daraus ergeben sich Vergleichsmöglichkeiten. Der DTK steht in guter Übereinstimmung mit dem CDI, ergänzt den aber noch durch zwei zusätzliche Skalen. Eine ausführlichere Darstellung der genannten Verfahren findet sich in den nachfolgenden verfahrensspezifischen Arbeitsteilen.

Durch die Auswahl von CDI und DTK als Messinstrumente wird die damit durchzuführende epidemiologische Untersuchung über Selbsteinschätzungen von Schulkindern spezifische Hinweise auf das Vorliegen unterschiedlicher Grade depressiver Belastungen sammeln. Das erlaubt zum einen die Unterscheidung der unterschiedlichen Grade von selbstdargestellten Belastungen sowohl über CDI als auch über DTK. Diese selbstangegebenen Belastungen sind inhaltsvalide für sich selbst. Dann gelten sie weiterhin als Indikatoren für das Auftreten unterschiedlicher Grade depressiver Störungen. Die Indikatorenfunktion ist in entsprechenden Validitätsstudien, zu denen sich in den nachfolgenden verfahrensspezifischen Arbeitsteilen weitere Angaben finden, hinreichend belegt. Die Sensitivität des DTK liegt bei 81%. (Rossmann, 2005). Ähnliche Werte finden sich in unterschiedlichen Untersuchungen für den CDI (Kovacs, 2004). Beide Verfahren erreichen keine fehlerfreie Trennung zwischen den mit klinischen Ausprägungen depressiver Störungen belasteten Schulkindern und anderen. Im Sinne der von Bortz und Döring (2002) unterschiedenen qualitativen Merkmale von Stichproben werden für die über Testwerte voneinander getrennten Teilstichproben spezifische Wahrscheinlichkeiten bekannt, die eine Abschätzung der mit ihnen als Gruppe identifizierten Kinder mit klinischen Ausprägungen an depressiven Störungen erlauben. Eine sichere individuelle Identifikation von Schulkindern mit klinischen Ausprägungen affektiver Störungen ist weder mit CDI noch mit DTK möglich. Jedes bekannte Kind könnte eines von denen mit irrtümlichen zu hohen oder zu niedrigen Gesamtpunktwerten sein. Nach Punktwerte aus dem DTK käme beispielsweise eins von fünf Schulkindern mit hohen Gesamtpunktwerten in Gefahr irrtümlich als im klinischen Sinne belastet zu gelten.

Die bekannt werdenden Daten sind epidemiologisch nutzbar. Sie erlauben eine statistisch korrekte Abschätzung der Auftretenshäufigkeit depressiver Störungen, ohne sie aber verlässlich in Major Depression und Dysthymia unterscheiden zu können.

Die Untersuchungsarbeit mit zwei ähnlichen Verfahren zur Selbsteinschätzung der Ausprägung depressiver Symptome bietet Möglichkeiten der wechselseitigen Ergebniskontrolle auf unterschiedlichen Ebenen, also sowohl im Hinblick auf Gesamtpunktwerte als auch auf Itemebene.

Ein drittes Verfahren, die Skala ESDM nach Domènech et al. (1985) setzt der auf Selbsteinschätzungen beruhenden Verfahren eine andere Perspektive entgegen. ESDM ist ein dem CDI auf Itemebene ähnliches aber für Fremdeinschätzungen von Schulkindern durch ihre Lehrkräfte benutztes Instrument. Zum Untersuchungsgegenstand werden damit depressive Symptome an Schulkindern in den Perspektiven von Selbst- und Fremdeinschätzung.

Diesen Messinstrumenten werden zwei weitere Verfahren zur Informationsnahme zur Seite gestellt. Das eine, SDQ von Goodman (1997) erlaubt das Aufnehmen von Indizien über das potentielle Auftreten komorbider Störungen. Das Verfahren hat Hinweischarakter.

Eine weitere Informationsnahme findet über einen kurzen Erhebungsbogen statt, der im Sinne nach speziellen Lebensumständen der Schülerinnen und Schüler fragt. Diesem Bogen ist zum einen eine die Durchführung der Untersuchung für die beteiligten Schulkindern erleichternde Funktion zugeordnet. Die Fragen beziehen sich auf alltägliche Lebensumstände. Die Antworten auf diese Fragen geben aber zum anderen die Möglichkeit, Aspekte alltäglichen Lebensumstände zu den Daten aus DTK und CDI in Beziehung setzen zu können.

Die über die Wahl der Messinstrumente betriebene operationale Bestimmung des Untersuchungsgegenstandes der durchzuführenden epidemiologischen Untersuchung zum Auftreten depressiver Störungen bei Schulkindern führt ein Aussagennetz von Selbstauskünften und Fremdeinschätzungen ein, mit dem in ihrem Aussagewert statistisch einschätzbare Indizien für das Auftreten depressiver Störungen in der untersuchten Stichprobe bekannt werden. Der epidemiologische Wert der Untersuchung liegt in den durch die gewonnenen Daten abzusichernden Hochrechnungen für das Auftreten depressiver Störungen bei Schulkindern in Griechenland.

Der für diese Messinstrumente auszuarbeitende Untersuchungsplan soll zum einen die Aussagemöglichkeiten für die untersuchte Stichprobe optimieren und zum anderen die Grundlage für eine in ihren Risiken abschätzbare Hochrechnung der Kennwerte dieser Stichprobe auf die griechische Population an Kindern sichern, die diese Stichprobe durch ihren Anteil an ihr partiell vertritt.

Kapitel 4

Der unter gegebenen Restriktionen gewählte Ansatz zur Suche nach Hinweisen zum potentiellen Auftreten depressiver Störungen bei griechischen Schulkindern

Nachdem im vorangegangenen Kapitel unterschiedliche Möglichkeiten der Bestimmung eines dem Untersuchungsziel angemessenen Untersuchungsobjektes erörtert worden sind, ist nun darzustellen, wie der Ansatz für die dann auszuführende Untersuchung vollzogen wird. Da das eigentlich Objekt, die depressive Störung einschließlich ihrer Varianten, nicht selbst, unmittelbar, erfassbar ist, sondern nur mittelbar erschlossen werden kann, sind bei der Wahl des Ansatzes die entsprechenden Restriktionen zu beachten. Als Ansatz wird hier die konzeptuelle Verbindung des Untersuchungsanlasses mit dem Beginn der Datenaufnahme, verstanden. Der Ansatz erklärt, warum die Untersuchung so geführt werden soll, wie das nun bestimmt wird. Die dazu benutzte konzeptuelle Verbindung ist eine als spezifisch angesehene Struktur in der kognitiven Repräsentation der angesprochenen Realität.

Die durchzuführende Untersuchung soll einen bestimmten Realitätsausschnitt nach definierten Objekten absuchen. Die Ausbeute dieser Suche ist das Material, aus dem informationstragende Merkmale zur Gewinnung eines Datensatzes verwendet werden. Über die Analyse dieses Datensatzes sollen sich die Ergebnisse der Untersuchung ergeben.

An dieser etwas verfremdenden Darstellung des Vorhabens lassen sich drei Bereiche des dann zu beschreibenden Ansatzes unterscheiden. Das sind 1. die konzeptuelle Verbindung des Untersuchungsinteresses mit dem danach zu bestimmenden Untersuchungsobjektes. Dem entsprechend ist 2. das Untersuchungsobjekt darzustellen. Dem hat 3. die Beschreibung der geplanten Vorgehensweise in allen ihren Aspekten zu folgen. Danach wäre mit der Durchführung der Datenaufnahme innerhalb dieses empirischen Vorhabens zu beginnen.

Die aufgeführten Inhalte des Ansatzes korrespondieren ausnahmslos, einschließlich ihrer Subkategorien, mit methodischen Fachtermini (s. etwa Bortz & Döring, 2002). Die in Abweichung davon gewählte Form der Beschreibung soll helfen, zwei qualitative Wechsel im Tätigkeitsablauf zu verdeutlichen.

Der erste Wechsel ist der aus der Ebene kognitiver Repräsentationen zur Tätigkeitsebene bei der Suche nach dem Untersuchungsobjekt. Auf der Ebene kognitiver Repräsentation wird der Begriff der „depressiven Störung bei Kindern und Jugendlichen“ verwendet. Niemand wird dieses Konstrukt an den Kindern und Jugendlichen einer Stichprobe finden. Gesucht wird oder werden die Untersuchungsobjekte, die im Ansatz der Untersuchung als Träger depressionsspezifischer Informationen bezeichnet und bestimmt worden sind. In der Terminologie der Konstruktvalidierung sind es die Korrelate des hypothetischen Konstrukts.

Der zweite qualitative Wechsel wird vollzogen, in dem die in der jeweiligen Stichprobe gesammelten Untersuchungsobjekte im jeweils durchzuführenden Analyseverfahren sequentiell in bedeutungstragende kognitive Repräsentationen transformiert werden.

Auf diesen Sachverhalt wird nochmals ausdrücklich aufmerksam gemacht, weil er, zumindest bisher und gegenwärtig, ein methodisches Grundproblem aller empirischen Untersuchungen zum Auftreten depressiver Störungen enthält. Dass diese Schwierigkeit alle Untersuchungen betrifft,

die auf Konstrukte bezogen sind und nur deren Korrelate erfassen können, hebt das methodische Problem nicht auf.

4.1 Von der Ausgangsproblematik zur Bestimmung depressionsspezifischer Informationsträger als Untersuchungsobjekte

Eine Studie über griechische Schulkinder, welcher speziellen Zielsetzung sie auch folgen mag, ist zunächst nur eine von vielen anderen ausübenden wissenschaftlichen Tätigkeiten. Popper (2005) führt sie alle in ihrer handlungsregulatorischen Entwicklung auf einen trotz unterschiedlicher Inhalte formal übereinstimmenden Ursprung zurück. Nach ihm gilt, „...dass jede wissenschaftliche Entwicklung nur so zu verstehen ist, dass ihr Ausgangspunkt ein Problem ist, oder eine Problemsituation, das heißt, das Auftauchen eines Problems in einer bestimmten Situation unseres Gesamtwissens.“ (Karl Popper, 2005, 9. Auflage, S. 19). Mit dem Begriff „Problem“ meint Popper jede subjektiv oder objektiv wahrgenommene Erscheinung, die zum gegebenen Zeitpunkt einer Klärung bedarf, aber nicht ausreichend erklärt werden kann. Das jeweils vertretene Erkenntnisinteresse ist nicht befriedigt, insoweit ist der fragliche Sachverhalt nicht unter der durch das Weltwissen repräsentierten kognitiven Kontrolle.

Solche Probleme werden durch die individuellen oder kollektiven Interpretationen von Wahrnehmungen konstituiert und beschrieben. Zum Ausgangspunkt wissenschaftlichen Handelns entwickeln sie sich, wenn die Lösung des erfassten Problems als persönliche Aufgabe übernommen wird. So erklärt Hacker (1986) den jeweiligen Beginn der Regulation von Tätigkeiten und sinngemäß dann auch den Beginn wissenschaftlicher Arbeit.

Das Bemühen um die Lösung erfasster Probleme kann sehr unterschiedlich motiviert sein. Das Suchen nach Problemlösungen mag etwa durch ein Menschen häufig und in der Wiederholung lebenslang bewegendes Erkenntnisinteresse veranlasst sein. In anderen Fällen könnte auch das Erleben existenzieller Befürchtungen oder schon bestehender Notlagen zur Lösung von damit verbundenen Problemen drängen.

Die vom Wissenschaftler erlebte Veranlassung zur Übernahme von Aufgaben im Sinne Hackers (s.o.) unterscheidet sich inhaltlich wahrscheinlich stark, aber in der ihr zu Grunde liegenden psychischen Struktur nur unwesentlich von den durch alle Menschen erlebten Anlässen zu eigener Tätigkeit.

Die Regulation von Tätigkeiten ist eng mit der Individualentwicklung des Denkens und Handelns verbunden. Tätigkeiten auf vorwissenschaftlicher und auf wissenschaftlicher Ebene haben grundlegende psychische Gemeinsamkeiten. So gehen zwar keinesfalls alle, wohl aber zahlreiche Anlässe für wissenschaftliches Arbeiten auf das Erleben von Problemen im vorwissenschaftlichen Bereich zurück. Das Erleben der Gefährdung menschlicher Existenz durch Krankheiten gibt bislang immer wieder neue Beispiele dafür. Erkennbar wird in diesem Zusammenhang auch die dynamische Bedeutung der Emotionen in der Genese von Tätigkeiten. Potentielle regulatorische Bedeutung haben alle Emotionen, sowohl die so genannten positiven als auch die ihnen als gegenüber stehend gedachten negativen. Das Erleben von Leiden, Enttäuschungen und Hilflosigkeit kann die wissenschaftliche Suche nach Möglichkeiten zur Überwindung von Krankheiten ebenso voran treiben wie etwa eine durch den wissenschaftlich Tätigen antizipierte Genugtuung über seinen individuellen Beitrag zu kollektiven Arbeiten an der Schaffung oder Stabilisierung gesellschaftlich getragener protektiver Systeme.

Die regulative Entwicklung der wissenschaftlichen Tätigkeit von dem durch Popper beschriebenen Erfassen eines speziellen Problems und dessen individueller Interpretation, bis hin

zur Übernahme einer Aufgabe in dem von W. Hacker vertretenen Verständnis und dann weiter bis zur nachfolgenden Planung, der Realisierung eines Vorhabens und dessen Evaluierung ist selbst ein psychologischer Gegenstand. Damit hätten sich sowohl die Allgemeine Psychologie als auch die Handlungspsychologie oder die Arbeitspsychologie zu befassen. In besonderer Weise ist empirische wissenschaftlicher Tätigkeit aber Gegenstand der psychologischen Methodik (Bortz und Döring, 2002), an deren Positionen der Ansatz dieser Untersuchung orientiert ist.

Ein psychologisches Vorhaben, dass mit dem Auftreten depressiver Störungen in Stichproben und den Möglichkeiten der Übertragung entsprechender Aufschlüsse auf Populationen befasst ist, geht, wie hier schon anfänglich unter Hinweis auf Bortz und Döring (2002) erklärt, möglicherweise auf die Wahrnehmung einer, wissenschaftliche Arbeit erforderlich werden lassenden, Notfallsituation zurück. Von der Möglichkeit dazu ist auszugehen, weil die Notsituation mangels weiterreichender Informationen nicht erwiesen ist. Eben dieser Umstand aber drängt zur Durchführung einer entsprechenden Studie. Nicht der Notstand selbst, sondern die Ungewissheit darüber, ob er besteht, legen diese Studie nahe, weil sie helfen könnte, die Ungewissheit zu beenden. In diesem Sinne ist der Ausgangspunkt der anstehenden Arbeit nach Festinger eine kognitive Dissonanz und als solche ein nach Popper am Ausgangspunkt der Arbeit zu erwartende Problem. Dementsprechend ist auch hier die weiter oben als unumgänglich beschriebene und von oberen kategorialen Ebenen der Problembeschreibung her fortschreitend zu führender Vergegenständlichung des Untersuchungsgegenstandes zu beginnen. Die ist gelungen, wenn der bei Durchführung der Datenaufnahme zweifelsfrei erkennbaren ist.

Nach einem ersten Arbeitsschritt zur Lösung dieser Aufgabe ist das Aufnehmen von Informationen über das Auftreten von depressiven Störungen bei Kindern und Jugendlichen das Untersuchungsziel. Das Erkennen von Häufigkeiten ist wesentlich, weil die wissenschaftliche, gesellschaftliche und soziale Gewichtung der potentiellen Problematik in Relation zu quantitativen Aspekten der eventuell gegebenen Notlage steht. Ein in Griechenland auftretender Einzelfall einer psychischen Störung wäre für die betroffene Persönlichkeit eine Notlage, die vervielfachte sich bei entsprechenden Prävalenzraten zu einer griechischen Notlage. Depressive Störungen bei griechischen Kindern und Jugendlichen sind, nicht grundsätzlich anders als bei betroffenen Erwachsenen, spezielle Belastungen der somatischen und psychischen Gesundheit. Solche Störungen ziehen, wegen der akuten und der prospektiven Auswirkungen für die Betroffenen, zu denen mittelbar auch Personen in deren sozialer Umgebung gehört, sowohl individuelle als auch soziale und gesellschaftliche Probleme nach sich. Die demnach wissenschaftlich notwendigen populationsbeschreibenden Untersuchungen zum Auftreten depressiver Störungen im Kindes- und Jugendalter haben innerhalb der Bemühungen um das Erkennen des Ausmaßes der gesellschaftlichen Belastung durch psychische Störungen ihre spezielle und nicht ersetzbare Funktion. Die erfüllen sie durch die Erarbeitung von Informationen über das Ausmaß und die Bedingungen der Verbreitung emotionaler Störungen in Stichproben und Populationen. Auch im Hinblick auf depressive Störungen bei Kindern und Jugendlichen dienen sie der realitätsnahen Orientierung über die Verbreitung der Störungen in einer Gesellschaft. Damit schaffen sie Voraussetzungen für das gesellschaftliche Erkennen der mit dem Auftreten von Depressionen verbundenen gesellschaftlichen Probleme, aber sie bewirken oder sichern es nicht. Das mit ihnen gesammelte Wissen über die Prävalenzen der Störungen und über Bedingungen von Prävalenzvariationen vermittelt Beiträge zum Erlangen einer störungsspezifischen gesellschaftlichen Handlungsfähigkeit, stellt die aber nicht her.

Durch populationsbeschreibende Untersuchungen wird nicht zwangsläufig gesichert, dass eine Gesellschaft das erarbeitete Wissen zu nutzen weiß. Zu diesem Wissen gehören auch Nachweise über die mittelbar entstehende beträchtliche ökonomische Belastung der Gesellschaft aus dem gehäuften Auftreten depressiver Störungen. Dazu verweist Laux (in Laux, 2002) auf zahlreiche

für Deutschland geltende empirische Belege, zu denen auch die unverhältnismäßig hohen und permanent durch depressionsbedingte Arbeitsunfähigkeit entstehenden Kosten zu rechnen sind. Die in der epidemiologischen Perspektive erkennbaren Aufschlüsse lassen erfassbar werden, welchen Herausforderungen eine Gesellschaft gegenüber steht, die sich bemühen möchte, präventive und kurative Sorge für die psychische Gesundheit ihrer Kinder und Jugendlichen zu übernehmen. Sowohl die Geschichte zurückliegender Jahrzehnte und Jahrhunderte als auch die Gegenwartsgeschichte zeigen Staaten, die sich, sicherlich aus sehr unterschiedlichen Gründen, wenig um die Gesundheit ihrer Kinder und Jugendlichen sorgen und schon gar nicht um die Verhinderung oder die Überwindung psychischer Störungen. Die Existenz von Staaten sichert nicht zwangsläufig das Wohlergehen der in ihnen lebenden Menschen.

In der weiteren Entwicklung des Ansatzes der hier vorzubereitenden Untersuchung bleibt zu prüfen, ob die mit dem ersten Arbeitsschritt vollzogene Vergegenständlichung bereits dem weiter oben eingeführten Prüfkriterium der zweifelsfreien Erkennbarkeit nachzukommen vermag.

Der Absicht, mit Hilfe populationsbeschreibende Untersuchungen nach Möglichkeit mittelfristig und langfristig zur Reduktion des Ausmaßes der auftretenden Störungen beizutragen, ergibt sich aus dem Umstand, in einigen Ländern epidemiologische Untersuchungen zum Auftreten von depressiven (emotionalen) Störungen bei Kindern Jugendlichen schon seit einigen Jahrzehnten fortlaufend durchgeführt worden sind, ohne dass sich die Prävalenzen für das Auftreten emotionaler Störungen bei Kindern und Jugendlichen erkennbar gesenkt hätten, was eben diese Untersuchungen selbst nachweisen.

Die Beziehungen zwischen individuellen Notlagen und deren wissenschaftlicher Durchdringung einerseits und der gesellschaftlichen Praxis andererseits wird von epidemiologischen Untersuchungen offensichtlich nicht oder nur unzureichend erfasst. Dann bliebe die Frage aufzuwerfen, welcher Teilbereich einer wissenschaftlichen Disziplin auch diese Problematik erkundet und darstellt.

Das für die hier anzusetzende Untersuchung wahrgenommene spezielle griechische Forschungsinteresse konstituiert sich aus der Divergenz zwischen der kleinen Zahl bereits vorliegender depressionsspezifischer Studien über griechische Schulkinder einerseits und der aktuell wie auch entwicklungspsychologisch hohen Bedeutung depressiver Störungen bei Kindern und Jugendlichen andererseits. Die gesellschaftlichen Relevanz des Wissens über das Auftreten des Störungsbildes unter Schulkindern ergibt sich einerseits aus dem sozialen Gewicht der auftretenden individuellen Belastungen, andererseits hat das jeweils bestehende numerische Verhältnis zwischen den von Störungen belasteten und den unbelasteten Schulkindern auch prognostischen Wert für die Einschätzung der psychischen und physischen Gesundheit in der erwerbstätigen griechischen Bevölkerung der kommenden Jahrzehnte. Eng damit verbunden sind die daraus erwachsenden künftigen ökonomischen Belastungen (s. Laux, 2002).

Der Ansatz der durchzuführenden Untersuchung geht von dem unzureichenden Wissen über das Auftreten depressiver Störungen unter griechischen Schulkindern aus. Da depressionsspezifische Untersuchungen, die populationsbezogene Informationen vermitteln könnten, weitgehend fehlen, verbesserte jede daraufhin durchgeführte Untersuchung dieser Art mit ihren Ergebnissen den nutzbaren Wissensstand.

Die Zielsetzung der danach anzusetzenden Arbeit ist entsprechend der dargestellten Problemsituation die folgende:

„Die anzusetzende Untersuchung soll in einer hinreichend großen und möglichst repräsentativen Stichprobe von griechischen Schulkindern nach Informationen über das Auftreten von depressiven Störungen suchen, um mit deren Hilfe das Auftreten solcher Störungen in der Population griechischer Schulkinder einschätzen zu können.“

Die nachstehende Auflistung fasst den konzeptualisierten Zusammenhang zwischen dem als Problem wahrgenommenen Defizit an populationsbezogenem Wissen über depressive Störungen unter griechischen Schulkindern und dem spezifisch darauf reagierendem Ansatz einer Untersuchung zusammen.

1. Die in Griechenland betriebene wissenschaftliche Psychologie kann einen gesellschaftlich nutzbringenden Beitrag zur Verbesserung der Entwicklungsbedingungen griechischen Kinder und Jugendlicher leisten
2. Zum Ausmaß des Auftretens depressiver Störungen unter griechischen Schulkindern fehlen empirische Informationen. Damit besteht sowohl die Gefahr einer Verkennung der Betroffenheit Griechenlands als auch die der Fehleinschätzung des gesellschaftlichen und sozialen Handlungsbedarfs.
3. Dem kann durch das empirische Erfassen von Informationen über depressive Störungen bei griechischen Schulkindern entgegen gewirkt werden. Die hier angesetzte empirische Studie ist ihrer Zielsetzung nach geeignet, einen solchen Beitrag zu leisten.
4. Nach der Zielsetzung ist eine der Aufgabe entsprechende Stichprobengröße zu bestimmen. Die minimale Größe für eine zu Hochrechnungen auf Populationen nutzbare Stichprobe wäre nach Bortz und Döring (2002) mit 1300 Schulkindern eingestellt. Unter Einbeziehung eines durch die Forderung nach Repräsentativität potentiell erwachsenden Mehrbedarfs wird die Stichprobengröße auf mindestens 1900 Schulkinder festgelegt.
5. Der Forderung nach Repräsentativität wird nachzukommen versucht, indem die für das Auftreten depressiver Störungen bislang als relevant bekannt gewordenen persönlichen Merkmale von Alter und Geschlecht näherungsweise an den Bevölkerungszahlen orientiert werden. Als weitere, die Repräsentativität stützenden persönlichen Merkmale werden Schulklasse, Schule, Schulart und die Positionierung des Schulortes nach Einwohnerzahl und regionaler Lage mitberücksichtigt.
6. Das Untersuchungsziel verlangt die Erfassung von Informationen, die Hinweise zum Auftreten depressiver Störungen unter den Angehörigen der untersuchten Stichprobe zu geben vermögen. Ausführlich dargelegt wurden bereits zuvor die auch mit dieser Untersuchung nicht auszuräumenden methodischen Hindernisse beim Versuch der Aufnahme möglichst zweifelsfreier und valider Informationen zum Auftreten depressiver Störungen.
7. Das Hauptproblem bei der jeweils erforderlichen Aufnahme von Informationen über das aktuelle Auftreten einer depressiven Störung besteht in der die Durchführung von objektiven Messungen verhindernden Dominanz der bislang ausschließlichen subjektiv vermittelten Daten über eines der beiden Hauptsymptome einer depressiven Störung, der depressiven emotionalen Situation einer Persönlichkeit.
8. Die Subjektivität der verfügbaren Informationen ist aber nicht nur auf Angaben über den emotionalen Zustand einer Persönlichkeit beschränkt. Weitere Symptome und vor allem auch die im Störungsbild verwendeten zeitlichen Vorgaben, nach der beispielsweise Hauptsymptome ein Jahr bestehen müssten, um diagnostisch verwertbar zu werden, bewirken eine allgemeine Dominanz subjektiver Angaben in der Aufnahme störungsspezifischer Informationen.

9. Für Untersuchungen größerer Stichproben, in denen die relevanten störungsspezifischen Angaben während der zeitlich begrenzten Untersuchungsphase erhoben werden sollen, ergibt sich daraus die Nötigung, die Dominanz subjektiver Angaben hinzunehmen. Das scheint die Praxis aller bisher mit Screening-Verfahren durchgeführten Untersuchungen zu sein. Aber auch für diagnostische Interviews ist nicht zu erkennen, dass sie der Dominanz subjektiv vermittelter Informationen entgehen können. Der Verdacht besteht, dass auch die diagnostische Praxis vom gleichen Sachverhalt betroffen ist.
10. Die auch für diese Untersuchung unumgängliche Bestimmung des Untersuchungsobjektes oder der Untersuchungsobjekte vermag die gegebenen methodischen Restriktionen nicht zu umgehen. Für die vorzunehmende Bestimmung wird wegen der bestehenden methodischen Restriktionen das Ziel gesetzt, eine möglichst weitgehende Annäherung an valide Informationen über das Auftreten depressiver Störungen unter den Schulkindern einer zu bildenden Stichprobe anzustreben.
11. Eine solche Annäherung wird erreicht, wenn für die Informationssuche die nachstehenden Vorschriften eingehalten werden.
 - 11.1 Die aufzunehmenden Informationen sollen über unterschiedliche Messinstrumente erhoben werden (multiple instruments)
 - 11.2 Sie sollen von Messinstrument zu Messinstrument nach unterschiedlichen Konzepten erhoben werden (multiple traits)
 - 11.3 Sie sollen von unterschiedlichen Personen, einschließlich entsprechender Selbsteinschätzungen, gegebene Informationen über das jeweilige Mitglied (Schulkind) der Stichprobe geben
 - 11.4 Die zu erhebenden Informationen sollen auf die Symptomliste des Störungsbildes der >major depression< und der >Dysthymia< bezogen sein.
 - 11.5 Anders als bei diagnostischen Interviews soll die Zahl der einbezogenen Symptome zugunsten einer messtheoretisch hinreichend aussagekräftigen Auswahl begrenzt werden. Die wäre jeweils erreicht, wenn der Informationsgewinn aus einem einzelnen Item keine signifikante Erhöhung der Gesamtinformation nach sich zieht.
12. Die über solche Aufnahme depressionsspezifischer Daten gewonnenen Informationen werden als zu wuchtende Indizien für das Bestehen depressiver Störungen verstanden
13. Das dargestellte Messkonzept folgt der Annahme, dass durch die Aufnahme zahlreicher depressionsspezifischer Informationen aus unterschiedlichen Quellen bei entsprechend hoher intrapersonaler Übereinstimmung in einer Vielzahl von Daten Ergebnisse bekannt werden, die sich statistisch von Zufallsvariationen unterscheiden und Indizien für eine in der Stichprobe durchzuführende Unterscheidung von mutmaßlich mit depressiven Symptomen belasteten und nicht belasteten Schulkindern gestattet.
14. Die Ergebnisse aus der dafür hinreichend großen und in mehreren Aspekten auch repräsentativen Stichprobe können zu gleichsinnigen Schätzungen des jeweiligen Anteils der unterschiedenen Teilgruppen in der dieser Stichprobe entsprechenden Population griechischer Schulkinder genutzt werden.
15. Die mit den Ergebnissen einer solchen Untersuchung mitzuteilenden Schätzungen des Anteils der unterschiedenen Teilgruppen mit und ohne mutmaßliche Belastung durch depressive Störungen an der Population bieten keine Prävalenzraten zum Auftreten depressiver Störungen, wohl aber denen angenäherte quantitative Informationen.
16. Der damit erreichte Wissensstand über das Ausmaß des Auftretens depressiver Störungen unter griechischen Schulkindern wäre deutlich höher als der zum Zeitpunkt des Beginns dieser Untersuchung.

Nach diesen Entscheidungen kann die operante Bestimmung des Untersuchungsgegenstandes vorgenommen werden.

Diese Bestimmung wird hier durch die für die Durchführung der Untersuchung verbindliche Festlegung der Datenaufnahme vollzogen. Untersuchungsgegenstand ist das, was die eingesetzten Messinstrumente in der realisierten Messprozedur erfassen.

Die auszuwählenden Messinstrumente müssen zum einen den allgemeinen methodischen Anforderungen genügen und sie sollen zum anderen den hier zuvor aufgeführten speziellen Forderungen entsprechen.

Den allgemeinen methodischen Anforderungen kommen am ehesten international bewährte Messinstrumente nach. Speziell sollen sie Informationen zum Auftreten der Symptome geben, die in den Symptomlisten zum Störungsbild der >major depression< und der >dysthymen Störung< aufgeführt werden.

Dabei wird auf die diagnostischen Manuale DSM-IV und ICD-10 Bezug genommen.

Von allen aufzunehmenden Informationen wird erwartet, dass sie jeweils Informationen über jedes Mitglied (Schulkind) der zu untersuchenden Stichprobe geben. Weiterhin sollen aber unterschiedliche Personengruppen als unterschiedliche Informationsgeber mitwirken. Die Zahl der Personengruppen wird auf 2, die der einzusetzenden Verfahren auf 4 beschränkt. Beide Bestimmungen folgen verfahrensökonomischen Überlegungen. Als dritte Bedingung waren für die einzusetzenden Messinstrumente unterschiedliche Messkonzepte gefordert worden.

Die nachstehende Tabelle veranschaulicht, welche Messinstrumente nach den zu beachtenden Forderungen gesucht werden.

Tabelle 4.1: Anforderungsschema für die einzusetzenden Messinstrumente

Depressionsspezifische Messverfahren	<u>Informationsgebende Person</u>			
	Kind selbst	Kind selbst	Kind selbst	Lehrkraft
Verfahren* Konzept A	X			
Verfahren* Konzept B		X		
Verfahren* zur Einschätzung von Schulkindern				X
Verfahren zur Einschätzung psychischer Gesundheit			X	

* = depressionsspezifisches Screeningverfahren, x GESUCHT

Den in der obigen Tabelle genannten Anforderungen soll durch die Auswahl geeigneter und international bewährter Messinstrumente nachgekommen werden.

Damit sind zunächst zum Erfassen von Informationen über das Auftreten depressiver Symptome bei Schulkindern konstruierte und international bewährten Messinstrumenten zu suchen, die im Übrigen unterschiedliche Konstruktionsmerkmale tragen sollten. Das dritte Messinstrument solle sich von den beiden ersten in seinen Zielsetzungen bei der Datenaufnahme unterscheiden. Benötigt wird ein Verfahren zur Erfassung von Informationen über die psychische Gesundheit von Schulkindern. Ein solches Verfahren würde nicht nur Informationen über das Auftreten depressiver Symptome, sondern darüber hinaus auch Informationen zum Auftreten von Hinweisen auf weitere psychische Störungen sammeln. Das vierte Verfahren diene der Aufnahme depressionsspezifischer Informationen, die von Lehrkräften über das Auftreten depressiver Symptome bei den Kindern ihrer Schulklasse gegeben werden können.

Als Fragebögen zur Selbsteinschätzung des Ausmaßes der eigenen Betroffenheit durch depressive Symptome wurden im Verlaufe dieser Arbeiten der CDI von Kovacs (1985, 2004) und der DTK von Rossmann (1993, 2005) ausgewählt. Beide Verfahren entsprachen den gesetzten Kriterien.

Für die von Lehrkräften aufzunehmenden Informationen über depressive Symptome bei den Schulkindern der eigenen Schulklasse zeigte sich der ESDM von Domènech und Mitarbeitern (1985) geeignet. Bei seiner Wahl wurden die Berichte über dessen Bewährung in einem mediterranen Land, nämlich bei Untersuchungen mit mehreren Tausenden spanischer Schulkindern, als Empfehlung für dessen Bevorzugung gewertet.

Der gesuchte Fragebogen zur allgemeinen Erfassung psychischer Störungen wurde in dem SDQ von Goodman (1997) gefunden. Seine Wahl empfahl sich nach Bekanntwerden seines Einsatzes in anderen griechischen Studien.

Tabelle 4.2: Liste der für die Untersuchung mit griechischen Schulkindern ausgewählten Messinstrumente

Kennung	Angaben	Funktion in der Untersuchung
CDI	Kovacs (1985; 2004), Selbstrating des Auftretens depressiver Symptome, Kinder und Jugendliche	Aufnahme von Informationen über das Auftreten depressiver Symptome
DTK	Rossmann (1993; 2005), Selbstrating des Auftretens depressiver Symptome, Kinder und Jugendliche	Aufnahme von Informationen über das Auftreten depressiver Symptome
SDQ	Goodman (1994; 1997), Selbstbericht über psychische Störungen bei Jugendlichen; Hinweise auf Symptome weiterer Störungen	Aufnahme von Informationen über das Auftreten psychischer Störungen
ESDM	Skala zur Einschätzung depressiver Symptome bei Kindern durch Lehrkräfte	Aufnahme von Informationen über von Lehrkräften an Schulkindern wahrgenommene depressive Symptome

Von den entsprechend den Angaben in der vorausgegangenen Tabelle ausgewählten Messverfahren liegt zum Zeitpunkt des Planungsbeginns für diese Untersuchung nur eines, das Verfahren SDQ, in griechischer Sprache vor.

Daraus ergibt sich die unabwiesbare Aufgabe, vor Aufnahme der eigentlichen Untersuchung, für die Verfahren CDI, DTK und ESDM jeweils eine Version in griechischer Sprache zu entwickeln, zu erproben und gegebenenfalls auch bis hin zur Einsetzbarkeit in seinen Messqualitäten zu verbessern.

Dieser Aufgabe wurde durch entsprechende Vorarbeiten und dann nachfolgend mit der Erprobung der griechischen Versionen in einer in Thessaloniki durchgeführten Voruntersuchung nachgekommen.

Die von Ch. Kampisiou in Zusammenarbeit mit F. Pistofidou, Prof. Dr. Stogiannidou, Prof. Dr. Bibou-Nakou, K. Gkantinas, S. Tassi, T. Segovia und Dr. M. Casanovas übersetzten Verfahren CDI, DTK und ESDM wurden in einer Stichprobe mit 350 griechischen Schulkindern im Alter von 8 bis 16 Jahren in Thessaloniki von Ch. Kampisiou erprobt und in ihren Messqualitäten geprüft. Alle drei Verfahren erreichten mit ihren griechischen Versionen die Messqualitäten der anderssprachigen Originalfassungen. Die Messqualitäten aller eingesetzten Verfahren werden in nachfolgenden Kapiteln gesondert und ausführlich behandelt.

Mit der Verfügbarkeit griechischer Versionen der CDI, des DTK, der ESDM und des SDQ wurden die Voraussetzungen für die Weiterführung der Planung geschaffen.

In einer verkürzten Schreibweise wäre festzustellen, dass mit der Wahl von vier Messverfahren das gesuchte Untersuchungsobjekt bestimmt worden sei.

Nachstehend wird versucht, die Beschreibung des Untersuchungsobjektes zu präzisieren. Es tritt, da es bei allen an der Untersuchung teilnehmenden Mädchen oder Jungen und allen teilnehmenden Lehrkräften gesucht wird, im Plural auf.

„Untersuchungsobjekte sind die von den teilnehmenden Kindern, Jugendlichen und Lehrkräften auf die über Fragebogen vollzogene Darbietung von depressionsrelevanten Fragen markierten Varianten vorgegebener Antworten.“

Zum Ansatz der Untersuchung gehört die begründete Annahme, diese als Beantwortung von Fragen zu verstehende Markierung von Varianten vorgegebener Antworten sei im jeweiligen Fall kein Zufallsprodukt, sondern vermittele jeweils spezifische Informationen über jedes teilnehmende Schulkind der Stichprobe und über die beteiligten Lehrkräfte.

Diese aufzunehmenden Informationen werden von den Schulkindern in der Perspektive der Selbsteinschätzung und von den Lehrkräften in der von teilnehmenden Beobachtern gegeben. Beide Perspektiven sind durch Kenntnisse über das einzuschätzende Kind ausgezeichnet. In der Perspektive der Selbsteinschätzung sind es die aus dem Selbsterleben der Schulkinde. In der Perspektive der teilnehmenden Beobachter sind das Kenntnisse aus dem Erleben von Lehrkräften mit dem jeweiligen Schulkinde als Mitglied in einer von der betreffenden Lehrkraft unterrichteten Schulklasse.

An dieser Stelle der Erörterungen ist nach der eindeutigen Bestimmung der im Plural auftretenden Untersuchungsobjekte erstmals auch Gelegenheit die darauf zu beziehende Forschungsfrage zu präzisieren. Zwar ist das Forschungsinteresse vorab schon mehrfach dargestellt und erläutert worden, aber die Entwicklung des Forschungsvorhabens von einer allgemein begründeten Absicht zu einer eindeutig objektivierten Beschreibung der zu vollziehenden Tätigkeit nötigt auch zu einer darauf bezogenen Forschungsfrage. Fragen sollen nach Möglichkeit beantwortet werden. In einer empirischen Untersuchung können nur Fragen beantwortet werden, die direkt auf die nach der Durchführung der Untersuchung vorliegenden empirischen Daten bezogen sind. Wie der Untersuchungsansatz ist die einzusetzende Forschungsfrage vom Allgemeinen zum Speziellen zu entwickeln. Die nachstehende Forschungsfrage ist die speziell auf die zu erhebenden Daten im Sinne des vorab beschriebenen Forschungsinteresses bezogene und so späterhin auch zu beantwortende Frage:

„Welche Hinweise auf die Belastung der untersuchten griechischen Schulkinde durch depressive Symptome und depressive Störungen enthalten die mit dieser Untersuchung aufgenommenen Daten und welche Schlüsse ergeben sich daraus für die griechischen Schulkinde insgesamt?“

4.2 Die auf den eingesetzten Untersuchungsgegenstand und auf die dazugehörige Forschungsfrage zu beziehende Durchführung der Datenaufnahme

Nachdem der Untersuchungsgegenstand und die Untersuchungsfrage spezifiziert und präzisiert worden sind, ist, soweit noch nicht geschehen, auf weitere Details der geplanten Durchführung

der Datenaufnahme einzugehen. Die wurden durch die im gewählten Ansatz aufgenommenen Bestimmungen verlangt. Der Ansatz determiniert die Durchführung. So folgt aus dem entworfenen Messmodell, welche Messinstrumente gewählt werden könnten. Die dann vollzogene Wahl der Untersuchungsinstrumente CDI, DTK, SDQ und ESDM verlangt die Organisation ihrer Anwendung. Die Beantwortung der Fragebögen verlangten Schulkinder, die sich darauf einlassen. Insoweit kommt der Bestimmung der Stichprobe eine die gesamte Untersuchung determinierende Bedeutung zu. Das zentrale Geschehen in der geplanten Untersuchung ist der jeweils persönliche Dialog der Schulkinder und Lehrkräfte dieser Stichprobe mit den Fragebögen der Messinstrumente. Damit diese Kommunikation, aus der die Datenbasis der Untersuchung entsteht, zustande kommen kann, wird die helfende Vermittlung vieler Tausende von Menschen benötigt, von Eltern, Lehrkräften, Schulleitungen, Gebäudepersonal und von Menschen, die Funktionsträger innerhalb der Untersuchung von der Gesellschaft getragenen Institutionen sind, die für die Abhaltung von Schule zu sorgen haben, was erst seit der Einführung der Schulpflicht in Griechenland eine öffentliche Aufgabe ist. Wenn insoweit die geplante Durchführung der Datenaufnahme gesichert scheint, wäre die sach- und fachgerechte Durchführung der Datenaufnahme der letzte Schritt zur Erarbeitung der Datenbasis dieser Untersuchung.

Die an die Stichprobe zu stellenden Forderungen sind mitgeteilt und begründet worden. Sie soll repräsentativ im Hinblick auf mehrere Merkmale sein, durch die der Anteil der Schulkinder an der griechischen Gesellschaft aktuell gekennzeichnet ist.

Für die Stichprobe war bislang deren Größe bestimmt worden. Der Umfang der Stichprobe sollte so gewählt werden, dass die Hochrechnung der Ergebnisse auf eine Population griechischer Schulkinder methodisch begründbar realisiert werden kann. In Anlehnung an Bortz und Döring (s.o.) war unter Einbeziehung einiger spezieller Überlegungen eine Stichprobengröße von mindestens 1900 Schulkindern festgelegt worden.

Die Stichprobe sollte, dem allgemeinen Zielsetzungen folgend, repräsentativ sein.

Repräsentativität (siehe auch Bortz und Döring, 2002) wird hergestellt, in dem die Relation von Merkmalen in der Population mit den Relationen dieser Merkmale in der Stichprobe nachgebildet wird. Daraus ergibt sich, dass die Merkmale und deren Relation in der Population vor Untersuchungsbeginn bekannt sein müssen. Ob diese Merkmale für die Messungen in der Untersuchung bedeutsam sein werden, ist unbekannt. Weiterhin müssen aus einer Fülle von potentiellen Merkmalen der Population die ausgewählt werden, die in der Untersuchung berücksichtigt werden sollen. Welche Bedeutung diese Auswahl für die durchzuführenden Messungen haben werden oder haben könnten, ist ebenfalls weitgehend unbekannt. Die Herstellung von Repräsentativität schützt aber zumindest für die ausgewählten Merkmale vor extremer Stichprobenbildung, in denen eventuell in der Population auftretende Teilgruppen nicht oder nicht hinreichend berücksichtigt wären. Diese Vorsorge kann aber nur im Hinblick auf eine meist relativ geringe Zahl ausgewählter Merkmale getroffen werden. Eine Ausnahme bildeten Verhältnisse, von denen bereits bekannt ist, dass sie mit dem gesuchten zentralen Merkmal in bestimmbareren Zusammenhängen auftreten.

Die zu bildende Stichprobe der hier geplanten Untersuchung soll im Hinblick auf die nachstehend aufgeführten Merkmale repräsentativ werden:

1. Geschlecht der Schulkinder,
2. Altersverteilung in den Altersklassen 8 bis 15 Jahre,
3. Verteilung auf Grundschule und Gymnasium (Schultypen),
4. Verteilung der Stichprobe auf Schulorte in drei Hauptregionen Griechenlands, (nördliches Festland, südliches Festland, griechische Inseln)
5. Verteilung der Stichprobe auf Ortsgrößen in Griechenland

Die Feststellung der Relationen in der Population wurde unter Zuhilfenahme des amtlichen Statistischen Handbuchs für Griechenland (Εθνική Στατιστική Υπηρεσία, 2001) vorgenommen. Die entsprechenden Relationen in der zu ziehenden Stichprobe sollen den in der Population möglichst nahe kommen. Die am jeweiligen Schulort vorfindbare Zahl von Schulkindern ist vorab nicht mit Sicherheit zu bestimmen. Die Zahl der Schulkinder einer Klasse ist nicht immer identisch mit der Zahl der an einem bestimmten Tag anwesenden Kinder. Vorgesehen ist, Disproportionen in der sich jeweils einstellenden Verteilung möglichst durch Auffüllung der Stichprobe und nicht durch Streichungen zu überwinden. Durch die in der Vorbereitung der Untersuchung zu vollziehende Einsichtnahme in die jeweilige Schulstatistik wird versucht, die angestrebten Relationen in der Stichprobe bereits vorab durch Auswahl unter den verfügbaren Klassen angenähert herzustellen. Erwartet wird, dass eine angenäherte Repräsentativität hergestellt werden kann.

Da auch die Relationen der Verteilung von Schulkindern der entsprechenden Altersgruppen von Schulkindern auf drei Landesteile Griechenlands und da wiederum auf Schulstandorte mit unterschiedlichen Bevölkerungszahlen in der Stichprobe repräsentativ abgebildet werden soll, waren entsprechende Berechnungen notwendig, um die Schulstandorte auszuwählen, die in die untersuchte Stichprobe einbezogen werden sollen. Die schließlich gebildete und dann untersuchte Stichprobe an griechischen Schulkindern wird an einer in diesen Ausführungen nachfolgenden Stelle numerisch exakt beschrieben werden. Hier folgen zunächst die nach den mitgeteilten Vorschriften bestimmten diversen Relationen, die in der Stichprobe abgebildet werden sollen. Weiterhin wird die Liste der ausgewählten Schulstandorte wiedergegeben.

Tabelle 4.3: Liste der für die ausgewählten Merkmale in der Stichprobe herzustellenden Relationen

1. Geschlecht der Kinder
Mädchen : Jungen
49 : 51
2. Altersverteilung in den Altersklassen 8 bis 15
Jahre: 8 9 10 11 12 13 14 15
9 : 11 : 12 : 13 : 17 : 17 : 16 : 5
3. Verteilung auf Grundschulen und Gymnasien (Schultypen)
Grundschulen : Gymnasien
49 : 51
4. Verteilung der Stichprobe auf Schulorten in drei Hauptregionen Griechenlands (nördliches Festland, südliches Festland, griechische Inseln)
Nord : Süd : Inseln
5. Verteilung der Stichprobe auf Ortgrößen in Griechenland
Großstädte : Mittelstädte : Kleinstädte
40 : 32 : 28

Tabelle 4.4: Liste der ausgewählten Schulorte

Relationen in Hundertstel		
Athen	20	
Thessaloniki	13	
Heraklio	7	
Großstädte insgesamt	Teilsomme	40
Veria	10	
Ksanthi	9	
Karditsa	7	
Rethymno	6	
Mittelstädte insgesamt	Teilsomme	32
Galatista	6	
Neo Agioneri Kilkis	5	
Stavroupoli	4	
Itea	6	
Episkopi	7	
Kleinstädte insgesamt	Teilsomme	28

Die Annäherung an alle die Menschen, die dafür gewonnen werden sollen, die Datenaufnahme zu unterstützen, folgt zweckmäßiger Weise den gegebenen gesellschaftlichen Strukturen. Die Untersuchung soll Schulkinder in ihren Schulklassen und innerhalb des Schulalltags zur Mitarbeit gewinnen. Der Zugang zu den Schulen, den Lehrkräften und den Schulkindern wird gesellschaftlich durch das Ministerium in Athen verwaltet, das für Schulbildung zuständig ist. Die Organisation der Durchführung beginnt dementsprechend mit der Stellung eines Antrags, in dem um die Genehmigung der geplanten Untersuchung gebeten wird. In dem Antrag ist darzulegen, welchem Interesse die Untersuchung dient, welche Leistungen von der Schulverwaltung, den Lehrkräften und den Kindern erwartet werden und wie dem Einspruchsrecht der Eltern nachgekommen werden kann. Darzulegen ist ebenfalls, wie die Anonymität der Durchführung gesichert werden soll. Die Anonymität soll den Kindern, den Eltern und den Lehrkräften dienen und gilt insoweit als eine zu verteidigende Errungenschaft. In Untersuchungen wie der durchgeführten, wo Daten über die Gesundheit von Kindern erhoben werden, schlägt die distriktale Sicherung der Anonymität in das Gegenteil des angestrebten Zieles um. Die Anonymität führt dazu, dass den Kindern, die auf Grund der Untersuchungsdaten als hilfsbedürftig gelten sollten, nicht geholfen. Die Anonymität der Durchführung überlässt die Kinder dem ungewissen Fortgang der Entwicklung depressiver Störungen. Hier wird ausdrücklich die Position vertreten, dass diese Konsequenz einer anonymen Durchführung nicht hinnehmbar ist. Wenn ernsthaftere Anlässe zur Hilfe bekannt werden, muss erste Hilfe vermittelt werden. Die Regeln für die Sicherung der Anonymität können und müssen dann auch entsprechend modifiziert werden.

Die organisatorischen Arbeiten werden in den späterhin zu analysierenden Daten nirgendwo unmittelbar repräsentiert sein, aber sie haben zweifellos Einfluss auf die Datenaufnahme, weil sie die Untersuchungsbedingungen beeinflussen. Die Schul Kinder sind strukturell Mitglieder des Schulsystems. Die Untersuchenden dringen mit ihrem Vorhaben unabweisbar in die örtlichen Schulsysteme ein. Sie schaffen damit soziale Abläufe, deren Ereignisse die jeweilige aktuelle Situation in der Schule und der teilnehmenden Schulklasse beeinflussen. Diese Einflussnahme teilt sich den Schulkindern mit und beeinflusst auch sie in ihrer Teilnahmebereitschaft. Im Interesse der Untersuchenden und der Untersuchung liegt es, ein möglichst günstiges soziales Untersuchungsklima entstehen zu lassen. Die besondere Schwierigkeit für ein Erreichen dieser Zielsetzung besteht zum einen darin, dass vorab nicht bekannt ist, wie ein solches günstiges Klima beschaffen sein sollte. Zum anderen wäre es herzustellen und müsste potentiell herstellbar sein. Daraus resultiert die kaum auflösbare Schwierigkeit bei Versuchen, ein untersuchungsfreundliches Sozialklima in den beteiligten Schulen herzustellen. Andererseits wird mit dem Erkennen solcher Schwierigkeiten der Einfluss des problematisierten schulischen Sozialklimas am Untersuchungstag nicht aufgehoben. Auch der Hinweis, in vielen vergleichbaren Untersuchungen fände sich das Problem nicht erwähnt. Damit ist das Problem aber auch nicht gelöst. Vielleicht wären die Ergebnisse aus Untersuchungen zu entsprechenden Einflussvariablen aufschlussreich und könnten möglicherweise dazu beitragen, unerwartete Untersuchungsergebnisse verständlich werden zu lassen.

Für die hier darzustellende Untersuchung wurde festgelegt, neben den vorgeschriebenen Genehmigungen vor der Untersuchung auch persönlichen Gesprächskontakt zu den an der Durchführung der Untersuchung mittelbar beteiligten Personen mit der Zielsetzung aufzunehmen, das Untersuchungsvorhaben in allen relevanten Aspekten transparent werden zu lassen und die Freiwilligkeit der Teilnahme zu sichern. Diese Vorgehensweise löst die hier zuvor beschriebene Problematik nicht auf, aber sie versucht, vorab bestimmte Bedingungen an allen Untersuchungsorten einwirken zu lassen.

Dem diene auch die Entscheidung, alle Datenaufnahmen in allen Schulklassen an allen unterschiedlichen Schulstandorten von einer Person, der hier berichtenden Untersuchungsleiterin, vornehmen zu lassen. Das erhöhte den einzusetzenden Aufwand, reduzierte dafür aber die Zahl unkontrolliert einwirkender Durchführungseffekte. Zugleich öffnete das die Möglichkeit einer nach der Datenaufnahme wünschenswerten Kontaktnahme zu den Schulkindern in den teilnehmenden Klassen und zu ihren Lehrkräften. Kontakte zu Schulkindern über die Durchführungsregeln hinaus wurden aber auf die Zeit nach der Durchführung beschränkt.

Da die Datenaufnahme über die Durchführung der ausgewählten Messinstrumente vorgenommen wird, sind die Regeln für die Durchführung dieser Verfahren auch die Vorschriften für die Datenaufnahme insgesamt. Die werden ergänzt durch einen für alle beteiligten Schulklassen identischen Durchführungsrahmen. Der betrifft die vorab an die Schulen und die Schulklassen zu gebenden Informationen, die Einleitung der Datenaufnahme in der Schulklasse, die Reihenfolge in der Durchführung der Messverfahren, die Zeitvorgaben für die Durchführung und den Abschluss der Datenaufnahme.

Die Regelung der Durchführung ist in Termini der Arbeitspsychologie ein hierarchisch-sequentiell strukturierter Tätigkeitsplan, der an dem übergeordneten Untersuchungsziel ausgerichtet ist und für alle zu erkennenden Teilziele erklärt, auf welchen Wegen mit welchen Mitteln das Hauptziel, die Datenaufnahme erreicht werden soll.

Kapitel 5

Die Realisierung der empirischen Studie zum Auftreten depressiver Belastungen unter griechischen Schulkindern

In dem nachstehenden Arbeitsteil wird nun dargestellt, wie der für die Suche nach Hinweisen zum potentiellen Auftreten depressiver Störungen bei griechischen Schulkindern ausgewählte theoretische Ansatz realisiert wurde.

Die Realisierung einer jeden empirischen Untersuchung erfolgt durch Untersuchungsarbeit, die ihrerseits dem speziell für sie entworfenen Untersuchungsplan folgt, durch dessen Umsetzung die Daten gewonnen werden, deren Analyse idealerweise Ergebnisse hervorbringt, die zur Grundlage einer Beantwortung der Forschungsfrage und aller beachteten Nebenfragen geeignet erscheinen.

Kernstück der Untersuchungsplanung für die Studie 2034GR ist, wie in Kapitel 4 dargestellt, das im Hinblick auf die Forschungsfrage konzipierte Messmodell, nach dem durch die Anwendung von 4 Messinstrumenten in einer für griechische Schulkinder vom Schuleintritt bis zum Alter von 15 Jahren hinreichend umfänglichen und dann auch partiell repräsentativen Stichprobe die erforderlichen empirischen Daten für die Beantwortung der Forschungsfrage gewonnen werden sollen.

Nachdem in Kapitel 4 auf den Messprozess und die eingesetzten Verfahren eingegangen worden ist, muss nachfolgend eine Beschreibung der untersuchten Stichprobe aus Kindern, Jugendlichen und Lehrkräften gegeben werden. Da diese Stichprobe einerseits zwar Gegenstand der Planung war, andererseits aber erst durch die Untersuchungsarbeit real gebildet worden ist, finden sich vor deren Beschreibung zunächst Angaben zum Ablauf der Untersuchungsarbeit an den ausgewählten griechischen Schulen.

5.1 Die Durchführung der Untersuchung

Die Untersuchungsarbeit wird im Folgenden in drei Arbeitsphasen unterschieden, die ihrerseits wiederum mehrere Aufgaben umfassen. Eine erste Phase der Untersuchungsarbeit umfasst alle Vorbereitungen, eine zweite die unmittelbare Interaktion mit der Stichprobe während der Datenaufnahme, der dann die Datenanalyse und die Befundbildung folgen. Diese Gliederung folgt ausschließlich Aspekten der Zweckmäßigkeit und würde nach anderen Absichten auch anders zu strukturieren sein.

Zur ersten Phase, der Vorbereitung der Untersuchung, gehörten einerseits alle die Untersuchung betreffenden psychologischen und soziokulturellen Orientierungen und andererseits, dann eben als spezielle Vorbereitung, vor vielem anderen die Bereitstellung der Messinstrumente, ohne die auch keine Datenaufnahme erfolgen könnte. Für drei der benötigten Messinstrumente fehlten griechische Versionen. Die damit erforderlich werdende Adaption der ausgewählten Instrumente zu deren Anwendung in einer griechischen Stichprobe umfasst weitaus mehr als die Übersetzung einer Fragenliste und mitunter auch die Bewältigung kaum erwartbarer rechtlicher und merkantiler Probleme. In zwei Fällen erfolgte eine direkte Kontaktaufnahme mit Testautorin und Testautor, in einem weiteren Fall half der durch Maria Casanovas mittelbar gehaltene Kontakt bestehende Sprachbarrieren zu überwinden. Die Absicherung gegenüber rechtlichen Problemen grenzte an Überforderung und verdiente im Interesse der im internationalen Urheberrecht eher weniger gut qualifizierten Psychologen einer speziellen und spezifisch ausdifferenzierenden Darstellung und kritischen Erörterung. Wenn von Studierenden selbstfinanzierte und ohne

Gewinnabsichten mit eigenem Arbeitsaufwand geleistete Untersuchungsarbeit abgabepflichtig wird, wirft das mehr als ein Problem auf. Obschon diese Vorbereitungen zeitaufwendiger und schwieriger zu realisieren waren als das den Anschein haben könnte, sind die mit der Erarbeitung von griechischen Versionen der ausgewählten Messinstrumente verbundenen fachlichen Probleme die psychologisch gewichtigeren gewesen. Eine der zu bewältigenden Schwierigkeiten war die Sicherung der an psychologischen Kriterien zu orientierenden inhaltlichen Übereinstimmung zwischen der jeweils benutzten fremdsprachlichen Vorlage und der zu konzipierenden griechischen Version. Um die sprachliche Qualität der Übersetzungen gewährleisten zu können, wurde auf die dafür empfohlene Methode der „font-back translation“ zurückgegriffen.

Die jeweils erarbeitete griechische Version von DTK, CDI und ESDM, sowie eine schon existierende griechische Version des SDQ, wurden durch ihren Einsatz in einer Pilotstudie mit 350 Kindern aus Thessaloniki, unter Einbeziehung ihrer Lehrkräfte, einer ersten Bewährungsprobe unterzogen. Die Messqualitäten der ersten Versionen wurden statistisch geprüft. Einige gegenüber der Eichstichprobe abweichenden Kennwerte im DTK wurden mit dem Testautor besprochen und führten zu sprachkritischen Korrekturen an einzelnen griechischen Items. Für die inhaltliche Übereinstimmung der Items der Skala ESDM wurden die durch Maria Casanovas (2003) abgesicherten deutschsprachigen Übereinstimmungen mit der spanischen Version als Kriterium gewählt und entsprechend überprüft.

Nach dem Vorliegen einer mehrfach geprüften griechischen Version eines jeden der vier gewählten Verfahren und nach deren Bewährung in Probedurchläufen, erfolgte in Zusammenarbeit mit Prof. Dr. Kioseoglou von der Aristoteles Universität zu Thessaloniki und unter Nutzung der vom Statistischen Amt Griechenlands publizierten demoskopischen Daten eine im Hinblick auf Alter, Geschlecht, Schultyp, Einwohnerzahl des Schulortes und dessen regionale Lage in Griechenland vorgenommene Stratifizierung der insgesamt zu wählenden Besetzung der Stichprobe. Die „Soll-Werte“ sind im Abschnitt 6.2 dargestellt, die „Ist-Werte“ folgen im nachstehenden Abschnitt 5.2.

Aus der Stratifizierung der aufzunehmenden Stichprobe folgte die Nötigung, in Absprache mit dem zuständigen griechischen Ministerium die Genehmigung zur Durchführung der Untersuchung an ausgewählten Schulorten zu erbitten.

Die dafür zuständige griechische Institution (Ministerium für Bildung und Religion, Pädagogisches Institut) prüfte das Ersuchen auf der Grundlage eines ausführlichen Antrags zur Genehmigung der geplanten Untersuchung und stellte von sich aus Erstkontakte mit den ausgewählten Schulen und den regionalen Schulverwaltungen her. Trotz der wohlwollenden Unterstützung des Ministeriums verweigerten zwei der per Zufall ausgewählten Schulen die Durchführung der Untersuchung. Als Begründung führten sie an, dass die Kinder aufgrund von schon wiederholt durchgeführten ähnlichen Untersuchungen den Ausfall vieler Unterrichtseinheiten hätten hinnehmen müssen, weitere Beeinträchtigungen seien deshalb pädagogisch nicht verantwortbar. Die Schulen lagen in räumlicher Nähe zu Universitäten, was als ein für die Glaubwürdigkeit der Aussagen sprechendes Argument gewertet werden könnte. Die abschließende Sicherung der Durchführung der Untersuchung fand dann jeweils vor Ort über den persönlichen Kontakt zwischen der Untersuchungsleiterin und der Schulleitung statt.

Die zweite Phase der Realisierung der geplanten Studie war dann die dem Plan folgende Interaktion mit Kindern, Jugendlichen und Lehrkräften zur Durchführung der Messprozeduren. Die Messungen wurden zwischen März und Juni 2003 vorgenommen. Die Untersuchungsleiterin führte jede dieser Datenaufnahmen selbst durch. Dem stratifizierten Untersuchungsplan folgend

wurden 112 Schulklassen in 30 Schulen (15 Grundschulen und 15 Gymnasien) an 12 systematisch über Griechenland verteilten Schulorten mit unterschiedlichen Bevölkerungszahlen besucht. Die nachstehende Abbildung (Landkarte) zeigt die geographische Position der ausgewählten griechischen Orte.

Die in der jeweiligen Schulklasse durchzuführende Erhebung fand während einer normalen Schulstunde statt. Zu Beginn wurden die Kinder in einer mit dem Untersuchungsziel vereinbaren allgemeiner Form über das Vorhaben informiert („...*wir wollen erfahren, was die Kinder in eurem Alter denken und was sie fühlen...*“). Allen Kindern und Jugendlichen war zuvor das Recht eingeräumt worden, die Teilnahme abzulehnen. Weiterhin wurden sie über die völlige Anonymität der Teilnahme informiert. Bei der Durchführung von Untersuchungen mit Schulkindern in den Schulräumen kommt insbesondere der Verschwiegenheit gegenüber Lehrkräften, Schulleitung und Eltern eine besondere Bedeutung zu. Danach folgte eine Erklärung des Ablaufs der Untersuchungsprozedur und anhand von zwei Beispielen eine modellhafte Instruktion zum Ausfüllen der Fragebögen. Zur Erörterung von Fragen, die nicht unmittelbar mit dem Ablauf der Durchführung verbunden waren, stand, in der Regel aus schulischen Gründen recht begrenzte, Zeit nach der Unterrichtsstunde zur Verfügung. Die veranschlagte Bearbeitungszeit erwies sich allgemein als zureichen, für die jüngere Kinder (8- und 9jährige) war sie jedoch relativ knapp bemessen, obschon in diesen Schulklassen, wegen der Altersbeschränkungen für den SDQ, eine Skala weniger auszufüllen war. Darüber hinaus hatte diese Altersgruppe relative Schwierigkeiten beim Ausfüllen der Fragenliste FIL, speziell bei den Fragen, die das Bildungsniveau und den Beruf der Eltern betrafen. Diese Sachfragen forderten bei diesen jüngeren Kinder gelegentlich die Hilfe der Untersuchungsleiterin. Bei den älteren Mädchen und Jungen ergaben sich keine erwähnenswerten Ereignisse. In Gymnasien betrug die Bearbeitungszeit durchschnittlich nur etwa 30 Minuten. Abgesehen von einer gewissen Unzufriedenheit der Lehrkräfte, die sie verbal oder nonverbal ausgedrückt haben, als sie gebeten wurden, selbst auch einen ESDM-Bogen für jedes Kind ihrer Klasse auszufüllen, führte die Untersuchung zu keinen Komplikationen.

Abbildung 5: Lage der griechischen Untersuchungsorte



5.2 Die untersuchte Stichprobe

Der Untersuchungsplan war auf zwei unterschiedliche Personengruppen bezogen worden, auf Schüler und Schülerinnen einerseits und auf Lehrkräfte zum anderen. Aus der Gesamtheit der Lehrer und Lehrerinnen waren die um Teilnahme an der Untersuchung zu bitten, die mit der Funktion einer Klassenlehrerinnen oder eines Klassenlehrers betraut waren. waren.

5.2.1 Die Stichprobe der Schülerinnen und Schüler

Im Kapitel 4.2 wurden Informationen vermittelt, die über die Bedingungen unterrichteten, denen die zu ziehende Stichprobe genügen soll. Das Einhalten der dort genannten Bedingungen stellt die Repräsentativität der Stichprobe her. In den bisher dazu gegebenen Darstellungen wird von einer relativen Repräsentativität gesprochen, so denn auch hier, weil der Begriff selbst, siehe auch Bortz und Döring (2003), aus mehreren Gründen zu problematisieren ist. Die numerische Übereinstimmung zwischen einem in der Population auftretendem Merkmal, etwa Alter, mit der Häufigkeit des Merkmals in einer Stichprobe verweist nur auf eine numerische, nicht aber auf eine Gleichheit nach psychologischen Kriterien.

Die Durchführung der Untersuchung vom Februar bis Juni 2003 bei den ausgewählten Schulen führte zur Teilnahme einer Stichprobe von insgesamt 2034 Kindern und Jugendlichen. Diese Gesamtstichprobe ist, je nach dem zu betrachtenden Merkmal, in Teilstichprobe zu gliedern. Für jede Teilstichprobe sind Prozentsätze bestimmbar, die deren Anteil am an dem durch die Gesamtstichprobe repräsentierten Grundwert erkennen lassen. Für den Fall, dass, wie beim Geschlecht, jedes Mitglied der Gesamtstichprobe entweder dem einen oder dem anderen Merkmal zuzuordnen ist, präsentieren solche Prozenhtsätze die Relation der Teilgruppen zueinander und lassen, bezogen auf die betrachteten Merkmale, die Mischung der Gesamtstichprobe erkennen. Merkmale wie Geschlecht, Alter und Wohnort beschreiben dann zweifellos einige Eigenheiten der gezogenen Stichprobe und dann auch Aspekte der Realität in der jeweils besuchten Schulklasse und dort für den Tag der Durchführung der Untersuchung.

Unter Zuhilfenahme des amtlichen Statistischen Handbuches für Griechenland kann dann auch versucht werden, wie in der Studie 2034GR geschehen, die für die Population der griechischen Schülerinnen und Schüler geltenden Relationen zwischen den diese Population kennzeichnenden Merkmalen mit der gezogenen Gesamtstichprobe der Untersuchung proportional korrekt nachzubilden. So besteht die untersuchte Stichprobe beispielsweise zu 48% aus Jungen und dann zu 52% aus Mädchen (vgl. Abbildung 5.1). Nach den Zahlen des Statistischen Amtes für Griechenland sind die national für Kinder und Jugendliche der in Betracht zu ziehenden Altersspanne geltenden Prozentsätze 49% für Jungen und 51% für die Mädchen. Auch hinsichtlich des Alters der untersuchten Schulkindern ergaben sich aus der Durchführung der Untersuchung Prozentsätze von Teilstichproben, die den offiziell mitgeteilten nationalen Relationen nahe kommen (vergleiche Abbildung 2 mit entsprechenden Informationen in Tabelle 4.3). Geplant war die Untersuchung auf Kinder und Jugendliche im Alter von 8 bis 15 Jahren zu beschränken, da aber in den besuchten Klassen auch ältere Jugendliche zu finden waren, wurden auch einige 16-, 17- und 18jährige in die Stichprobe aufgenommen. Das durchschnittliche Alter der Kinder beträgt 11,6. Die leicht unterschiedliche Besetzung in jeder Alterstufe von Jungen und Mädchen ist statistisch nicht signifikant [siehe Tabelle 5.1; $\chi^2(8) = 7,434$; $p = 0,491$].

Abbildung 5.1: Verteilung nach Geschlecht

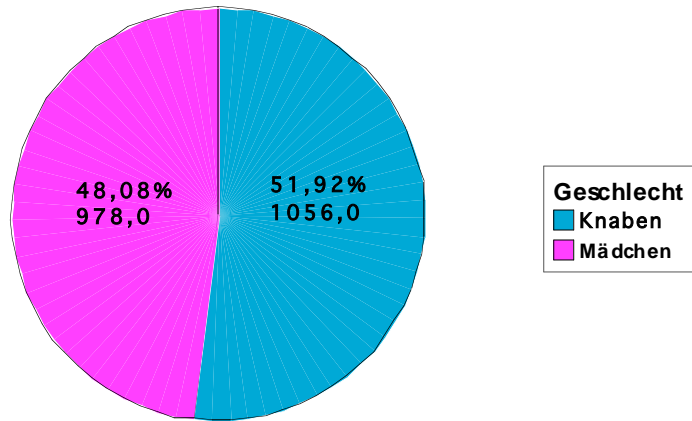


Abbildung 5.2: Verteilung nach Alter

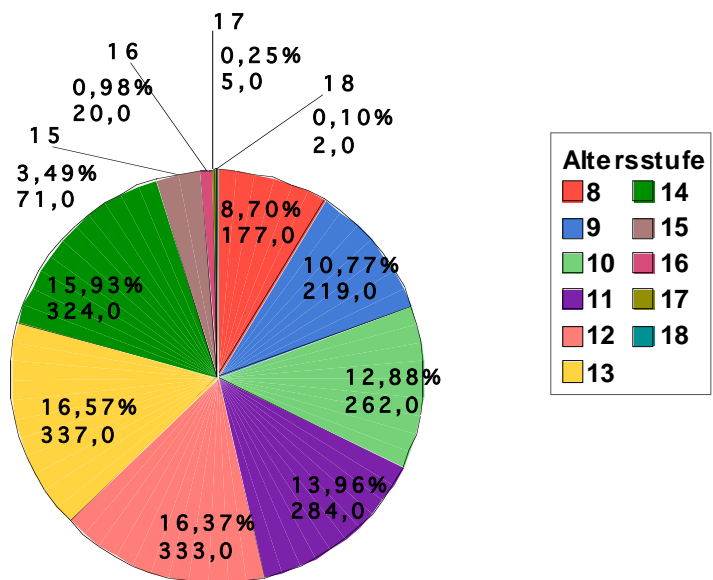
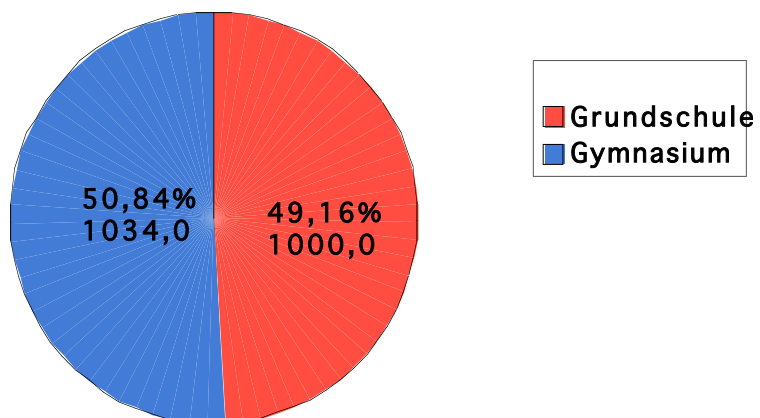


Tabelle 5.1: Geschlechtsspezifische Verteilung der untersuchten Kinder auf unterschiedlichen Altersgruppen

Altersgruppe		Jungen	Mädchen	Gesamt
8	N	98	79	177
	%	55,4	44,6	100,0
9	N	106	113	219
	%	48,4	51,6	100,0
10	N	146	116	262
	%	55,7	44,3	100,0
11	N	146	138	284
	%	51,4	48,6	100,0
12	N	161	172	333
	%	48,3	51,7%	100,0
13	N	170	167	337
	%	50,4	49,6	100,0
14	N	178	146	324
	%	54,9	45,1	100,0
15	N	35	36	71
	%	49,3	50,7	100,0
16+	N	16	11	27
	%	59,3	40,7	100,0
Gesamt		1056	978	2034
		51,9	48,1	100,0

Die folgende Abbildung zeigt die Verteilung der untersuchten Kinder auf die zwei von uns besuchten Schulformen (Gymnasium – Grundschule).

Abbildung 5.3: Verteilung nach Schulform



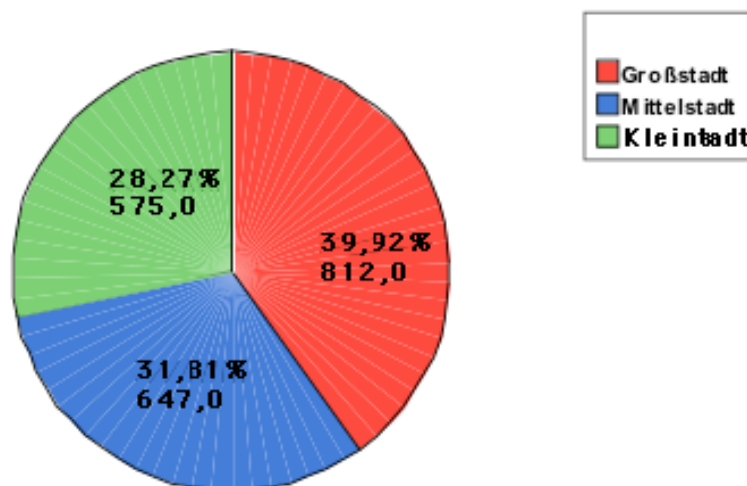
Aus der Abbildung 5.3 ist zu ersehen, dass etwa 51% der untersuchten Kinder die Grundschule besuchen und etwa 49% das griechische Gymnasium. Auch diese Prozentsätze entsprechen den vom Statistischen Amt gegebenen Relationen. Die geschlechtsspezifische Verteilung der untersuchten griechischen Schulkindern auf die genannten Schultypen ist statistisch nicht signifikant unterschiedlich [siehe Tabelle 5.2; $\chi^2(1) = 0,483$; $p = 0,487$].

Tabelle 5.2: Geschlechtsspezifische Verteilung der untersuchten Stichprobe auf die zwei besuchten Schulformen

Schulform		Jungen	Mädchen	Gesamt
Grundschule	N	527	473	1000
	%	52,7	47,3	100,0
Gymnasium	N	529	505	1034
	%	51,2	48,8	100,0
Gesamt	N	1056	978	2034
	%	51,9	48,1	100,0

Die folgende Abbildung (5.4) zeigt die Verteilung der untersuchten Kinder und Jugendlichen auf Orte mit unterschiedlichen Bevölkerungsgrößen.

Abbildung 5.4: Verteilung nach Siedlungsgröße



Der größte Teil der Kinder (39,92%) wohnt in einer Großstadt, dann folgen Kinder, deren Wohnort eine griechische Mittelstadt ist (31,81%) und die verbleibende Zahl von 28,27% besteht aus Kindern, die in einer Kleinstadt wohnen. Auch diese Verteilung darf als repräsentativ für diese Altersgruppe bezeichnet werden. Die amtlichen statistischen Angaben verweisen auf eine Relation von 40%, 32% und 28% in Groß-, Mittel- und Kleinstädten.

Die geschlechtsspezifische Verteilung der untersuchten Stichprobe auf die Wohnorte mit unterschiedlichen Bevölkerungszahlen weicht numerisch von den amtlichen Zahlen gering, aber nicht signifikant ab. [$\chi^2(2) = 3,121$; $p = 0,210$; siehe auch nachstehende Tabelle 5.3]

Tabelle 5.3: Geschlechtsspezifische Verteilung der untersuchten Stichprobe auf Orte unterschiedlicher Größe

Siedlungsgröße		Jungen	Mädchen	Gesamt
Großstadt	N	419	393	812
	%	51,6	48,4	100,0
Mittelstadt	N	322	325	647
	%	49,8	50,2	100,0
Kleinstadt	N	315	260	575
	%	54,8	45,2	100,0
Gesamt	N	1056	978	2034
	%	51,9	48,1	100,0

Die Verteilung der Mitglieder der untersuchten Stichprobe auf 12 Schulorte zeigt Abbildung 5.5. Auch hier treten numerische Abweichungen in der geschlechtsspezifischen Verteilung auf, die keine signifikanten Unterschiede darstellen [$\chi^2(11) = 11,481$; $p = 0,404$; siehe nachstehende Tabelle 5.4].

Abbildung 5.5: Verteilung nach Wohnort

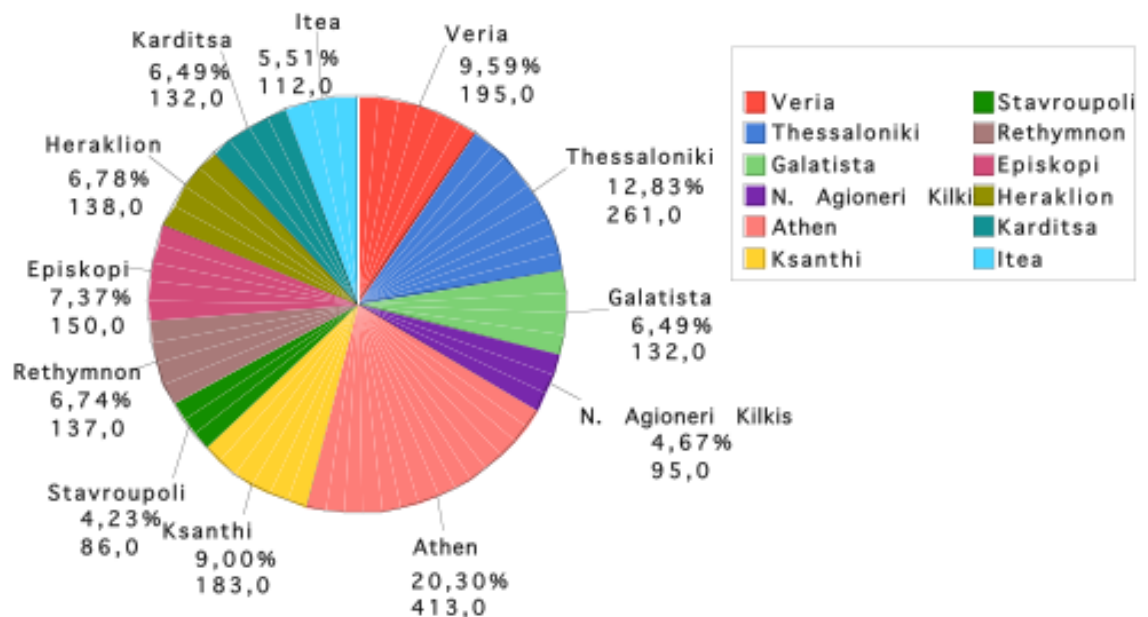


Tabelle 5.4: Geschlechtsspezifische Verteilung der untersuchten Stichprobe auf Orte unterschiedlicher Größe

Wohnort		Jungen	Mädchen	Gesamt
Veria	N	95	100	195
	%	48,7	51,3	100,0
Thessaloniki	N	121	140	261
	%	46,4	53,6	100,0
Galatista	N	74	58	132
	%	56,1	43,9	100,0
N. Agioneri Kilkis	N	57	38	95
	%	60,0	40,0	100,0
Athen	N	228	185	413
	%	55,2	44,8	100,0
Ksanthi	N	97	86	183
	%	53,0	47,0	100,0
Stavroupoli	N	47	39	86
	%	54,7	45,3	100,0
Rethymnon	N	64	73	137
	%	46,7	53,3	100,0
Episkopi	N	80	70	150
	%	53,3	46,7	100,0
Heraklion	N	70	68	138
	%	50,7	49,3	100,0
Karditsa	N	66	66	132
	%	50,0	50,0	100,0
Itea	N	57	55	112
	%	50,9	49,1	100,0
Gesamt	N	1056	978	2034
	%	51,9	48,1	100,0

Die zuvor gegebenen Zahlen beschreiben die untersuchte Stichprobe im Hinblick auf Merkmale, die schon bei der Planung der durchzuführenden Studie ausgewählt wurden, weil sie laut der aktuellen theoretischen und empirischen Wissenslage relevant für das Auftreten depressiver Symptome sein könnten. Selbstverständlich ist auch diese Stichprobe gekennzeichnet von ungezählten weiteren Merkmalen. Einige von denen wurden mit der in dieser Studie zusätzlich benutzten Fragenliste zu individuellen Lebensumstände (FIL) erfasst. Diese Fragenliste beinhaltet folgenden Items:

1. „welche der folgenden Personen wohnen mit dir zu Hause?“
2. „wie viele Geschwister hast du?“
3. „wie viele von deinen Geschwistern sind älter als du?“
4. „hast du ein eigenes Zimmer?“
5. „welchen Beruf übt deine Mutter aus?“
6. „welchen Beruf übt dein Vater aus?“
7. „welchen Schulabschluss hat deine Mutter?“
8. „welchen Schulabschluss hat dein Vater?“
9. „wie viele Stunde pro Tag arbeitet deine Mutter?“
10. „wie viele Stunde pro Tag arbeitet dein Vater?“
11. „hat jemand in deiner Familie ein ernstes gesundheitliches Problem?“

12. „*hast du ein gesundheitliches Problem?*“
13. „*Ist in der letzten Zeit deine Familie umgezogen? Wenn ja, wann?*“
14. „*Hast du in der letzten Zeit ein schlimmes Ereignis erlebt?*“
15. „*Wie oft triffst du deine Freunde?*“
16. *Was machst du gerne in deiner Freizeit?*
17. „*Welche Zensur gibst du deiner Schule?*“

Der mögliche Zusammenhang dieser Informationen mit dem selbstberichteten Auftreten depressiver Symptome wird an einer Stelle außerhalb dieser Arbeit dargestellt, deswegen werden innerhalb dieser Arbeit die Antworten, die die Kinder auf den Fragen des FIL gegeben haben, nicht vorgestellt.

5.2.2 Die Stichprobe der Lehrkräfte

Die Stichprobe der Lehrkräfte entstand aus der Realisierung der Untersuchung in 112 Klassen aus 30 griechischen Schulen und umfasst somit 112 Lehrkräfte. 67% von denen ($n = 75$) waren weiblich und 31% ($n = 35$) männlich. Von 2 Schulklassen kamen keine Fragebögen zurück (2%). Die Mehrheit der beteiligten Lehrkräfte sind Lehrerinnen und Lehrer in Grundschulen ($n = 61$) und die Resten ($n = 49$) unterrichten in einem der besuchten griechischen Gymnasien. Tabelle 5.5 zeigt die Zahl von männlichen und weiblichen Lehrkräften nach besuchten Ort. Die nachfolgende Tabelle (Tabelle 5.6) zeigt welches Fach die Lehrer aus Gymnasien am Tag der Untersuchung in der besuchten Klasse unterrichtet haben, sowie die Zahl der Kinder, für die diese Lehrkräfte jeweils die Fragen der Skala ESDM beantwortet haben.

Tabelle 5.5: Zahl der weiblichen und männlichen Lehrkräfte in Grundschulen und Gymnasien nach Untersuchungsort

	Grundschulen			Gymnasien		
	Männer	Frauen	Summe	Männer	Frauen	Summe
Veria	1	5	6	1	2	3
Thessaloniki	1	8	9	2	4	6
Galatista	1	3	4	1	2	3
Agioneri	1	3	4	3	-	3
Athina	4	8	12	2	7	9
Ksanthi	1	3	4	1	2	9
Stavroupoli	2	2	4	2	1	3
Rethymno	1	1	2	2	-	2
Episkopi	3	1	4	-	5	5
Heraklion	3	1	4	-	3	3
Karditsa	1	3	4	1	2	3
Itea	1	3	4	1	5	6
Gesamt	20	41	61	16	33	49

Tabelle 5.6: Zahl der in Grundschulen und Gymnasien arbeiteten Lehrkräfte, Ihr Unterrichtsfach am Tag der Untersuchung und die gesamte Zahl ihrer Schüler

	N der Lehrkräfte	N ihrer Schüler	Prozentsatz Schüler
Lehrkräfte in Gymnasien	49		
• Philologie	30	609	29,9
• Mathematik	6	120	5,9
• Soziologie	1	44	2,2
• Physik, Chemie	2	49	2,4
• Religion	1	21	1,0
• Fremde Sprache	6	88	4,3
• Geographie	1	29	1,4
• Kunst	2	48	2,4
Lehrkräfte in Grundschulen	61	981	48,2
Missing	2	45	2,2
Gesamt	112	2034	100,0

Die Stichprobe der Lehrkräfte ist dann im Weiteren mittelbar durch die Merkmale gekennzeichnet, die aus der Stichprobe der Schülerinnen und Schüler auf in Interaktion mit diesen Schülerinnen und Schülern stehen. die Lehrkräfte als individuell charakteristisch übertragen werden können. Das sind Angaben zur Schule, zur Klasse, zur Klassenfrequenz, zum Schulort und weiterhin auch alle, die über die Zusammensetzung der von ihnen unterrichteten Klasse vorliegen. Das sind alle Daten über die von ihnen betreuten Schulkinder. Die Lehrkräfte werden als Persönlichkeiten bekannt, die in unterrichtlicher Interaktion mit speziellen Schulkindern stehen. Schließlich kennzeichnen auch die von der jeweiligen Lehrkraft gegen ESDM Entscheidungen die Persönlichkeit. Das trifft dann auch auf die Relationen zwischen Selbstangaben der Jungen und Mädchen einerseits und den Fremdeinschätzungen durch die Lehrkraft zu. Jede Lehrkraft gibt auf diese Weise spezifische Belege für die Ausprägung der eigenen Empathie in dem Wahrnehmungsbereich, den die eingesetzten Messinstrumente ansprechen. Ein Teil dieser Informationen, insbesondere die letztgenannten, sollen in weiterführenden Analysen erneut bearbeitet werden. Eine der dann auf ihren Erklärungswert zu prüfenden Annahmen könnte sein, dass Zusammenhänge zwischen dem Grad der Empathie von Lehrkräften und den Ausprägungen depressiver Störungen unter den von ihnen unterrichteten Schulkindern bestehen.

Kapitel 6

Die Messqualitäten der in dieser Untersuchung verwendeten Untersuchungsinstrumente

Die nach der vorangegangenen Planung durchgeführte Datenaufnahme umfasst in ihrem empirischen Teil zunächst die Anwendung der ausgewählten fünf Messverfahren. Auf die mit deren Einsatz jeweils verfolgten unterschiedlichen Zielsetzungen wird in den nachfolgenden Kapiteln einzugehen sein, wenn das jeweilige Messinstrument und die mit seinem Einsatz gewonnenen speziellen Informationen darzustellen sind. Gemeinsam ist allen eingesetzten Verfahren, dass sie als Messinstrumente bezeichnet werden. Damit sind sie allesamt Instrumente, die dem jeweils intendierten Messvorgang dienen und nützen sollen.

Der Vorgang des Messens darf zum einen als alltäglicher Vorgang verstanden werden. Als solcher umfasst er in der Regel die Anwendung von Messinstrumenten und die Nutzung der Messwerte, zu welchem Zweck auch immer. Insoweit ist der Begriff des Messens eng mit speziellen Tätigkeiten und Instrumenten verbunden und in alltägliche Handlungszusammenhänge eingebunden. Die alltägliche Selbstverständlichkeit des Umgangs mit Messinstrumenten ruht aber, wie so vieles andere auch, auf historischen Entwicklungen von Denkmodellen und Techniken, die von der Urzeit bis in die Gegenwart die gesamte Messtechnik, in allen ihren Teilbereichen, zu Spitzenleistungen voran gebracht haben. Und dennoch werden in der Gegenwart neben hochtechnisierten und nur von Spezialisten auszuführenden Messungen, auch noch tradierte Verfahren eingesetzt, die Jahrtausende unverändert überstanden haben könnten. Die Entwicklung der Qualität von Messungen diente sehr wahrscheinlich immer der qualitativ gesteigerten Bewältigung von Aufgaben und damit menschlichen Tätigkeiten und Zielsetzungen. Beispielsweise verlangte selbst die etwa angestrebte Einigung über die Nutzung bestimmter Areale durch Nutzungsberechtigte zumindest elementare Messprozesse. Das setzte die Bestimmung der Form, Größe und Lage von Grundflächen voraus und regte die Entwicklung von Vermessungstechniken an. Inzwischen ist das Messen in den unterschiedlichen Anwendungsbereichen längst weit über die Sicherung einheitlicher Mess- und Maßstäbe hinaus und mit technischen Problemen befasst, deren Bewältigung Spezialisten erfordert. Die inzwischen verfolgten Qualitätsforderungen haben sich inhaltlich gewandelt. Geblieben sind aber, weil mit der Tätigkeit des Messens konstitutiv verbunden, elementare Grundforderungen wie etwa die nach der gesicherten Konstanz von definierten Qualitäten verwendeter Messinstrumente. So ist auch das Eichwesen aus der erlebten Notwendigkeit hervorgegangen, die Stabilität von Messungen durch die Konstanthaltung von Messinstrumenten zu sichern.

Die in der hier darzustellenden und zu erörternden Untersuchung verwendeten Messinstrumente sind, mit Ausnahme einer kurzen, nur in dieser Untersuchung verwendeten Fragenliste, alle aus der psychologischen Literatur bekannt. Die Prüfung der Messqualitäten eben dieser vier Instrumente ist als spezielle Aufgabe bereits mit deren Konstruktion verbunden gewesen. Aber selbst wenn dies respektvoll zur Kenntnis zu nehmen ist, besteht keine Gewähr dafür, dass die einmal bekannt gewordenen Messqualitäten sich in einer aktuell durchgeführten Untersuchung, so wie eben in der hier zu besprechenden, erneut in gewünschter Weise einstellen werden. Zur Tätigkeit des Messens gehört die fortlaufende, sich immer wiederholende Überprüfung der verwendeten Messinstrumente. Zu sichern ist, dass die für einen speziellen Messvorgang benötigten Qualitäten, die mit der Konstruktion eines verlässlichen Instrumentes auch erstmals erreicht worden waren, sich nachweislich für jede Neuanwendung reproduzieren lassen. Die niemals vorab auszuschließende Möglichkeit einer Veränderung des für stabil gehaltenen Messvorganges durch unbemerkte Veränderungen an den Messinstrumenten oder den von ihnen

benutzten Eigenheiten der Messwelt, zwingt, wenn den Messungen vertraut werden soll, zu deren fortlaufender Kontrolle, soweit wirklich Interesse am Erhalt der Qualität von Messungen besteht. In der Psychologie ist die Sicherung von Konstanz in den Qualitäten der verwendeten Messinstrumente im Vergleich mit anderen Anwendungsbereichen, allein schon wegen ihres besonderen Gegenstandes, mit speziellen Problemen verbunden. Die für Messvorgänge grundlegende Sicherung der Konstanz der Qualitäten von Messinstrumenten erfordert darauf zu beziehende, also möglichst spezifische Verfahrensweisen, folgt dabei aber den allgemeinen messtheoretischen Zielsetzungen.

Die eingesetzten Verfahren DTK, CDI, ESDM und SDQ sind international bekannt und dementsprechend häufig auf Messqualität geprüft worden. Die Kenntnisnahme dieser Messqualitäten ist die Aufgabe einer ersten Phase der Kontrolle der Messqualitäten dieser Verfahren.

Die eingesetzten Verfahren DTK, CDI, ESDM und SDQ sind aber auch alle für die Anwendung in Griechenland sprachlich verändert worden. Die Annahme, dass diese Veränderung nur eine Transposition von Inhalten ins Griechische sei und die zuvor bekannten Messqualitäten der Instrumente nicht verändert habe, ist empirisch ungeprüft. Die empirische Überprüfung der jeweils mit der Anwendung der Verfahren in dieser Untersuchung gegebenen Messqualitäten und deren Vergleich mit den bislang bekannten Werten ist die zweite Aufgabe innerhalb der aktuellen Kontrolle der Messqualitäten dieser Verfahren.

Die genannten Verfahren waren ursprünglich auch wegen der ihnen zugeschriebenen Messqualitäten zur Anwendung ausgewählt worden. Innerhalb des zu vollziehenden dritten Arbeitsschrittes in der aktuellen Kontrolle der Messqualitäten ist dann folglich zu entscheiden, ob die in der Untersuchung 2034GR erreichten Messqualitäten die erforderlichen Standards erreicht haben. Erforderlich ist das Erreichen von Standards, die empirisch begründetes Vertrauen in die mit den eingesetzten Verfahren durchgeführten Messungen rechtfertigen.

Indem der Nachweis erbracht wird, dass den Messqualitäten der eingesetzten Verfahren empirisch begründet vertraut werden darf, rechtfertigt das ein entsprechendes Vertrauen in die erhobenen Daten. Das ist noch nicht identisch mit einem Vertrauen in die zu vollziehenden Analyseschritte oder den mit ihnen gewonnenen Feststellungen, aber es rechtfertigt die danach durchzuführende Analyse des erhobenen Datensatzes.

6.1 Kriterien der Bewertung von Messqualitäten psychologischer Untersuchungsinstrumente

Theorien psychologischer Messungen sind genötigt, sich auf allgemeine Bedingungen und Gegebenheiten des Messens zu beziehen. Bortz und Döring (2002) sparen in ihren Darstellungen über Forschungsmethoden und Evaluation für Human- und Sozialwissenschaften die Befassung mit allgemeinen Aspekten des Messens aus. Stattdessen gehen sie jeweils gesondert auf messtheoretische Probleme im Zusammenhang mit den von ihnen erörterten Methoden ein. Dies geschieht beispielsweise unter der Überschrift zu den quantitativen Methoden der Datenerhebung (ebd., S. 137 uf.) in den einschlägigen Ausführungen zum Zählen, zur quantitativen Inhaltsanalyse, zum Urteilen, zum Testen, zum Befragen, zur Beobachtung und zu physiologischen Messungen. Die Verfahren DTK, CDI, ESDM und SDQ könnten unterschiedlichen Methoden zugeordnet werden, dürften aber auch als Tests bezeichnet werden. Dann müssen alle die messtechnischen Probleme für sie beachtet werden, die in den

Darstellungen zur Klassischen Testtheorie und zur Item-Response-Theorie (IRT) (ebd. S. 192 - 237) aufgeführt werden (s. auch Lienert und Raatz, 1994).

Zu den grundlegenden Inhalten der Klassischen Testtheorie gehört zunächst deren Axiomatik. In fünf Axiomen sind Grundzüge der Klassischen Testtheorie entwickelt. (Bortz und Döring, 2002, S. 193). Die hat sich nach Bortz und Döring, obschon empirisch nicht überprüfbar, anscheinend bewährt. Die Messqualitäten von Tests werden, der Klassischen Testtheorie folgend, im Wesentlichen durch drei Testgütekriterien repräsentiert. Das sind Objektivität, Reliabilität und Validität. Indem Tests, die diesen Gütekriterien genügen, als psychometrische Tests bezeichnet werden (s. Bortz und Döring, ebd., S. 194), scheinen sie grundlegende psychometrische Eigenschaften zu repräsentieren. Alle drei Testgütekriterien stehen in vielfachen Zusammenhängen mit Konzepten, Methoden und Strategien, die zu ihrer Optimierung beitragen können, selbst zwar keine Gütekriterien darstellen, deren Sicherung aber unterstützen. Ein Beispiel dafür ist die Multitrait-Multimethod-Methode (MTMM; Cambell and Fiske, 1959; Sullivan und Felman, 1979, S.17 ff.). Die Konstruktvalidität als eine spezielle Messqualität kann durch MTMM potentiell erhöht und gesichert werden. Die Vorgehensweise hilft beispielsweise konvergente und diskriminierende Validität nachzuweisen oder abzulehnen. (Bortz und Döring, 2002, S. 202 – 203) Abweichend von der Klassischen Testtheorie, die jedes Testergebnis als messfehlerbehaftet versteht, betrachtet die IRT jedes Testergebnis als Indikator für die jeweils vorhandenen Verhaltensdispositionen beziehungsweise als Indikator für ein latentes Merkmal. Nach der IRT kommt vor allem, damit anders als nach der Klassischen Theorie, den Itemcharakteristiken eine größere Bedeutung zu. Solche und andere Vergleiche unterschiedlicher Messmodelle oder Testtheorien verweisen allgemein darauf, dass die Beurteilung von Messqualitäten modellabhängig ist. Unterschiedliche Messmodelle nutzen unterschiedliche Kriterien zur Feststellung von Messqualitäten. Ein wesentlicher Unterschied ist nach Bortz und Döring (ebd. S. 211) darin zu sehen, dass probabilistisch konstruierte Tests, anders als klassisch begründete, auch jeweils in den ihnen zugrunde liegenden Modellannahmen geprüft werden. Nach der klassischen Testtheorie konstruierte Verfahren bleiben in ihren Grundlagen immer ungeprüft. Diese Aussage wiederholt das, was bei der Erwähnung der Axiome der klassischen Testtheorie über die Möglichkeiten zu deren Überprüfung festzustellen war. Sie sind im Gegensatz zu den probabilistischen Ansätzen nicht prüfbar.

Das exemplarische Eingehen auf die Bedeutung der zur Rechtfertigung von Testkonstruktionen herangezogenen Modellannahmen für die Einsetzung von Kriterien für Messqualitäten verweist auf die Notwendigkeit einer eingehenderen Befassung mit diesen Modellen. Dem wird hier an dieser Stelle aber nicht gefolgt und dies dann mit zwei Begründungen. Zum einen wird Bortz und Döring (ebd., S. 211 ff.) vertraut, die auch unter Hinweis auf Fan (1998) und Rost (1999) als gegeben ansehen, dass Tests auf probabilistischen Grundlagen messtechnisch nicht leistungsfähiger sind als die auf klassischen Testtheorien beruhenden Konstruktionen. Diese Überlegung ist im hier gegebenen Zusammenhang von Bedeutung, weil die auszuwählenden Messinstrumente für die Untersuchung 2034GR erst noch zu entwickeln waren. Zum anderen wird der Bezug zur Klassischen Testtheorie unerlässlich, weil die in den Originalsprachen existierenden Urformen der ausgewählten Messverfahren unter Verwendung klassischer Kriterien für Messqualitäten entwickelt worden sind. Ein drittes Argument wäre, dass in den USA bis zum Jahr 2001 die probabilistischen Messmodelle in der Psychologie kaum Beachtung gefunden haben. Das spricht zwar weder gegen noch für die Nützlichkeit dieser Ansätze, eine Inanspruchnahme probabilistischer Messmodelle würde aber das Erreichen der erwünschten internationalen Vergleichbarkeit der griechischen Untersuchung erheblich behindern oder gar unmöglich werden lassen.

Nach all dem sind Objektivität, Reliabilität und Validität der ausgewählten Messverfahren als Kriterien für deren Messqualitäten einzusetzen. Da andererseits die Qualität von Tests aber auch von der Qualität der verwendeten Items abhängig ist (Bortz und Döring, S. 217) sollen auch Itemanalysen zur Begutachtung der Messqualitäten der Instrumente herangezogen werden. Da weiterhin auch die Dimensionalität von Tests von messtechnischer Bedeutung ist, empfiehlt sich allein schon deshalb auch eine entsprechend anzusetzende Faktorenanalyse des Verfahrens. Diese Verfahrenspraxis findet sich beispielsweise auch in der bisherigen Anwendungsgeschichte von DTK und CDI.

Die Auswahl der in diesen Erörterungen beachteten Kriterien für die Messqualität der eingesetzten Verfahren wird unter Kenntnisnahme der mit den Ausgangsformen der Verfahren selbst eingeführten theoretischen Bindungen und dann unter Rückgriff auf die Klassische Testtheorie vorgenommen. Diese Entscheidung folgt im Übrigen entsprechenden Trends in der internationalen Diskussion von Testtheorien. Die probabilistischen Ansätze haben bislang international weniger Beachtung gefunden, als zunächst erwartet worden war.

Die nun nachstehend zu gebende Übersicht zu den Messqualitäten der eingesetzten Verfahren folgt zwei Ordnungsprinzipien. Erstens wird auf die zu erörternden Qualitätsmerkmale nach Verfahren getrennt eingegangen. Zweitens sind die Ausführungen dann auch auf die weiter oben genannten drei Arbeitsschritte bezogen. Dementsprechend werden zunächst Messqualitäten aus vorangegangenen Untersuchungen und nachfolgend die aus der griechischen Untersuchung 2034GR mitgeteilt und besprochen. Eine vergleichende Bewertung der Vertrauenswürdigkeit aller mit dem jeweiligen Verfahren aufgenommenen Daten schließt die Erörterung der Messqualitäten ab.

6.2 Die Messqualitäten des DTK

Über den von Rossmann entwickelten Depressionstest für Kinder (DTK, 1993) sind sowohl aus seiner Eichung als auch aus unterschiedlichen Anwendungen die seine Messqualitäten kennzeichnenden Werte bekannt geworden. Da zwischen den entsprechenden Daten aus der Eichstichprobe und späteren Anwendungen keine gravierenden Diskrepanzen berichtet wurden, beziehen sich die nachfolgenden Ausführungen in ihrem ersten Teil vor allem auf die Angaben zu den Gütekriterien für die Eichstichprobe.

6.2.1 Messqualitäten des DTK in Studien außerhalb der griechischen Untersuchung 2034GR

Informationen über die Messqualitäten des DTK in Anwendung außerhalb der griechischen Untersuchung 2034GR liegen aus mehreren Studien vor, insbesondere dann eben auch mit den Angaben im Testhandbuch selbst. Die Informationen betreffen nahezu alle zur Abschätzung der Messqualität relevanten Kennwerte. So finden sich vor allem Angaben zur Itemanalyse, zur Reliabilität des Verfahrens, zu seiner Faktorenstruktur und zur Validität vor. Verwiesen wird auf die Publikationen von Reicher und Rossmann (1991), Rossmann (1993), Rossmann (2005), Linsbogen (1996), Ofner, (1995), Schein (1996) und Slezak (1993).

Von den drei aus der klassischen Testtheorie (siehe Lienert und Raatz, 1994) bekannten Gütekriterien, dem der Objektivität, der Reliabilität und der Validität, sind vor allem die beiden letzteren zu beachten. Für die Objektivität liegen im Falle des DTK alle die Gegebenheiten vor,

die für ungezählte andere Testverfahren auch gelten. Die Struktur von standardisierten psychologischen Fragebögen, so auch die des DTK, bürgt für ein gewisses Maß an Objektivität. So haben alle Personen einen identischen Fragenkatalog und nehmen an relativ identischen Durchführungen teil. Von einem gewissen Maß an Objektivität muss einschränkend gesprochen werden, weil aus der entsprechenden Literatur bekannt ist (s. Lienert und Raatz, ebd.) dass eine Vielzahl denkbarer Versuchsleiter- und weiterer Situationseffekte die Vergleichbarkeit der Durchführungen einschränken kann. Den DTK trifft dieser Umstand ebenso wie alle anderen Testverfahren und dann leider auch ebenso unausweichlich, weil solche Objektivitätseinbußen auf aktuelle subjektive Determinationen der Interaktion zwischen Teilnehmern, Testleitern und Messinstrumenten zurück zu führen sind.

Die Erörterung der Messqualitäten hat sich insbesondere den Fragen nach Reliabilität und Validität zuzuwenden.

In den zur Kenntnis kommenden Erfahrungen aus der Anwendung des DTK liegen Die Reliabilitätswerte für die Skala 1 „Dysphorie/Selbstwertprobleme“ insgesamt zwischen 0,86 bis 0,89, die der Skala „Agitiertes Verhalten“ zwischen 0,76 und 0,82 und die Reliabilitätskoeffizient für die dritte DTK-Skala „Müdigkeit/autonome Reaktionen“ liegen zwischen 0,73 und 0,81. Zu wiederholen ist die Information, dass die drei DTK-Skalen aus einer jeweils unterschiedlicher Anzahl von Items bestehen (25, 16 und 14 Items), was helfen kann, die beobachtete Abnahme der Alpha-Werte von der ersten bis zur dritten Skala zumindest aus statistischer Sicht zu erklären. Da diese Abnahme in allen heranziehbaren Studien auftritt, ist nicht von einem stichprobenabhängigen, sondern eher von einem systematischen oder verfahrenstypischen Effekt auszugehen.

Angaben über die Retest-Reliabilität des Verfahrens nach unterschiedlichen Zeitintervallen, nämlich nach 1, 3, 5 und 8 Wochen, wurden sowohl aus einer unveröffentlichten Arbeit von Rossmann als auch aus den Studien von Gratz (1994), Renner (1993) und Slezak (1993) bekannt. Gemeinsam ist diesen Studien das Auftreten des so genannten „Attenuationseffektes“. Das ist die Bezeichnung für die Tendenz zur Abnahme der Alpha-Werte bei einer Messwiederholung der Untersuchung. Diese Tendenzen wurde beispielsweise auch bei der wiederholenden Anwendung sowohl des deutschen als auch des amerikanischen CDI bekannt (siehe dazu Rossmann, 2005).

Im Hinblick auf die Beziehungen zwischen DTK-Scores und dem Alter und dem Geschlecht der an den unterschiedlichen Studien beteiligten Kinder und Jugendlichen scheinen die Befunde weitgehend homogen mit den entsprechenden Daten aus der Normstichprobe des Verfahrens zu sein. Das betrifft sowohl die beobachteten Geschlechtseffekte in der ersten Skala, wo die Mädchen signifikant höhere Mittelwerte als die Jungen aufweisen, als auch die entsprechenden Werte aus der Durchführung der zweiten Skala, wo dann die Knaben signifikant höhere Werte als ihre Mitschülerinnen erreichen. Diese Tendenz wiederholt sich mit steigendem Alter der Schüler und Schülerinnen, aber jeweils auf einem höheren Punkteniveau. Für beide Skalen liegen die von den älteren Kindern erreichten DTK-Werte höher als die der jüngeren. Für die dritte DTK-Skala vermerken die meisten DTK-Studien ein Ausbleiben signifikanter Effekte im Hinblick auf Geschlecht und Alter der jeweils untersuchten Kinder. Die über unterschiedliche Untersuchungen als weitgehend stabil bekannt werdenden Geschlechtseffekte in den DTK-Skalen stellen nach Äußerungen des DTK Autors eine empirische Bestätigung für die Nützlichkeit des im DTK verwendeten Konstruktes der depressiven Verstimmungen bei Kindern und Jugendlichen dar, dessen Besonderheit die Verwendung dreier inhaltlich unterschiedlicher Subskalen ist. „...Die höheren Durchschnittswerte der Mädchen in der Skala, die Dysphorie/Selbstwertprobleme erfasst und die höheren Mittelwerte der Knaben in der Skala, welche agitierte Verhaltensweisen erfasst, stehen absolut in Übereinstimmung mit theoretischen Erwartungen. Der DTK ermöglicht

damit die getrennte Darstellung voneinander empirisch unterscheidbarer Aspekte der depressiven Selbstbeschreibung von Kindern, die weit über eine bloße Bestimmung eines allgemeinen Summenscores hinausgeht“ (Rossmann, 1993, S. 12). Weitere Informationen über Geschlechts- und Altersunterschiede in den Summenwerten auf den drei DTK-Skalen werden innerhalb der Ausführungen über die Ergebnisse der griechischen Untersuchung 2034GR im Kapitel 7 dieser Arbeit gegeben.

Im Hinblick auf Validitätsaspekte des Verfahrens DTK liegen Daten über die faktorielle Validität, die Kriteriumsvalidität und die diskriminante Validität des DTK vor.

Nach einem Vergleich der Ergebnisse aus einer Stichprobe von 1175 Kindern im Alter zwischen 9 und 15 Jahren und den Daten aus der Normstichprobe (2507 Kinder im gleichen Alter) ergab sich eine fast punktgenaue Übereinstimmung und somit eine fast perfekte Replizierbarkeit der ursprünglich identifizierten Faktoren. Das gilt auch noch, wenn die Daten getrennt nach Knaben und Mädchen analysiert werden (Rossmann, 1993). Der Autor stuft unter Hinweis auf diese Ergebnisse die faktorielle Struktur des DTK als stabil ein.

Angaben über die kriterienbezogene Validität des DTK liegen aus einer Reihe von Untersuchungen vor, die den zu erwartenden Beziehungen zwischen DTK-Werte und den Ergebnissen aus der Anwendung anderer Verfahren geprüft haben. Im Hinblick auf depressionsrelevante Messinstrumente ergab sich eine signifikante Korrelation zwischen dem DTK und dem CDI (Reicher und Rossman, 1991), zwischen dem DTK und der deutschen Fassung der „Youth Self-Report Form“ des „Child Behavior Checklist“ von Achenbach und Edelbrock (1987; deutsche Fassung von Lösel und Mitarbeitern, 1988). Diese Beziehung bestätigte sich mit einer Studie von Schein (1996). Zusammenhänge zwischen dem DTK und dem depressionsspezifischen Bildverfahren TONI wies Buchmann (2000) nach. Theoriekonforme signifikante Beziehungen wurden auch zu Messungen von Neurotizismus und Extraversion nachgewiesen (Janisch, 1994), zu depressionsspezifischen Attributionsstilen (Engel, 1994; Gratz, 1994; Reitbauer, 1999), zu Bewältigung von negativen Lebensereignissen und Separationserlebnissen (Acher mann, 1995; Fuchs, 1995), zu Stressbelastung und Belastungsverarbeitung (Acher mann, 1995; Hafner, 2001), zu Selbstwertgefühl und Selbstkonzept (Renner, 1993; Sitzenfrey, 1994), zu selbst- und fremdbeurteilter sozialer Kompetenz (Schein, 1996) sowie zur Integration in der Schulklasse (Linsbogen, 1996). Unter den gleichzeitig mit der Durchführung des DTK untersuchten Variablen nehmen die Schulnoten von Schülerinnen und Schülern insoweit eine Sonderstellung ein, als diese Variable als einzige nicht signifikant mit DTK-Werten korrelierte (Acher mann, 1995; Engel, 1994; Sitzenfrey, 1994). In der zweiten Auflage des DTK-Manuals (Rossmann, 2005) sind detaillierte Angaben über einige der oben genannten Studien zu finden, wie auch weitere Literaturangaben über DTK-Studien.

Von besonderer Bedeutung für die hier darzustellende Untersuchung 2034GR sind Angaben zur diskriminanten Validität des DTK. Obschon Angaben über die diskriminative Leistungsfähigkeit des DTK bei der Trennung von klinisch und nicht klinisch auffälligen Kindern mit depressiven Symptomen nur aus zwei Studien vorliegen, belegen deren Ergebnisse gute diskriminative Leistungen. Die Ergebnisse verweisen vor allem auf die Nützlichkeit des DTK als Screening-Instrument in Normalpopulationen. Auf diese zwei Studien (Zimmermann, in Rossmann, 1993; Pötz, 2003) wird ausführlich an anderer Stelle dieser Arbeit einzugehen sein (Kapitel 7).

Der mit den vorangegangenen Ausführungen angebotene Überblick über die Messqualitäten des DTK in deutschsprachigen Untersuchungen verschafft zum einen die Möglichkeit, sie verlässlich abzuschätzen. Der DTK erweist sich als ein reliables und valides Messinstrument, dessen

Messqualitäten seine Auswahl für eine Anwendung in Griechenland unterstützen. Zum anderen ergeben sich aus der Übersicht nützliche Vergleichsstandards für die Anwendung in der griechischen Untersuchung 2034GR. Der Nachweis von Messqualitäten bei einer deutschsprachigen Anwendung lässt entsprechende Qualitäten für eine griechische Anwendung erhoffen. Ob diese Hoffnungen sich erfüllen, kann nur über einen Vergleich der Messqualitäten aus der deutschsprachigen und einer anderssprachigen Anwendung empirisch verlässlich entschieden werden. Erst damit würde sich die Auswahl des DTK für eine griechische Untersuchung rechtfertigen.

Bis dahin bliebe die Entscheidung für seine Anwendung nur eine Hypothese über seine Nützlichkeit innerhalb einer Untersuchung, die um eine verlässliche Erfassung des Auftretens depressiver Symptome bei Kindern und Jugendlichen in Griechenland bemüht ist. Die verlässliche Erfassung setzte das Erreichen von Messqualitäten bei einer griechischen Anwendung des DTK voraus, die denen aus dem deutschsprachigen Einsatz gleichen. Ob dies erreicht werden konnte, bleibt nachfolgend zu prüfen.

6.2.2 Messqualitäten des DTK bei seiner griechischen Anwendung in der Untersuchung 2034GR

Da die Testgütekriterien letztlich alle auf der Anwendung der Items in den Verfahren beruhen, ist auch das an erster Stelle stehende Eingehen auf die für die Items bekannt werdenden Angaben zu ihren Messqualitäten sinnvoll.

6.2.2.1 Itemanalyse

Für die Itemanalyse sind nach der Klassischen Testtheorie (siehe Bortz und Döring, 2002; Lienert und Raats, 1994) mehrere Kriterien eingeführt worden, zu denen in der Regel der Schwierigkeitsgrad und die Trennschärfe gehören. Beide Kennwerte sind in den hier geführten Analysen neben anderen berechnet und beachtet worden. Um den Darstellungsaufwand bei der Wiedergabe dieser Werte zu Gunsten der Überschaubarkeit der Tabellen reduzieren zu können, wird die Auflistung der Itemschwierigkeiten in einen Tabellenanhang gegeben.

Der DTK umfasst 55 Items, die auf drei Subskalen verteilt sind. Die Skala 1 „Dysphorie/Selbstwertprobleme“ umfasst 25 Items, die Skala 2 „Agitiertes Verhalten“ 16 Items und die Skala 3 „Müdigkeit und Autonome Reaktionen“ weitere 14 Items.

Für jedes einzelne der 55 Items des DTK wurden der Mittelwert, die Standardabweichung und die Varianz berechnet. Darüber hinaus finden sich nachfolgend, um Informationen über die Beziehungen zwischen den einzelnen Items und dem Summenwert für jede der drei DTK-Subskalen als deren kennzeichnendes Gesamtergebnis bekannt werden zu lassen, auch die Korrelationswerte zwischen dem Gesamtscore jeder Subskala und ihren eigenen Items. Alle diese deskriptive Informationen – Kennwerte und Trennschärfe - der DTK-Items sind in den Tabellen T1, T2 und T3 für jede DTK-Subskala getrennt dargestellt.

Da sich aus der Summe der ersten zwei DTK-Skalen auch ein Gesamtergebnis (DTK-Skala „Depressive Belastung“) ergibt, sind auch die Korrelationen der Items, die zu diesen beiden Skalen gehören, mit der Gesamtscore der Skala „Depressive Belastung“ berechnet (siehe letzte Spalten in Tabellen 6.1 und 6.2).

Tabelle 6.1: Kennwerte und Trennschärfe der 25 Items der DTK-Subskala „Dysphorie/Selbstwertprobleme“

DTK-Item	N	Mittelwert	Standardabweichung	Varianz	Trennschärfe ¹ (auf Skala 1)	Trennschärfe ¹ (auf Skala 1+2)
DTK 33	2030	,75	,433	,188	,38	,29
DTK 50	2031	,64	,479	,229	,38	,34
DTK 4	2032	,43	,496	,246	,51	,46
DTK 51	2031	,41	,493	,243	,39	,39
DTK 44	2029	,38	,486	,237	,34	,32
DTK 20	2031	,38	,486	,236	,51	,44
DTK 49	2032	,38	,485	,236	,40	,37
DTK 14	2031	,36	,481	,232	,43	,39
DTK 23	2029	,34	,475	,226	,38	,33
DTK 55	2031	,32	,468	,219	,53	,47
DTK 13	2030	,31	,461	,213	,38	,34
DTK 31	2030	,30	,457	,209	,47	,41
DTK 43	2030	,29	,455	,207	,54	,50
DTK 48	2031	,28	,448	,200	,58	,52
DTK 24	2030	,27	,443	,196	,43	,50
DTK 36	2028	,24	,426	,182	,50	,54
DTK 19	2029	,23	,422	,178	,47	,41
DTK 34	2030	,22	,417	,174	,60	,53
DTK 21	2032	,22	,416	,173	,45	,39
DTK 52	2031	,22	,412	,170	,53	,49
DTK 42	2031	,22	,412	,170	,55	,49
DTK 54	2031	,16	,365	,133	,46	,42
DTK 35	2031	,16	,362	,131	,55	,49
DTK 18	2028	,15	,361	,130	,42	,40
DTK 40	2031	,08	,275	,076	,30	,27
Valid N	2016					

Trennschärfe¹: alle Trennschärfe sind signifikant auf $p = 0,000$

Insgesamt erbrachte die Itemanalyse für die DTK-Skalen 1 bis 3 Werte, die als in hohem Maße zufrieden stellend eingestuft werden dürfen, weil sie in dem von der klassischen Testtheorie bevorzugten Wertebereich zwischen 0,20 und 0,80 liegen. Die Itemmittelwerte weisen eine recht breite Streuung der Schwierigkeit nach.

Unter den zur ersten DTK-Skala („Dysphorie/Selbstwertprobleme“) gehörenden Items (vgl. Tabelle 6.1), liegen einige Mittelwerte sehr niedrig ($p \leq 0,20$). Es handelt sich um Item 40 („*Hast Du mehr Schmerzen als andere Kinder?*“) und auch um Item 18 („*Denkst Du oft, dass Du ein schlechter Mensch bist?*“), Item 35 („*Bist Du oft unglücklich?*“) und Item 54 („*Hast Du mehr Sorgen als andere Kinder?*“). Möglicherweise haben diese Fragen inhaltliche und formale Besonderheiten, die mit dem Erscheinen dieser niedrigen Kennwerte in Verbindung stehen. Zwei Fragen verlangen den Vergleich des befragten Individuums mit einer kaum überschaubaren Population „anderer Kinder“. Das rechtfertigt das Einbringen einer entsprechenden, an dieser Stelle aber nicht zu beantwortenden, Nachfrage: „Führen Fragen, die zu ihrer Beantwortung formal (eigentlich) den spezifischen Vergleich des befragten Kindes mit einer unüberschaubaren

Population anderer Kinder verlangen, zu relativen Einbussen an Kennwerten in der Itemanalyse?
(ZF1)

**Tabelle 6.2: Kennwerte und Trennschärfe der 16 Items der DTK-Subskala
„Agitiertes Verhalten“**

DTK-Item	N	Mittelwert	Standardabweichung	Varianz	Trennschärfe ¹ (auf Skala 2)	Trennschärfe ¹ (auf Skala 1+2)
DTK 32	2029	,49	,500	,250	,46	,38
DTK 12	2030	,45	,498	,248	,45	,40
DTK 3	2032	,44	,497	,247	,51	,41
DTK 37	2031	,38	,486	,236	,41	,35
DTK 5	2032	,35	,478	,229	,54	,41
DTK 6	2028	,31	,464	,215	,37	,41
DTK 29	2029	,30	,459	,210	,55	,39
DTK 30	2029	,29	,455	,207	,52	,33
DTK 11	2030	,29	,454	,206	,37	,31
DTK 16	2031	,20	,401	,161	,48	,31
DTK 27	2028	,18	,387	,150	,48	,39
DTK 39	2031	,18	,385	,148	,44	,34
DTK 17	2030	,15	,353	,125	,47	,39
DTK 15	2031	,13	,340	,115	,47	,27
DTK 41	2031	,13	,339	,115	,36	,25
DTK 2	2032	,13	,334	,112	,43	,33
Valid N	2016					

Trennschärfe¹: alle Trennschärfe sind signifikant auf p = 0,000

**Tabelle 6.3: Kennwerte und Trennschärfe der 14 Items der DTK-Skala
„Müdigkeit/Autonome Reaktionen“**

DTK-Item	N	Mittelwert	Standardabweichung	Varianz	Trennschärfe ¹
DT K 8	2033	,62	,485	,236	,40
DTK 45	2031	,60	,490	,240	,49
DTK 46	2031	,60	,490	,240	,45
DTK 10	2031	,43	,495	,245	,49
DTK 7	2030	,41	,493	,243	,44
DTK 47	2030	,41	,491	,241	,43
DTK 1	2032	,39	,489	,239	,48
DTK 22	2032	,31	,464	,216	,46
DTK 25	2028	,25	,433	,188	,49
DTK 9	2031	,23	,418	,175	,44
DTK 38	2031	,22	,412	,170	,44
DTK 26	2031	,20	,403	,162	,40
DTK 28	2030	,18	,386	,149	,50
DTK 53	2032	,17	,379	,143	,33
Valid N	2024				

Trennschärfe¹: alle Trennschärfe sind signifikant auf p = 0.000

Hinsichtlich der zur zweiten DTK-Skala („Agitiertes Verhalten“) gehörenden Items erscheinen einige Fragen aufgrund ihrer niedrigen Mittelwerte als schlechte Informationsträger (vgl. Tabelle 6.2). Das sind die Items 2 („*Macht Dir die Schule Spaß?*“), 41 („*Spielst Du oft in der Klasse den Clown?*“), 15 („*Gibst Du manchmal in der Schule freche Antworten?*“), 17 („*Bist Du zu Hause oft unfreundlich?*“), 39 („*Streitest Du viel mit anderen Kindern?*“) und 27 („*Möchtest Du manchmal jemanden sehr weh tun?*“).

Für einige dieser Fragen ließen sich begründete Vermutungen dafür aufstellen, dass sie wenig unterschiedliche Informationen bieten. So ist die Frage 2 nach Spaß in der Schule durch deren nicht nur unter Schulkindern relativ hohen Bekanntheitsgrad belastet. Man dürfte sie sprachlich als „Allgemeinplatz“ klassifizieren. So werden gelegentlich Redewendungen oder Begriffe bezeichnet, deren Bedeutungsgehalt durch ihren speziellen Gebrauch gemindert ist. So nutzt gelegentlich hoch frequente Benutzung den subjektiv erlebten Bedeutungsgehalt ab. Eine Redewendung, die diese subjektive Einschätzung exemplarisch wiedergibt ist: „Davon spricht doch schon jeder.“ Dem ist ein Bedeutungshof beigegeben, der als Argument dafür dienen kann, selbst nicht auch noch auf den thematisierten Sachverhalt einzugehen. Auch Frage 15 und Frage 17 tendieren in diese Richtung. Diese und andere Mutmaßungen zu der thematisierten Problematik ließen sich empirisch prüfen. Das regt die hier vermerkte Nachfrage an: „Ist die durch häufigen Gebrauch von Redewendungen beobachtbare Abwehr einer ernsthaften Befassung mit dem jeweils gemeinten Thema beobachtbar mit Minderungen der Kennwerte in Itemanalysen verbunden?“ (ZF2).

Schließlich weisen zwei Fragen der dritten DTK-Subskala („Autonome Reaktionen/Müdigkeit“) Mittelwerte auf, die am unteren Rand des bevorzugten Wertebereichs liegen. Es handelt sich um Item 53 („*Bleibst Du am liebsten den ganzen Tag im Bett?*“) und Item 28 („*Fühlst Du Dich oft krank?*“).

Eine Hypothese, die vermuten ließe, dass diese Fragen für Erwachsene besser geeignet seien als für Kinder wäre, dass der von den Kindern mit diesen Fragen verbundene Bedeutungsgehalt wegen des allgemein höheren Bewegungsdranges bei Kindern relativ weit außerhalb der eigenen Denkmuster liegt. Das führt hierzu Formulierung einer weiteren Nachfrage: „Ist für Fragen in Fragebögen für Kinder, die inhaltlich mit hoch frequent auftretenden Verhaltenstendenzen bei Kindern kollidieren, häufig eine Reduzierung von Kennwerten in Itemanalysen zu beobachten?“ (ZF3)

Zu bemerken bliebe, dass die inhaltliche Akzentuierung der aufgeführten Zusatzfragen 1 bis 3 messtheoretisch verhältnismäßig bedeutungslos ist, weil die niedrigen Kennwerte bereits als Anlass gesehen werden könnten, diesen Items nicht zu vertrauen. Immerhin wäre die empirisch gesicherte Beantwortung von ZF 1 bis 3 als potentielle Hilfe bei der Konstruktion von Fragebögen zu sehen.

In Tabelle 6.4 sind die Häufigkeitsauszahlungen für alle Items mit einem hohen Mittelwert dargestellt.

An dieser Stelle wäre es relevant auf eine Bemerkung des Autors vom DTK (Rossmann, 1993) hinzuweisen. Bei der Itemanalyse von Daten aus nichtklinischen Stichproben sollten die dort eventuell gefundenen niedrigen Itemmittelwerte aufgrund der erwarteten leichteren Ausprägungen der abgefragten Symptome bei einer Mehrheit der Personen in der Stichprobe nicht unbedingt als negativ bewertet werden. Dazu schreibt Rossmann bei der Darstellung der Daten aus seiner Normstichprobe folgendes: „Bei der vorliegenden Art der Skalenkonstruktion wurde bewusst relativ wenig Rücksicht auf die Itemschwierigkeit genommen. Grund für diese Zurückhaltung war die Tatsache, dass sich alle Items explizit auf Indikatoren depressiver Verstimmungszustände

beziehen sollten. Da die Prävalenz solcher Symptome in nichtklinischen Gruppen *per definitionem* nicht allzu hoch sein kann, durfte natürlich nicht erwartet werden, dass die relativen Lösungshäufigkeiten für die vorgegebenen Fragen in einem besonders weiten Bereich variierten. Insbesondere kann nicht angenommen werden, dass der Erwartungswert für die relativen Lösungshäufigkeiten von Items dieses Typs in der Nähe von 0,5 zu liegen käme“ (Rossmann, 1993, S.7).

Ähnliche Überlegungen werden auch von der Arbeitsgruppe von Stiensmeier-Pelster (2000) in Bezug auf die von ihnen gefundenen niedrigen Itemsmittelwerte ihrer deutschen CDI-Version ausgedrückt (siehe dazu im folgenden Kapitel 6.3.2.1., „Itemanalyse der griechischen CDI-Items“).

Die Entscheidung, ob und welche der Items mit niedrigen Mittelwerten in der griechischen Version des DTK bleiben dürfen, wird in Zusammenhang mit deren Beitrag zu dem Reliabilitätskoeffizienten der DTK-Subskalen entschieden. Einen Beitrag zur Entscheidungsfindung ist auch aus der Analyse der faktoriellen Struktur des Instruments zu erwarten.

Tabelle 6.4: Häufigkeitsauszahlungen für Items mit hohem Mittelwert

DTK 2	Frq.	Percent	Valid Percent	Cum. Percent	DTK 15	Frq.	Percent	Valid Percent	Cum. Percent
0 „Nein“	1772	87,1	87,2	87,2	0 „Nein“	1761	86,6	86,7	86,7
1 „Ja“	260	12,8	12,8	100,0	1 „Ja“	270	13,3	13,3	100,0
Total	2032	99,9	100,0		Total	2031	99,9	100,0	
Missing System	2	,1			Missing System	3	,1		
Total	2034	100,0			Total	2034	100,0		

DTK 17	Frq.	Percent	Valid Percent	Cum. Percent	DTK 18	Frq.	Percent	Valid Percent	Cum. Percent
0 „Nein“	1733	85,2	85,4	85,4	0 „Nein“	1716	84,4	84,6	84,6
1 „Ja“	297	14,6	14,6	100,0	1 „Ja“	312	15,3	15,4	100,0
Total	2030	99,8	100,0		Total	2028	99,7	100,0	
Missing System	4	,2			Missing System	6	,3		
Total	2034	100,0			Total	2034	100,0		

DTK 27	Frq.	Percent	Valid Percent	Cum. Percent	DTK 28	Frq.	Percent	Valid Percent	Cum. Percent
0 „Nein“	1657	81,5	81,7	81,7	0 „Nein“	1661	81,7	81,8	81,8
1 „Ja“	371	18,2	18,3	100,0	1 „Ja“	369	18,1	18,2	100,0
Total	2028	99,7	100,0		Total	2030	99,8	100,0	
Missing System	6	,3			Missing System	4	,2		
Total	2034	100,0			Total	2034	100,0		

DTK 35	Frq.	Percent	Valid Percent	Cum. Percent	DTK 39	Frq.	Percent	Valid Percent	Cum. Percent
0 „Nein“	1716	84,4	84,5	84,5	0 „Nein“	1663	81,8	81,9	81,9
1 „Ja“	315	15,5	15,5	100,0	1 „Ja“	368	18,1	18,1	100,0
Total	2031	99,9	100,0		Total	2031	99,9	100,0	
Missing System	3	,1			Missing System	3	,1		
Total	2034	100,0			Total	2034	100,0		

DTK 40	Frq.	Percent	Valid Percent	Cum. Percent	DTK 42	Frq.	Percent	Valid Percent	Cum. Percent
0 „Nein“	1864	91,6	91,8	91,8	0 „Nein“	1763	86,7	86,8	86,8
1 „Ja“	167	8,2	8,2	100,0	1 „Ja“	268	13,2	13,2	100,0
Total	2031	99,9	100,0		Total	2031	99,9	100,0	
Missing System	3	,1			Missing System	3	,1		
Total	2034	100,0			Total	2034	100,0		

DTK 53	Frq.	Percent	Valid Percent	Cum. Percent	DTK 54	Frq.	Percent	Valid Percent	Cum. Percent
0 „Nein“	1680	82,6	82,7	82,7	0 „Nein“	1709	84,0	84,1	84,1
1 „Ja“	352	17,3	17,3	100,0	1 „Ja“	322	15,8	15,9	100,0
Total	2032	99,9	100,0		Total	2031	99,9	100,0	
Missing System	2	,1			Missing System	3	,1		
Total	2034	100,0			Total	2034	100,0		

*** Frq. = Frequency, Cum. Percent = Cumulative Percent

Im Hinblick auf die beobachteten Trennschärfen weisen alle 55 DTK-Items mittlere ($0,3 \leq r < 0,5$) bis hohe ($r \leq 0,5$) Trennschärfen auf. Aus dieser Sicht wären alle DTK-Items im Hinblick auf die messtheoretischen Anforderungen gegenüber einer empirischen Studie (siehe Bortz und Döring, 2002, S. 192, ff.) akzeptabel und somit messtheoretisch uneingeschränkt einsetzbar.

Da die Betrachtung der nachgewiesenen Messqualitäten des DTK auf einen Vergleich zwischen den für die deutsche Version schon bekannten und den für die griechischen Fassung nun erreichten Werte abzielt, bietet die nachstehenden Tabellen (6.5) die Trennschärfen der DTK-Items aus drei Stichproben an (Kampisiou= Untersuchung 2034GR, Rossmann= deutschsprachige Untersuchungen).

Tabelle 6.5: Trennschärfe der DTK-Items in drei Stichproben

Skala 1 Items	Kampisiou (N=2034)	Rossmann (N=1175)	Rossmann (N=2507)
4	,51	,49	,36
13	,38	,44	,38
14	,43	,44	,38
18	,42	,59	,50
19	,47	,49	,40
20	,51	,43	,38
21	,45	,45	,37
23	,38	,38	,32
24	,43	,47	,38
31	,47	,48	,42
33	,38	,46	,44
34	,60	,57	,55
35	,55	,59	,54
36	,50	,49	,34
40	,30	,43	,31
42	,55	,53	,54

Skala 2 Items	Kampisiou (N=2034)	Rossmann (N=1175)	Rossmann (N=2507)
2	,43	,43	,31
3	,51	,46	,65
5	,54	,52	,59
6	,35	,46	,45
11	,45	,43	,44
12	,46	,43	,20
15	,47	,52	,22
16	,48	,35	,16
17	,47	,44	,32
27	,48	,31	,15
29	,55	,52	,27
30	,52	,54	,32
32	,46	,41	,37
37	,41	,39	,34
39	,44	,34	,19
41	,36	,31	,17

Skala 1 Items	Kampisiou (N=2034)	Rossmann (N=1175)	Rossmann (N=2507)
43	,54	,49	,45
44	,34	,37	,37
48	,58	,36	,45
49	,40	,40	,33
50	,38	,34	,37
51	,39	,44	,38
52	,53	,45	,47
54	,46	,51	,42
55	,53	,52	,51

Skala 3 Items	Kampisiou (N=2034)	Rossmann (N=1175)	Rossmann (N=2507)
1	,48	,41	,40
7	,44	,34	,17
8	,40	,38	,45
9	,44	,35	,25
10	,49	,48	,24
22	,46	,39	,23
25	,49	,47	,30
26	,40	,41	,21
28	,50	,45	,21
38	,44	,48	,17
45	,49	,46	,57
46	,45	,42	,50
47	,43	,42	,26
53	,33	,33	,30

Der Vergleich der Trennschärfen in den drei Stichproben verweist auf die relativ hohe Homogenität der Werte aus den drei unterschiedlichen Untersuchungen.

6.2.2.2 Interne Konsistenz der DTK-Subskalen

Im Folgenden wird die Reliabilität der vier DTK-Skalen überprüft (siehe Tabellen 6.6, 6.7, 6.8 und 6.9). Im Weiteren finden sich dann auch Bestimmungen zur der Höhe der Reliabilität jeder DTK-Subskala in Abhängigkeit von Geschlecht, Schulform und Schulklasse, wie weiterhin dann auch für ausländische Kinder und ältere Jugendlichen (ab 16 Jahren alt) in der untersuchten Stichprobe (Tabelle 6.10). Nach dem Konzept der Reliabilität wird ein Test im Hinblick darauf geprüft, wie gut er misst. Hohe Reliabilitätskoeffizienten verweisen auf Tests, die, unabhängig von dem, was sie eigentlich feststellen, gut messen. Wenn dann weiterhin geprüft wird, ob dieser Test auch bei der Anwendung in speziellen Stichproben gut misst, könnten Informationen darüber gewonnen werden, ob die Qualität der Messungen mit den unterschiedenen speziellen Teilnehmergruppen variiert. Sollten beispielsweise ausländische Kinder signifikant niedrigere Reliabilitätswerte als Kinder mit griechischer Muttersprache haben, ließe sich die Hypothese rechtfertigen, dass unzureichende Sprachkompetenz die Messqualität des geprüften Verfahrens beeinflusst. Solche Prüfungen gehen über das Standardprogramm der Reliabilitätsprüfungen hinaus, bleiben zunächst aber auf messtechnische Probleme beschränkt und sparen den Aspekt aus, den der jeweilige Test speziell misst, was seine Validitätswerte bezeugen können. Immerhin ließe sich auch denken, dass ein Depressionstest für Kinder durch unterschiedliche Reliabilitätskoeffizienten ausgewiesene unterschiedliche Messqualitäten bei Kindern mit niedrigen und mit hohen Depressionswerten haben könnte.

Tabelle 6.6: Reliabilität der Skala 1. „Dysphorie/Selbstwertprobleme“

Cronbach's Alpha	N of Items
,842	25

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
d4	7,33	22,867	,432	,834
d13	7,45	23,609	,299	,839
d14	7,40	23,324	,346	,838
d18	7,61	23,782	,355	,837
d19	7,53	23,362	,398	,836
d20	7,38	22,897	,437	,834
d21	7,54	23,448	,383	,836
d23	7,42	23,562	,298	,839
d24	7,49	23,444	,355	,837
d31	7,46	23,224	,393	,836
d33	7,01	23,713	,299	,839
d34	7,54	22,819	,545	,830
d35	7,61	23,309	,493	,833
d36	7,52	23,202	,433	,834
d40	7,68	24,421	,247	,840
d42	7,54	23,050	,490	,832
d43	7,47	22,930	,465	,833
d44	7,38	23,739	,251	,841
d48	7,48	22,741	,520	,831
d49	7,38	23,452	,314	,839
d50	7,12	23,558	,296	,840
d51	7,35	23,472	,303	,839
d52	7,54	23,136	,469	,833
d54	7,60	23,627	,395	,836
d55	7,44	22,890	,459	,833

Die für Cronbachs Alpha ermittelten korrigierten Werte sprechen dafür, keines der zu dieser Skala gehörigen Items zu entfernen. Das gilt auch für die Items 40, 18, 35, 54, die bei der Itemanalyse niedrige Mittelwerte erbrachten.

Auch bei der zweiten Skala sollte, entsprechend den ermittelten Reliabilitätswerten, keines der aufgeführten Items entfernt werden, auch nicht die Items mit niedrigen Mittelwerten (2, 41, 15,17, 39 und 27).

Ähnliches gilt dann weiterhin für die 14 Items, die zur dritten Skala des DTK gehören.

Auffallend ist, dass die Alpha-Werte in den Tabellen für Skala 1 fast durchgehend höher liegen als die für Skala 2.

Tabelle 6.7: Reliabilität der Skala 2 „Agitiertes Verhalten“

Cronbach's Alpha	N of Items
,751	16

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
d2	4,30	9,251	,336	,740
d3	3,98	8,680	,383	,735
d5	4,07	8,631	,423	,730
d6	4,11	9,246	,209	,751
d11	4,14	8,974	,319	,741
d12	3,98	8,838	,325	,741
d15	4,29	9,151	,381	,736
d16	4,23	8,981	,377	,736
d17	4,28	9,115	,378	,736
d27	4,24	9,022	,377	,736
d29	4,13	8,653	,439	,729
d30	4,14	8,749	,407	,732
d32	3,94	8,830	,326	,741
d37	4,05	9,038	,265	,747
d39	4,25	9,118	,337	,739
d41	4,30	9,391	,263	,745

Tabelle 6.8: Reliabilität der Skala 3 „Müdigkeit/ autonome Reaktionen“

Cronbach's Alpha	N of Items
,687	14

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
d1	4,64	6,899	,324	,668
d7	4,62	6,996	,280	,674
d8	4,41	7,106	,243	,679
d9	4,81	7,112	,306	,671
d10	4,60	6,835	,344	,665
d22	4,72	6,969	,320	,668
d25	4,78	6,966	,357	,664
d26	4,83	7,220	,272	,675
d28	4,85	7,038	,383	,662
d38	4,82	7,115	,313	,670
d45	4,43	6,854	,341	,665
d46	4,43	6,974	,292	,672
d47	4,63	7,033	,267	,676
d53	4,86	7,406	,204	,682

Wie bei der ersten und zweiten DTK-Skala führt auch die Eliminierung keines der zu der dritten Skala gehörenden Items zur Erhöhung des entsprechenden Reliabilitätswertes. Aus dieser Sicht sollten alle Items der Skala 3 beibehalten werden. Deutlich ist, dass die Alpha-Werte für die Skala 3 noch einmal unter denen für Skala 1 und dann auch denen für Skala 2 liegen. Das Absinken der Alpha-Werte wird vom Absinken der Anzahl der Items begleitet.

Die nachstehende Tabelle enthält die Werte, die nach einer Zusammenlegung von Skala 1 und Skala 2 eine neue und gemeinsame Skala bilden, die der durch den DTK ermittelte depressive Belastung. Gefragt wird, welche Reliabilitätswerte sich für diese neu gebildete Skala ermitteln ließen.

Tabelle 6.9: Reliabilität der DTK-Skala "Depressive Belastung"

	Cronbach's Alpha	N of Items
	,859	41

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
DTK 2	12,06	46,924	,285	,857
DTK 3	11,75	45,773	,345	,856
DTK 4	11,76	45,429	,399	,855
DTK 5	11,84	45,862	,348	,856
DTK 6	11,88	46,596	,243	,858
DTK 11	11,90	46,618	,244	,858
DTK 12	11,74	45,862	,331	,857
DTK 13	11,89	46,359	,282	,858
DTK 14	11,83	45,983	,326	,857
DTK 15	12,06	47,177	,227	,858
DTK 16	11,99	46,756	,260	,858
DTK 17	12,05	46,562	,343	,856
DTK 18	12,04	46,448	,358	,856
DTK 19	11,96	46,147	,352	,856
DTK 20	11,81	45,610	,381	,855
DTK 21	11,97	46,283	,333	,856
DTK 23	11,85	46,395	,266	,858
DTK 24	11,92	45,712	,406	,855
DTK 27	12,01	46,372	,346	,856
DTK 29	11,89	46,093	,328	,857
DTK 30	11,90	46,487	,266	,858
DTK 31	11,90	45,961	,351	,856
DTK 32	11,70	45,949	,317	,857
DTK 33	11,44	46,781	,232	,859
DTK 34	11,97	45,433	,488	,853
DTK 35	12,04	46,019	,447	,855
DTK 36	11,95	45,320	,495	,853
DTK 37	11,81	46,232	,284	,858
DTK 39	12,01	46,661	,292	,857
DTK 40	12,11	47,388	,233	,858
DTK 41	12,06	47,282	,205	,859

(Fortführung Tabelle 6.9)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
DTK 42	11,98	45,696	,445	,854
DTK 43	11,90	45,377	,451	,854
DTK 44	11,81	46,398	,258	,858
DTK 48	11,91	45,316	,469	,853
DTK 49	11,81	46,113	,302	,857
DTK 50	11,55	46,286	,281	,858
DTK 51	11,78	45,913	,328	,857
DTK 52	11,98	45,776	,430	,854
DTK 54	12,03	46,354	,373	,856
DTK 55	11,87	45,540	,409	,855

Zusammenfassend ist festzustellen, dass alle vier DTK-Skalen den in der klassischen Testtheorie als Mindestmass eingesetzten Alpha-Wert 0,60 erreichen oder übertreffen. Demnach darf auch davon ausgegangen werden, dass die Reliabilität des DTK den messtheoretischen Anforderungen nachkommt. Alle vier Skalen messen, was immer sie auch messen sollten, mit zufrieden stellender oder gar guter Qualität.

Entsprechend der zuvor gegebenen Übersicht werden nun nachfolgend die Reliabilitätswerte des DTK bei seiner Anwendung in unterschiedlichen Teilgruppen geprüft. Tabelle 6.10 gibt die ermittelten Reliabilitätswerte der DTK-Subskalen für unterschiedliche Teilgruppen der Stichprobe an.

Tabelle 6.10: Interne Konsistenzen der DTK-Skalen in verschiedenen Subgruppen

	1. D/S (25 Items)	N	2. AV (16 Items)	N	3. M/aR (14 Items)	N	Skala 1+2	N
Gesamtstichprobe	,84	2016	,75	2016	,69	2024	,86	2005
- Mädchen	,84	969	,73	970	,69	974	,87	965
- Jungen	,83	1047	,76	1046	,69	1050	,86	1040
Grundschule	,84	986	,69	989	,71	992	,86	980
3e Klasse	,82	216	,61	216	,70	219	,84	212
4e Klasse	,83	245	,67	247	,71	247	,85	244
5e Klasse	,84	281	,73	282	,74	282	,86	280
6e Klasse	,84	244	,69	244	,68	244	,85	244
Gymnasium	,84	1030	,76	1027	,68	1032	,85	1025
1e Klasse	,85	349	,73	347	,68	350	,86	346
2e Klasse	,83	352	,76	352	,68	352	,84	352
3e Klasse	,85	329	,77	328	,68	331	,84	327
Spezielle Teilgruppen								
ausländische Kinder	,85	157	,70	156	,66	158	,86	156
16-, 17-, und 18jährige	,87	27	,63	27	,72	27	,81	27

Zeichenerklärung: 1. D/S = DTK- Skala 1 „Dysphorie / Selbstwertprobleme“, 2. AV = DTK-Skala 2 „Agitiertes Verhalten“, 3. M/aR = DTK- Skala 3 „Müdigkeit / autonome Reaktionen“

Die in Tabelle 6.10 aufgeführten Alpha-Werte zeigen sowohl in den Spalten als auch in den Zeilen Variationen. Zum einen sind diese Variationen in den Spalten mit der Zahl der jeweils

eingesetzten Items verbunden. Diese Variation ist allgemein bekannt und in der ersten Zeile, der für die Gesamtstichprobe, spaltenweise vorgegeben und in so weit auch als in den Zeilen erwartbar angekündigt. Die Variationen in den Zeilen folgen dem. Von ihnen dürfte erwartet werden, dass sie in ihrer Stärke bemerkenswert sein könnten. Schließlich wird mit den Zeilen die Frage beantwortbar, ob die DTK-Skalen von Klassenstufe 1 der Grundschule durch alle Altersstufen hindurch bis zu den 16-18 jährigen Schülern ähnlich verlässlich gemessen haben sollten oder nicht. Die Antwort ist keineswegs selbstverständlich. Umso erstaunlicher könnte gefunden werden, dass die DTK-Skala 1, in allen Altersstufen gleich reliabel gemessen hat.

Aus der Berechnung der Reliabilitätswerte bei unterschiedlichen Stichproben ergeben sich für die Untersuchung 2034GR im Allgemeinen nur kleinere und insoweit unwesentliche nicht systematischen Schwankungen in der Höhe der Reliabilitätswerte. Diese innere Homogenität ist bei der Skala 2 nicht gegeben. Werden zur Quantifizierung der Unterschiede die jeweiligen Divergenzen zur ersten Zeile, der Gesamtstichprobe, summiert, ergeben sich die nachstehenden Summenwerte: Summe dieser Divergenzen für Skala 1 = 11, Summe dieser Divergenzen für Skala 2= 62 , für Skala 3= 21. Wovon die Güte der Messungen selbst beeinflusst wird, soll an dieser Stelle nicht erörtert werden. Immerhin ließe sich eine Hypothese bilden, nach der die Skala 2 mit den Teilgruppen der Untersuchung, insbesondere dann mit den Altersgruppen, den relativ höchsten Güteschwankungen unterworfen ist.

Wie aus Tabelle 10 zu ersehen, sind die Reliabilitätskoeffizient bei der ersten Skala des DTK bei allen Subgruppen fast identisch. Ähnliches gilt auch für die zusammen gelegte Skala 1+2.

In Bezug auf die Reliabilitätsschwankungen der zweiten DTK-Skala ist zu erkennen, dass mit einer Alterszunahme auch die Alpha-Werte anwachsen. Umgekehrt kann man für die dritte DTK-Skala merken, dass bei den älteren Kindern die Reliabilitätswerte eher geringer sind als bei den jüngeren Schulkindern, obwohl die beobachteten Schwankungen nicht besonders scharf sind.

Für die sich aus diesen Feststellungen ergebenden Nachfragen findet sich an dieser Stelle der Erörterungen kein Raum zu entsprechenden Nachforschungen. Deshalb wird hier auf die dafür vorgesehene Verfahrensroutine zurückgegriffen, mit der dem Übergehen dieser und anderer Nachfragen entgegen gewirkt werden soll. Solche Fragen werden, wie schon weiter oben, als Nachfragen mit dem Kürzel NF, also auch als neu formulierte Fragen NF1 bis NF_n, aufgeführt und inhaltlich gekennzeichnet.

NF4: Sind die mit den Altersgruppen auftretenden Reliabilitätsschwankungen für die DTK-Variable 2 empirisch gestützt erklärbar?

NF5: Womit könnten diese Schwankungen zusammenhängen, und was könnten sie bedeuten?

6.2.2.3 Faktorielle Struktur des DTK

Die Faktorenanalyse basiert jeweils auf der Korrelationsmatrix eines entsprechend strukturierten Datensatzes. Das Auftreten hoher Korrelationen zwischen den Variablen legt die Vermutung nahe, dass sie beobachtete und folglich beobachtbare Indikatoren für eine nicht selbst gemessene, aber eben durch angezeigte Indikatoren vertretene latente Variablen sind. Bei Tests, die als Fragebögen aufgebaut sind, könnte beispielsweise angenommen werden, dass die einzelnen Fragen in ihrer Bedeutung durch den Leser interpretiert werden müssen. Alle diese Fragen zielen auf das Erfassen spezieller Sachverhalte ab, der Leser muss jeweils die Bedeutung einer Frage interpretieren und sie dann nach subjektivem Ermessen beantworten. Dabei könnte geschehen,

dass drei unterschiedlich gemeinte Fragen unter Bezugnahme auf ein und dasselbe subjektive Ermessen beantwortet werden. Wenn diese Art der Beantwortung der drei Fragen von zahlreichen Angehörigen einer Stichprobe praktiziert wird, ergeben sich sehr wahrscheinlich hohe Korrelationen zwischen den fraglichen drei Variablen. Der jeweils von vielen Angehörigen einer Stichprobe in Anspruch genommene übereinstimmende subjektive Bezug hätte den Status einer latenten Variablen. Sie wird nicht selbst gemessen, wohl aber durch Indikatoren erkennbar vertreten. Die Faktorenanalyse ist ein Verfahren, das dazu beiträgt, latente Klassen als Faktoren erkennbar werden zu lassen. Die Faktorenladung gibt dazu dann an, welche Korrelation zwischen einer Messvariablen und dem von der Faktorenanalyse entdeckten Faktor, der latenten Variablen, besteht. Dieser Faktor ist ein Konstrukt und als solches permanent latent und nie selbst zu messen. Da die Beantwortung eines als Fragebogen konstruierten Tests in unterschiedlichen Stichproben durchaus verschiedenartig erfolgen kann, dürfte folglich die jeweils für eine neue Stichprobe gewonnene Datenmatrix auch Unterschiede gegenüber den Matrizen aus anderen Versuchen aufweisen. Wenn sich über die Ergebnisse aus durchgeführten Faktorenanalysen zeigen würde, dass entgegen solchen Erwartungen das wiederholte Auftreten entsprechender Faktoren auf deren Konstanz hindeutete, berechtigte das zu der Hypothese, mit dem jeweiligen Test auch konstante latente Variablen angesprochen zu haben.

Für den DTK haben die Ergebnisse aus Faktorenanalysen für bislang bekannt gewordene deutschsprachigen Anwendungen nahe gelegt, von einer relativ konstanten Faktorenstruktur auszugehen. Indem nun in der Untersuchung 2034GR eine Anwendung in griechischer Sprache erfolgt, ist durchaus zweifelhaft, ob die Ergebnisse aus Faktorenanalysen der griechischen DTK-Daten noch Ähnlichkeiten mit Ergebnissen aus deutschsprachigen Anwendungen haben könnten. Dieser Frage wird im Folgenden nachgegangen.

Für die Prüfung der Faktorenstruktur des DTK wurde dieselbe Methode verwendet, die auch für die ursprüngliche Entwicklung der drei DTK-Subskalen von Rossmann (1993) angewendet worden ist, nämlich eine Hauptkomponentenanalyse mit darauffolgender nicht-orthogonaler (Oblimin) Rotation. Rossmann begründet unter Berufung auf Gorsuch (1988) die Wahl dieses Verfahrens wie folgt: „Diese Vorgangsweise stellt im Grunde die Anwendung einer exploratorischen Faktorenanalyse zu konfirmatorischen Zwecken dar. Es ist nicht nur die rechentechnisch ökonomischste Variante sondern ist zudem, verglichen mit einer konfirmatorischen Faktorenanalyse, auch ein konservativerer Test der Hypothesen in Bezug auf die erwartete Faktorenstruktur, weil die gefundene Lösung von den Hypothesen des Untersuchers weitgehend unabhängig ist.“ (Rossmann, 1993, S. 14).

Im Folgenden werden die Ergebnisse aus der griechischen Untersuchung 2034GR wiedergegeben. Im Kommentar unterhalb dieser Tabelle finden sich die einen Vergleich ermöglichenden Hinweise auf die von Rossmann (1993) gefundenen Faktoren.

Tabelle 6.11: Oblimin-Faktorladungen für die 55 Items des DTK in der griechischen Untersuchung 2034GR

Skala 1: Dysphorie/ Selbstwert				Skala 2: Agitiertes Verhalten				Skala 3: Müdigkeit/ Autonome Reaktionen			
	F1	F2	F3		F1	F2	F3		F1	F2	F3
DTK 4	.52			DTK 2		.48		<i>DTK 1</i>	.36	.40	
<i>DTK 13</i>			.51	DTK 3	.32	.46		<i>DTK 7</i>	.38	.34	
DTK 14	.43			DTK 5		.56		DTK 8			.33
DTK 18	.41			DTK 11		.47		DTK 9	.28		.41
DTK 19	.51			DTK 12		.40		DTK 10			.40
DTK 20	.55			DTK 15		.52		DTK 22			.50
DTK 21	.47			DTK 16		.52		DTK 25	.33		.38
<i>DTK 23</i>			.44	DTK 17		.45		DTK 26			.46
<i>DTK 24</i>	.34	.36	.30	DTK 27		.49		DTK 28	.31		.50
DTK 31	.50			DTK 29		.56		DTK 38			.53
DTK 33	.36			DTK 30		.51		DTK 45	.29		.32
DTK 34	.64			DTK 32		.39		<i>DTK 46</i>		.36	
DTK 35	.53		.39	DTK 37		.33		DTK 47			.51
DTK 36	.52	.43		DTK 39		.43	.29	<i>DTK 53</i>		.37	
<i>DTK 40</i>			.42	DTK 41		.35					
DTK 42	.54		.33								
DTK 43	.55	.27									
DTK 44	.29										
DTK 48	.60		.27								
DTK 49	.34										
DTK 50	.36										
DTK 51	.31										
DTK 52	.51		.31								
DTK 54	.39		.36								
DTK 55	.51		.34								

Ladungen kleiner als .25 werden nicht angeführt

Entgegen allen, auch durch Hinweise auf die beträchtlichen Sprachunterschiede zwischen deutscher und griechischer Anwendung plausibel erscheinende Erwartung divergierender Faktorenstrukturen, überraschten die Ergebnisse der Faktorenanalyse für die DTK-Daten aus der Untersuchung 2034GR durch eine unerwartet hoher Übereinstimmung mit den Ergebnissen von Rossmann (1993). Insgesamt ist festzustellen, dass die erwartete Faktorenstruktur in hohem Maße repliziert worden ist. Die darzustellenden Abweichungen beschränken sich auf die nachfolgenden Daten.

Vier Items, die bei Rossmann dem ersten Faktor zugeordnet waren, finden in 2034GR ihre Hauptladungen im zweiten und dritten Faktor. Das sind die nachstehend aufgeführten Fragen:

Frage 13 („*Wirst Du oft von anderen Kindern verspottet?*“),

Frage 23 („*Tun Dir manchmal andere Kinder absichtlich weh?*“),

Frage 40 („*Hast Du mehr Schmerzen als andere Kinder?*“)

zeigen ihre jeweils höchsten Ladungen gegenüber dem dritten Faktor, und die

Frage 24 („*Bist Du oft unglücklich wegen der Schule?*“) kann allen drei Faktoren zugeordnet, relativ stärker allerdings zum zweiten Faktor.

Interessanterweise kommt Rossmann (1993) bei seinen Analysen für die Fragen 24 und 40 zu ähnlichen Ergebnissen, wie sie sich in 2034Gr zeigen. Rossmann behielt diese zwei Fragen dennoch beim ersten Faktor, „zwecks Erzielung maximaler Reliabilität der ersten Skala...“ (Rossmann, 1993, S. 14).

Der von Rossmann (s.o.) dargestellte Effekt stellte sich im Übrigen auch in der griechischen Version ein. Bei einer erneuten Prüfung des Reliabilitätskoeffizienten für die Daten der ersten Skala in der griechischen Stichprobe sank nach Eliminierung der problematisierten Items 13, 23, 24 und 40 der ursprünglichen Alpha-Wert 0,842 auf den neuen Wert von 0,833. Die Veränderung ist minimal, aber sie liegt in der erwarteten Richtung.

In Bezug auf den zweiten Faktor repräsentiert das griechische Ergebnis eindeutig und ohne nennenswerte Nebenladungen die von Rossmann (1993) für die deutsche Version errechneten Werte zur zweiten Skala des DTK „Agitiertes Verhalten“.

Schließlich ist über vier Items zu berichten, die nach Rossmann (1993) und somit ursprünglich dem dritten Faktor zuzuordnen waren. In der griechischen Version zeigen sie Ladungen zum ersten und zweiten Faktor. Das sind die nachstehenden Fragen:

Frage 1 („*Bist Du morgens meistens gut ausgeschlafen?*“),

Frage 46 („*Stehst Du am Morgen schwer auf?*“) und

Frage 53 („*Bleibst Du am liebsten den ganzen Tag im Bett?*“), die ihre Hauptladungen auf den zweiten Faktor geben, und um die

Frage 7 („*Schläfst Du in der Nacht immer sehr gut?*“), die sich stärker an den ersten Faktor bindet.

Eine daraufhin vorgenommene erneute Prüfung der Reliabilität der dritten Skala nach Elimination der kritischen Items 1, 7, 46 und 53 erbrachte wiederum einen niedrigeren Alpha-Wert (0,66) als hier zuvor <(s.o.) berichtet worden ist (0,69).

Bei drei dieser Fragen (1, 7 und 46) ist ihre inhaltliche Ähnlichkeit vom Interesse, die sich auf das Schlafverhalten der gefragten Kinder bezieht. Ob das gefundene Strukturmodell direkt mit speziellen Eigenschaften der untersuchten Stichprobe zusammenhängt oder doch etwas Allgemeineres, etwa eine gegenüber dem Deutschen erfasste griechische Besonderheit des Schlafens darstellt, das ist eine Frage, die hier gestellt, aber doch nicht zugleich beantwortet werden kann (NF 6).

Die über die Faktorenanalyse gefundenen drei Faktoren korrelieren positiv miteinander. Die höchste der drei Korrelationen ist die zwischen dem ersten und dem dritten Faktor (0,33). Zwischen dem ersten und zweiten Faktor besteht eine Korrelation von 0,28. Die niedrigste Korrelation besteht zwischen dem zweiten und dem dritten Faktor (0,18).

8.2.3 Zur Vertrauenswürdigkeit der mit dem DTK in dieser griechischen Untersuchung 2034GR erhobenen Daten

Die Beantwortung der Frage nach den Messqualitäten eines psychologischen Messinstrumentes ist ein gewichtiges Argument für die Beurteilung der Vertrauenswürdigkeit von Daten. Vertrauenswürdig sind nach der klassischen Testtheorie alle Datensätze, die mit der kontrollierten

Anwendung von Messverfahren gewonnen worden sind, die, empirisch geprüft, als objektiv, reliabel und valide bezeichnet werden dürfen. Für Datensätze ist zu verlangen, dass sie in diesem Sinne jeweils vertrauenswürdig sind, bevor deren Analyse aufgenommen wird. Die Vertrauenswürdigkeit des analysierten Datensatzes ist eine unerlässliche Voraussetzung für die Verlässlichkeit der späterhin mitzuteilenden Analyseergebnisse, wenn auch noch keine für sich selbst hinreichende.

Die deutschsprachige Version des DTK ist ein Messverfahren, das im Hinblick auf seine Objektivität, seine Reliabilität und seine Validität als ein vertrauenswürdige Screening-Instrument zu klassifizieren ist, dessen Nutzbarkeit und Nützlichkeit innerhalb von Diagnosen depressiven Störungen bei Kindern und Jugendlichen als erwiesen gelten darf.

Die in der Untersuchung 2034GR verwendete griechische Version des DTK steht in ihren bislang nachgewiesenen Messqualitäten der deutschen Ursprungsversion nicht nach. Sie erreicht die Objektivität und die Reliabilitätswerte der deutschen Version und stimmt zudem auch in der Faktorenstruktur überein.

Das alles rechtfertigt die Annahme einer messtechnisch im Sinne der klassischen Testtheorie hohen Homogenität der deutschen und der griechischen Version des DTK.

Insoweit ist den Messqualitäten des griechischen DTK unter dem Hinweis zu vertrauen, dass die Validität der griechischen Version bislang nur auf einer Homogenitätsabnahme mit der deutschen Version beruht und empirisch nachzuweisen bleibt.

Mit der Untersuchung 2034GR wird ein erster Schritt dazu unternommen, indem ein zweiter Test, der CDI, ebenfalls eingesetzt werden wird. Sollten die griechische Version des DTK und die griechische Version des CDI in der erwarteten Weise, so wie in deutschsprachigen Untersuchungen, korrelieren, wäre dies ein erstes Validitätsargument sowohl für griechischen DTK als auch für den griechischen CDI.

Dementsprechend wird für die Anwendung des DTK in dieser griechischen Untersuchung von dessen nachgewiesenen Messqualitäten ausgegangen. Angenommen wird, dass er nicht nur empirisch verlässlich misst, sondern, seiner messtechnischen Entsprechung mit der deutschen Version folgend, auch entsprechend valide. Das sollte sich erstmals in der entsprechenden Übereinstimmung der Messdaten aus dem DTK und der griechischen Version des CDI bestätigen. Eine solche Bestätigung ist in der Folge durch weitere empirische Nachweise für die Validität des DTK in einer griechischen Version zu erweitern, zu ergänzen und zu vervollständigen.

6.3 Messqualitäten des CDI

Nachdem die Messqualitäten des DTK in seiner Anwendung innerhalb der Untersuchung 2034GR geprüft worden sind, ist dies auch für die griechische Version des CDI durchzuführen. Dieses Verfahren hat einerseits innerhalb dieser griechischen Untersuchung einerseits als international erprobter Test seine eigene Bedeutung, andererseits kommt ihm gegenüber dem DTK auch die Funktion eines potentiellen Korrektivs zu, wobei dies wechselseitig gilt. Beide Verfahren werden zum Aufspüren von Hinweisen auf das Auftreten depressiver Störungen bei Kindern und Jugendlichen benutzt. Sie stützen ihre Informationsaufnahme beide auf Selbstangaben von Kindern und Jugendlichen, unterscheiden sich aber in ihrem Testaufbau. Der DTK benutzt 54 Items auf drei unterschiedliche Skalen zur Erfassung von Selbstberichten über das Auftreten depressiver Symptome. Der CDI arbeitet mit einer Skala. Ein messtheoretisches Korrektiv stellt jedes dieser beiden Verfahren in dieser griechischen Untersuchung für das jeweils andere dar,

wenn von den messtheoretischen Übereinstimmungen beider Verfahren in ihren deutschen Versionen ausgeht. DTK und CDI werden in ihren deutschen Fassungen als miteinander vergleichbare Instrumente zur Erfassung von Informationen über das Auftreten depressiver Störungen bei Kindern und Jugendlichen gesehen. Für die Anwendung ihrer griechischen Versionen könnte die Erwartung gelten, dass sich die messtheoretischen Ähnlichkeiten der deutschen Versionen bei der Anwendung griechischer Versionen wiederholen würden. Falls das eintreten sollte, könnte dies als eine erste, wenn auch bedingte Aussage zur Validität der Verfahren gewertet werden. Eine solche Validitätsaussage würde allerdings auch darauf aufmerksam machen, dass die für Griechenland beanspruchte Validität beider Verfahren eingehend empirisch geprüft werden muss.

6.3.1 Messqualitäten des CDI in Studien außerhalb der griechischen Untersuchung 2034GR

Die erste Version des CDI wurde nach Kovacs (2004, S. 51) im Jahr 1975 zur Anwendung gebracht. Ihr folgte 1976 eine zweite Version. Die bis jetzt gültige Form des CDI geht auf das Jahr 1979 zurück. Im Vorwort zu dem Handbuch „Children´s Depression Inventory (CDI) - Technical Manual“ (2004) verweist Maria Kovacs als Autorin unter anderem auf die große Zahl der Anwendungen des CDI seit der ersten Skalenkonstruktion und weiterhin dann auch seit 1979. Allein schon die Verbreitung des CDI sichert dem Verfahren einen Platz unter den messtheoretisch außergewöhnlich intensiv überprüften Testverfahren für Kinder und Jugendliche. Die große Zahl der überaus unterschiedlichen Einsätze des CDI verschaffte dem die Möglichkeit der Bewährung in wechselnden Zusammenhängen. Wenn einmal mehr die Testgütekriterien der klassischen Testtheorie bemüht werden dürfen, die weltweit verfügbaren Informationen über Objektivität, Reliabilität und Validität gestatten eine sichere Einschätzung der Messqualitäten des Verfahrens.

Über die testtheoretisch geforderte Objektivität verfügt das Verfahren so gut wie jeder andere Fragebogen für Kinder und Jugendliche. So wie bei jedem anderen Verfahren werden auch Diskrepanzen zwischen der testtheoretisch gemeinten Objektivität und der Objektivität in den jeweiligen aktuellen Anwendungen auftreten, ohne dass danach die Nutzbarkeit des Verfahrens angezweifelt werden müsste.

Das Erreichen der testtheoretisch für Items geforderten Messqualitäten durch den CDI Items belegen so gut wie alle bekannt gewordenen Anwendungen. Kovacs (2004, S. 59) zitiert CDI Item-Total Korrelationen, die alle in dem von der klassischen Testtheorie geforderten Bereich liegen. Sie variieren zwischen 0,13 und 0,60, wobei 0,13 eine extreme Ausnahme unter den zur Kenntnis zu nehmenden Daten darstellt.

Angaben zur Reliabilität des CDI finden sich in der Mehrzahl der Veröffentlichungen über die Anwendung des CDI. Sie weisen mit wenigen Ausnahmen hohe Übereinstimmungen in den berichteten Reliabilitätswerten auf.

Eine der Diskrepanzen wird von Kovacs selbst ermittelt. In einem Bericht über eine mit dem Ziel der Überprüfung seiner Messqualitäten durchgeführte Anwendung des CDI teilte M. Kovacs (1983) für eine psychiatrische und eine pädiatrische Stichprobe sowie für eine Stichprobe von Schulkindern, alle Teilnehmer im Alter von 8 bis 15 Jahren, drei Werte für Cronbachs Alpha mit, 0,85, 0,71 und 0,87. Das erlaubt eine Klassifikation der Messqualitäten des Verfahrens als befriedigend bis gut, wobei die verwendeten Begriffe in ihrer Bedeutung den Zahlenwerten nachzuordnen sind. Die unterschiedlichen Alpha-Werte beziehen sich auf unterschiedliche

Gruppen von Versuchsteilnehmern. Der relativ niedrige Wert von 0,71 bezieht sich auf Teilnehmer, die von Kovacs als „pediatric-outpatients“ bezeichnet werden. Nach deutschem Sprachgebrauch wäre diese Stichprobe als eine von Kinderärzten in Kliniken ambulant behandelte Gruppe von Kindern und Jugendlichen zu bezeichnen. Welche Diagnose der jeweiligen Behandlung unterlag ist nicht bekannt, depressive Störungen könnten innerhalb dieser Störungen aufgetreten sein, sie waren aber nicht das Merkmal dieser Gruppe. Immerhin existieren durch die Anwendung des CDI spezifische Gruppenkennwerte, so ein niedriger Alpha-Wert. Weiterhin lagen nach Kovacs (ebd.) die Item-Total-Korrelationen für diese Gruppe zu über 50% unter 0,25. Diese Werte zeigen, dass der CDI von Kindern und Jugendlichen mit unspezifischen pädiatrischen Diagnosen so beantwortet wird, dass seine Messqualitäten sinken. Empirische Nachweise für die Gründe dieses Stichprobeneffektes sind nicht bekannt.

Nach einer von Kovacs (2004, S.58) gegebenen Übersicht, fanden sich in zahlreichen Untersuchungen überwiegend Alpha Reliabilitätskoeffizienten deutlich über .80.

Der Test-Retest Wert für eine Stichprobe von Schulkindern aus Toronto war nach Kovacs (ebd.) 0,84, für eine Gruppe depressiver Kinder dagegen nur 0,72. Nach einer von Kovacs (2004, S. 60) gegebenen tabellarischen Übersicht variierten Test-Retest Werte für den CDI in 12 Untersuchungen zwischen 1982 und 1991 zwischen $r = 0,38$ und $r = 0,87$. Die Intervalle zwischen den Messzeitpunkten t_1 und t_2 variierten von 1 Woche bis zu 1 Jahr. Für das Intervall von 1 Jahr wurden aus einer Untersuchung mit Schülerinnen und Schülern r -Werte von 0,44 bis 0,69 mitgeteilt (Smucker et al., 1986). Der höchste Wert in dieser Übersicht beträgt $r = 0,87$ und wurde von Saylor, Fint, Spirito und Bennett (1984) für ein Messintervall von 1 Woche aus einer Untersuchung mit ambulant behandelten psychiatrischen Patienten berichtet.

Zur Validität des CDI führte Kovacs (ebd.) Korrelationen von 0,65 zu einem Angsttest (Rev. Children's Manifest Anxiety Scale) und von 0,59 mit einem Selbstwertwerttest (Coopersmith Self-Esteem Inventory). Zu diesem Verfahren fand Green (1980) in einer Stichprobe mit Kindern aus Pennsylvania höhere Korrelationen, nämlich 0,72 für Mädchen und 0,67 für Jungen. Friedmann und Butler (1979) berichteten über signifikante Korrelationen des CDI mit einem Selbstkonzept-Test (Pies-Harris Children's Self-Concept Scale) in Höhe von 0,66. Da der CDI als Test zur Aufnahme von Informationen über das Auftreten depressiver Störungen konstruiert worden ist, kommt den angeführten Korrelationen wegen des Konzeptes der Komorbidität eine gewisse Bedeutung auch in der Beantwortung der Frage nach der Validität des Verfahrens zu, aussagekräftig sind aber erst Daten, die sich speziell auf depressive Kinder beziehen. Unterschiede in der Bearbeitung des CDI durch Kinder mit Major Depression, mit remittierender Depression und Kindern und ohne Depression fand Kovacs (1982). Der Mittelwertsunterschied zwischen Kindern mit Major Depression zum einen und remittierender Depression zum anderen wurde mit $t = 3,42$ und $p < 0,002$ angegeben. Für Kinder ohne Depression ist dann ein Wert von $t = 2,82$ und $p < 0,006$ mitgeteilt worden. Diese Kinder mit Major Depression (Kovacs, 1982) erreichten wiederum höhere Mittelwerte an CDI Punkten als die von Friedman und Butler (1979) untersuchte Stichprobe mit Schulkindern aus Toronto ($t = 3,05$; $p < 0,002$).

Eine Übersicht zur Validitätsproblematik findet sich bei Kovacs (2004, S. 63 ff.). Die reicht von Nachweisen für Vergleichsuntersuchungen mit anderen Messinstrumenten zur Aufnahme von Informationen über das Auftreten depressiver Störungen bei Kindern und Jugendlichen bis hin zu speziellen klinischen Validitätsnachweisen und Validitätsproblemen. Auf Seite 66 (ebd.) schreibt Kovacs eine Entgegnung auf Kritik an Aspekten der Diskussion im CDI Handbuch von Kovacs (1992) mit der auch die Validität des CDI als Screening Verfahren bewertet wird. Dazu schreibt Kovacs (ebd.): „Matthey and Petrovski (2002) correctly critique the discussion of base rates presented in the original manual (Kovacs, 1992) that poorly explicates the measure's value as a screening tool. However, the predictive utility of the CDI has been well reported in research

literature (see Aronen & Soininen, 2000; Reynolds, 1986; Stavrakaki et al., 1991; Timbremont & Braet, 2001). As a result, the CDI should be considered an effective instrument for screening purposes.”

Die Sensitivität und Spezifität des CDI untersuchten Craighead et al. (1995). Ihnen zufolge wird bei einem Cut-off Punkt von 17 mit 80% die höchste Sensitivität bei einer Spezifität von 84% erreicht.

Zahlreiche Angaben finden sich auch über die Faktorenstruktur des CDI. Eine Varimax Rotation der Daten aus der so genannten Toronto Stichprobe führte zur Entdeckung nur einen Faktors (aufgeklärte Varianz = 63,7%). Hodges, Kline et. al. (1982) berichteten über vier Faktoren (aufgeklärte Varianz= 70,8%).

Kovacs (2004) kommt im Hinblick auf den zu diesem Zeitpunkt bestehende Wissensstand über die Faktorenstruktur des CDI zu der nachstehenden Folgerung: „A scree test (Cattel, 1966),...,suggested that as many as six factors are present....It is evident that the CDI is capturing one major second-order factor of depression.Considering factor analytic results obtained by other investigators (e.g. Weiss et a., 1991) and the interpretability of the factors, a five-factor solution was selected.” (Kovacs, ebd., S. 52) Die Erörterungen um die Faktorenstruktur des CDI dürften damit noch nicht abgeschlossen sein. Die Analysen haben einzubeziehen, dass gravierende Unterschiede zwischen Stichproben und klinischen Stichproben bestehen (s. Kovacs, ebd., S. 53). Das Item „Sadness“, im CDI die Frage 1, lädt auf den Faktor NM (negative mood) in einer so genannten normativen Stichprobe mit 0,34. in einer klinischen Stichprobe demgegenüber aber mit 0,22. Wäre dies ein Einzelfall, müsste der anders beurteilt werden als die vorgefundenen Unterschiede. Die sind, jedes für, teilweise sehr markant und in der Struktur der Divergenzen bedenkenswert. Die nachstehende Tabelle zeigt dies exemplarisch für zwei Fragen des CDI in Anlehnung an die Tabellen bei Kovacs (2004, S. 53).

Tabelle 6.12: CDI Faktorenladungen zweier Items im Vergleich einer normativen und einer klinischen Stichprobe (s. Kovacs, 2004, S. 53)

Item	Nr.	Stpr	NM	IP	IE	AH	NSE
Sadness	1	N	.34	.12	.12	.13	-.14
Sadness	1	KL	-.22	-.23	-.03	.23	.17
Reduced social interest	12	N	.01	.28	.11	.25	.22
Reduced social interest	12	KL	-.53	.08	.12	.03	-.07

Zeichenerklärung: Nr. = Position des Items im CDI; Stpr = Stichprobe (N = normative; KL = klinische); NM = Faktor Negative Mood; IP = Interpersonal Problems; IE = Ineffectiveness; AH = Anhedonia; NSE = Negative Self-Esteem

Obschon in der griechischen Untersuchung 2034GR keine dieser Versionen zum Einsatz kommt, wird darauf verwiesen, dass sowohl eine Kurzform, als auch jeweils Versionen des CDI für Eltern und für Lehrkräfte konstruiert worden sind.

6.3.2 Messqualitäten des CDI bei seiner griechischen Anwendung in der Untersuchung 2034GR

Da eine Übertragung der Messqualitäten des CDI aus seiner Originalform in eine griechische Übersetzung nicht als selbstverständlich angesehen werden kann, sind die Messqualitäten der griechischen Übersetzung entsprechend zu prüfen.

6.3.2.1 Itemanalyse der griechischen Version des CDI

Die nachstehende Tabelle bietet Informationen über einige Kennwerte für die griechischen CDI Items aus der Anwendung des Verfahrens in der Untersuchung 2034GR.

Tabelle 6.13: Mittelwerte, Standardabweichungen und Varianzen der griechischen CDI Items

	N	Mean	Std. Deviation	Variance
CDI 24	2025	,82	,63	,40
CDI 13	2026	,80	,52	,27
CDI 2	2027	,69	,59	,35
CDI 15	2027	,67	,78	,61
CDI 11	2026	,66	,66	,43
CDI 14	2027	,57	,64	,41
CDI 17	2026	,56	,62	,38
CDI 19	2026	,54	,63	,39
CDI 23	2024	,50	,66	,43
CDI 26	2028	,47	,53	,28
CDI 18	2027	,44	,55	,31
CDI 16	2028	,39	,56	,31
CDI 8	2026	,37	,61	,37
CDI 3	2027	,37	,52	,27
CDI 6	2026	,33	,60	,36
CDI 20	2026	,32	,56	,31
CDI 9	2027	,30	,66	,43
CDI 25	2025	,29	,53	,28
CDI 7	2027	,27	,52	,27
CDI 21	2027	,26	,52	,27
CDI 22	2024	,25	,48	,23
CDI 4	2029	,24	,46	,22
CDI 5	2028	,22	,51	,27
CDI 27	2028	,20	,44	,19
CDI 10	2023	,18	,43	,18
CDI 1	2028	,16	,41	,16
CDI 12	2027	,12	,37	,14
Valid N	2012			

Die Tabelle 6.13 lässt erkennen, dass die Items 10, 1 und dann insbesondere das Item 12 niedrige Mittelwerte aufweisen und zudem eine geringe Varianz zeigen. Die mit diesen Items gestellten

Fragen sind die folgenden „Jeden Tag ist mir zum Weinen zumute“ (Item 10), „Ich bin immer traurig“ (Item 1) und „Ich bin ungern mit anderen zusammen“ (Item 12). Die Kennwerte lassen erkennen, dass Forderungen der klassischen Testtheorie nicht erwartungsgemäß erfüllt werden. Das Testziel der Differenzierung von Persönlichkeiten scheint gefährdet. So berichten zum Beispiel 85,4% der untersuchten Kinder, dass sie während der letzten zwei Wochen nur selten traurig waren. Die Häufigkeitswerte aller Items mit einem niedrigen Mittelwert sind in der nachstehenden Tabelle 6.14 angegeben.

Tabelle 6.14: Antworthäufigkeiten für Items mit einem niedrigen Mittelwert (griechischer CDI)

cdi 1	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,000	1737	85,4	85,7
	1,000	260	12,8	98,5
	2,000	31	1,5	100,0
	Total	2028	99,7	100,0
Missing	System	6	,3	
	Total	2034	100,0	

cdi 10	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,000	1699	83,5	84,0
	1,000	288	14,2	98,2
	2,000	36	1,8	100,0
	Total	2023	99,5	100,0
Missing	System	11	,5	
	Total	2034	100,0	

cdi 12	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,000	1801	88,5	88,9
	1,000	200	9,8	98,7
	2,000	26	1,3	100,0
	Total	2027	99,7	100,0
Missing	System	7	,3	
	Total	2034	100,0	

Die messtechnische Problematik der in der vorangegangenen Tabelle aufgeführten Items mit einem niedrigen Mittelwert besteht in den Grundzügen darin, dass die Kennwerte für diese Fragen im CDI defizitär sind. Damit ist gesagt, sie seien niedriger als erwartet. Daraus geht die Funktion der Erwartungen hervor. Sie bedingen das erkannte Defizit. Es entsteht aus dem Vergleich von Ist- und Sollwerten. Die Werte, die erreicht werden sollen oder sollten, sind allgemein unabhängig von den Ist-Werten, die sind so, wie sie sind und genügen sich gewissermaßen selbst. Das Defizit entsteht durch die theoretisch begründeten Forderungen, wobei Theorie im weitesten Sinne jede Vorstellung umfassen könnte, die zur Setzung eines Erwartungswertes führt. Da der CDI als Test verstanden wird, ist nicht irgendeine Theorie für die Einführung von Sollwerten heranzuziehen, sondern eine Testtheorie. Die klassische Testtheorie setzt für die von ihr betrachteten Verfahren das

hierarchisch übergeordnete Ziel der möglichst optimalen Differenzierung von Persönlichkeiten. Die Persönlichkeiten sind Merkmalsträger, ein Test wäre optimal, wenn er die kennzeichnenden einzelnen Merkmale objektiv und reliabel erfassen könnte und dabei eine optimale Differenzierung erreicht. Optimal ist dann eine Funktion der Zahl der Merkmalskategorien, die verwendet werden. Bei zwei Merkmalen, die als ja oder nein verstanden werden könnten, sind die Persönlichkeiten auch nur in zwei Gruppen zu teilen. Die oben in Tabelle 6.14 aufgeführten Items trennen die Persönlichkeiten der Stichprobe in zumindest drei Untergruppen. Zum einen sind dies diejenigen die 0 Punkte erlangen, zum anderen aber auch diejenigen, die 1 oder 2 Punkte zugewiesen bekommen. Das Item hätte eine Trennaufgabe optimal erfüllt, wenn nach seinem Einsatz die Zahlen der Mitglieder dieser drei Gruppen bekannt geworden sein sollten. Wenn gemessen wird, soll der wirklich bestehende Zustand erfasst werden. Jede Vorgabe, wie die Relationen zwischen den Teilgruppen zu sein hätten, wäre nur durch Rückgriff auf eine Theorie zu begründen, die vorschreiben wollte, wie die vorgefundene Realität im Falle der erfassten Stichprobe zu sein habe. Die klassische Testtheorie, auf nicht prüfbar, aber sich in der Praxis bewährenden (Bortz und Döring, 2002, S. 192 ff.) Axiomen gründend, benutzt solche Vorgaben. Die Frage, ob alle Angehörigen einer Stichprobe gesund seien, ist beispielsweise eine Frage, die für einen zuständigen Gesundheitsdienst von hoher Bedeutung sein kann. Wenn alle Teilnehmer dies wahrheitsgemäß bejahen können, könnte diese einhellig gegebene Antwort einen außerordentlich hohen Informationswert für den beispielhaft eingeführten Gesundheitsdienst haben. Wenn nur ein Mitglied der Stichprobe die Frage verneinen muss, ließe sich denken, dass auch dies wiederum für den Gesundheitsdienst von außerordentlicher Bedeutung sein könnte. Der Informationswert wäre sogar außerordentlich hoch, wenn sich in diesem Einzelfall potentiell der Beginn einer erwarteten Epidemie ankündigte. Für die klassische Testtheorie sind dessen ungeachtet, und in durchaus sinnvollem Bezug zu ihren Zielsetzungen, zu „leichte“ wie auch zu „schwere“ Items ungeeignet. „Extrem schwierige Items, denen kaum jemand zustimmt, oder extrem leichte Items, die von fast allen Probanden gelöst werden, sind wenig informativ, da sie keine Personenunterschiede sichtbar machen.“ (Bortz und Döring, 2002, S. 218) Welche Kriterien gesetzt werden sollten, kann nur im Hinblick auf die verfolgten Zielsetzungen entschieden werden. Wenn die klassische Testtheorie für Testitems Schwierigkeitsgrade von 0,2 bis 0,8 als empfehlenswert ausgibt, dient sie ihrem Ziel der möglichst ausgiebigen Differenzierung von Persönlichkeit. Je differenzierter sie erfasst werden, um so differenzierter könnte auf sie eingegangen werden. Sowohl der CDI als auch der DTK werden aber eingesetzt, um diejenigen zu erfassen, die, wegen ihrer psychischen Eigenheiten, anders als eine majorisierende Mehrheit zu außergewöhnlichen und in der Grundgesamtheit selten auftretenden Antworten neigt. Rossmann (2005, S. 15) hat explizit darauf verwiesen, andere Autoren klinischer Testverfahren äußern sich ähnlich und handeln dann entsprechend (s. Kovacs, 2004, S. 51 ff.). So werden gegebenenfalls auch Items in den Test aufgenommen, die in der Normstichprobe im Schwierigkeitsgrad unter 0,2 bleiben. Im klinischen Interesse könnten beispielsweise Items dann für nützlich gehalten werden, wenn sie den geringen Prozentsatz behandlungsbedürftiger Kinder mit hoher Treffsicherheit von anderen trennen würden. Sie sollten eine möglichst hohe Sensitivität des Verfahrens begründen helfen.

Die in Tabelle 6.14 aufgeführten CDI Items mit niedrigem Mittelwert sind im Zusammenhang mit solchen Überlegungen anders als unter strikter Bindung an die klassische Testtheorie zu sehen. Stiensmeier-Pelster (2000) äußert sich in ähnlichem Sinne zu dem Auftreten niedriger Kennwerte: „Dies ist jedoch nicht verwunderlich, da die Schulstichproben psychopathologisch eher unauffällige Schüler(innen) umfassen, bei denen im Regelfall keine und nur in Ausnahmefällen eine mäßige Symptomausprägung erwartet werden darf.“ (Manual des DIKJ, S. 24)

Nach allem ist das Auftreten niedriger Mittelwerte oder eines niedrigen Schwierigkeitsgrades allein noch kein Anlass für die Eliminierung eines Items aus einem auf die Erfassung depressiver Störungen gerichteten Testverfahrens. Die weiteren Kennwerte des Items sind zu berücksichtigen. Die nun nachfolgende Tabelle 6.15 zeigt sowohl die Korrelationen aller CDI Items der griechischen Version mit dem CDI-Gesamtpunktwert als auch die entsprechenden Werte aus der „Toronto“ Stichprobe (Butler, 1983, in Kovacs, 2004) und aus einer Studie von Reicher und Rossmann (1993) mit einer deutschen CDI-Version von Nevermann (1989). Die „Toronto“ Studie wurde gewählt, weil die aus dieser Studie herrührenden Itemkorrelationen in dem CDI-Manual (Kovacs, 2004) als repräsentativ für die Messqualitäten des CDI dargestellt sind. Die Studie von Reicher und Rossmann wurde unter mehreren Studien mit dem CDI im deutschsprachigen Raum ausgewählt, weil sie auch Informationen über die Beziehungen zwischen dem CDI und dem DTK bietet. Auf diese Studie wird an anderer Stelle dieser Ausführungen erneut eingegangen werden.

Tabelle 6.15: Trennschärfe des CDI in einer griechischen, einer nordamerikanischen und einer deutschen Stichprobe

CDI-Item*	Kampisiou (2003)	Kovacs (1983)	Reicher & Rossmann (1991)
1 Sadness	,50	,54	,36
2 Pessimism	,35	,43	,30
3 Self-deprecation	,50	,52	,44
4 Anhedonia	,45	,45	,35
5 Misbehavior	,31	,37	,21
6 Pessimistic worrying	,36	,43	,19
7 Self-hate	,50	,53	,45
8 Self-blame	,44	,33	,38
9 Suicidal ideation** „will nicht zur Schule gehen“****	,43**		
10 Crying spells	,42	,47	,40
11 Irritability	,45	,55	,40
12 Reduced social interest	,31	,45	,24
13 Indecisiveness	,32	,30	,33
14 Negative body image	,51	,40	,46
15 School-work difficulty	,45	,39	,31
16 Sleep disturbance	,49	,37	,34
17 Fatigue	,44	,36	,33
18 Reduced appetite	,27	,25	,34
19 Somatic concerns	,54	,28	,11
20 Loneliness	,51	,55	,52
21 School dislike	,49	,48	,29
22 Lack of friends	,42	,38	,39
23 School performance decrement	,50	,44	,35
24 Self deprecation (via peer comparison)	,42	,46	,39
25 Feeling unloved	,44	,42	,48
26 Disobedience	,43	,34	,31
27 Fighting	,48	,42	,43

*Die mit den CDI Items gestellten Fragen sind in dieser Tabelle durch die Begriffe ersetzt, die von Kovacs, 2004, S. 59, als inhaltlich kennzeichnend verwendet werden. ** = Item, verwendet bei Kovacs in der originalen Version des CDI, *** = Item, verwendet durch Kampisiou

Die Items des griechischen CDI weisen ausnahmslos signifikante Trennschärfen auf, die im Bereich von $r = 0,27$ bis $r = 0,54$ liegen, wobei die Mehrheit der Koeffizienten jeweils Werte zwischen $r = 0,40$ und $r = 0,50$ erreicht. Wichtig zu bemerken ist auch, dass die wegen ihres niedrigen Mittelwertes als problematisch bezeichneten Items 1, 10 und 12 durchaus befriedigende Trennschärfen besitzen. Zugleich wird die interne Konsistenz der Skala nicht gesteigert, wenn die Items entfernt werden (s.u.). Aus beiden Gründen wäre die Elimination dieser Items aus der CDI-Skala nicht empfehlenswert.

Der Vergleich der Werte für die Items der griechischen Version mit denen aus den anderen Anwendungen (s. obige Tabelle) erweist die relativ gute Messqualität der griechischen CDI Version. Sie ist relativ gut, weil als Qualitätsmerkmal das Erreichen oder Übertreffen der Werte aus den beiden anderen Anwendungen zu verwenden ist. Diese Feststellung ist ein gewichtiges Argument für die Feststellung der messtechnischen Gleichwertigkeit der CDI Versionen. Die Übersetzung der CDI Items ins Griechische hat die Messqualitäten der Skala, gemessen an Merkmalen der Reliabilität nicht beeinträchtigt. Demnach dürfte, falls nicht noch andere Fakten dem widersprechen sollten, den Daten aus der griechischen Anwendung ebenso vertraut werden, wie denen aus anderen Untersuchungen mit dem CDI.

Die zuvor getroffene Entscheidung gilt mit ähnlicher Argumentation auch für das Item 18, das nur niedrige Trennschärfe erreicht. Das Item zeigt ähnlich niedrige Trennschärfe (s. obige Tabelle) auch in den Studien von Kovacs und von Reicher & Rossmann.

6.3.2.2 Interne Konsistenz

Die interne Konsistenz der verwendeten CDI-Übersetzung, berechnet mittels Cronbach's Alpha, beträgt in der Gesamtstichprobe 0,83. Dieser Wert gilt als zufrieden stellend und erlaubt die weitere Verwendung der aus dieser Arbeit gewonnenen Informationen.

Tabelle 6.16: Reliabilität der griechischen Version des CDI

	Cronbach's Alpha	N of Items
	,831	27

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
c1	10,74378	39,962	,453	,823
c2	10,21143	40,084	,268	,829
c3	10,53785	39,346	,437	,823
c4	10,65835	39,964	,385	,825
c5	10,68156	40,598	,241	,829
c6	10,57736	40,037	,272	,829
c7	10,63267	39,335	,431	,823
c8	10,53686	39,357	,356	,825
c9	10,72501	40,321	,359	,826
c10	10,24797	39,022	,363	,825
c11	10,77736	41,100	,255	,829
c12	10,10131	40,556	,241	,829
c13	10,32896	38,642	,428	,822
c14	10,23464	38,460	,349	,827
c15	10,53341	39,186	,421	,823
c16	10,46279	40,799	,188	,831
c17	10,34279	39,290	,358	,825
c18	10,36699	38,419	,467	,821
c19	10,58625	39,032	,445	,822
c20	10,63760	39,367	,429	,823
c21	10,65538	40,044	,360	,825
c22	10,41044	38,568	,422	,823
c23	10,08797	39,339	,339	,826
c24	10,61736	39,707	,365	,825
c25	10,43316	39,714	,364	,825
c26	10,70378	39,904	,422	,824
c27	10,60106	39,171	,345	,826

Die Reliabilitätskennwerte wurden auch für verschiedene Subgruppen berechnet. Die sich für beide Geschlechter und für alle Schulstufen ergebenden Werte weisen eine in hohem Maße homogene Konsistenz nach. Nur in der Teilstichprobe der jüngeren Kinder ist der Reliabilitätswert etwa niedriger (0,78), erreicht aber den erwünschten Wert für die Verwendung eines Tests als Forschungsinstrument. Offen für weitere Untersuchungen bleibt die Frage, ob dieser niedrige Wert mit Schwierigkeiten der Kinder dieses Alters oder mit einer besonderen Teilstichprobe in Verbindung zu bringen ist.

Tabelle 6.17: Interne Konsistenz des CDI in der Gesamtstichprobe und in einzelnen Teilstichproben

Stichprobe	Cronbach's Alpha
Gesamt (N=2025)	,83
Knaben (N=1050)	,82
Mädchen (N=975)	,84
Schüler der 3. Grundschuleklasse (N=221)	,78
Schüler der 4. Grundschuleklasse (N=246)	,81
Schüler der 5. Grundschuleklasse (N=283)	,81
Schüler der 6. Grundschuleklasse (N=244)	,81
Schüler der 1. Gymnasiumsklasse (N =351)	,86
Schüler der 1. Gymnasiumsklasse (N =351)	,84
Schüler der 1. Gymnasiumsklasse (N =329)	,82

Die aus der Reliabilitätsprüfung der griechischen Version des CDI mitzuteilenden Kennwerte sprechen ausnahmslos für eine Übereinstimmung der Reliabilität der griechischen Version mit den englisch- und den deutschsprachigen Formen des CDI.

6.3.2.3 Faktorenstruktur der griechischen Form des CDI

Im Folgenden wird auf die Ergebnisse der durchgeführten Faktorenanalyse einzugehen sein. Das erfolgt dann vor dem Hintergrund der durch Kovacs (2004) mitgeteilten Lösung. Einzubeziehen ist dann aber auch die Erwartung, dass Faktorenstrukturen durch das unterschiedliche Antwortverhalten der Teilnehmer aus verschiedenen Stichproben im Vergleich von Stichprobe zu Stichprobe variieren können. Dem steht eine andere Sichtweise gegenüber, die in Betracht zieht, dass von Stichprobe zu Stichprobe identische psychische Strukturen angesprochen worden sein könnten, was dann auch zu relativ konstanten Faktorenstrukturen führen sollte.

Als Beispiel wird hier zunächst das Ergebnis der Faktorenanalyse aus einer deutschsprachigen Untersuchung angeführt, der von Stiensmeier (1988). Seine Stichprobe umfasste 153 Gymnasialschüler- und Schülerinnen im Alter zwischen 10 und 13 Jahren. Stiensmeier berichtete zum einen für diese Untersuchung über einen Alpha-Wert von 0,86. Dann teilt er das Auffinden von 9 Faktoren mit, von denen aber nur 4 interpretierbar seien. Diese Faktoren kennzeichnet der Autor wie folgt: 1. „mangelnder Sozialkontakt oder Ablehnung durch andere“, 2. „negative Stimmung oder Unlustgefühle“, 3. „mangelnde Kompetenzerwartungen“ und 4. „negative Sicht eigener schulischer Leistungen“. Erwartungsgemäß stehen die aufgeführten Faktoren in inhaltlicher Nähe zu den von Kovacs beschriebenen, aber ein realistischer Vergleich ist allein über die jeweils unterliegenden Daten möglich. Die Verbalisierungen der Faktoren sind kommunikative Hilfen zu deren Kennzeichnung, verleiten aber andererseits auch zu irrtümlichen Akzentuierungen von vermeintlichen Übereinstimmungen oder Divergenzen in Analyseergebnissen.

Für die griechische Version des CDI in der Untersuchung 2034GR ergaben sich die nachstehend aufgeführten Daten.

Nach der Durchführung einer Hauptkomponentenanalyse mit anschließender Varimax-Rotation haben sich nach dem Eigenwertkriterium (Eigenwert > 1) sieben Faktoren ergeben. Der erste Faktor (Eigenwert = 5,248) erklärt 19,44% Prozent der Gesamtvarianz. Alle sieben Faktoren erbringen zusammen eine Aufklärungsrate von 46,57 der Varianz. Die für die einzelnen Faktoren gefundenen Kennwerte zeigt die nachstehende Tabelle.

Tabelle 6.18: Faktorenanalyse der CDI-Daten aus der griechischen Untersuchung 2034G

Component	Total Variance Explained		
	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	5,248	19,436	19,436
2	1,676	6,207	25,643
3	1,376	5,095	30,737
4	1,121	4,153	34,890
5	1,113	4,123	39,013
6	1,021	3,780	42,794
7	1,009	3,737	46,530
8	,978	3,622	50,153
9	,898	3,327	53,480
10	,866	3,207	56,687
11	,863	3,195	59,882
12	,828	3,068	62,950
13	,807	2,989	65,939
14	,794	2,942	68,881
15	,772	2,858	71,739
16	,749	2,774	74,513
17	,721	2,670	77,182
18	,692	2,564	79,747
19	,686	2,541	82,287
20	,675	2,502	84,789
21	,663	2,455	87,244
22	,643	2,382	89,626
23	,634	2,349	91,974
24	,582	2,157	94,131
25	,573	2,122	96,253
26	,539	1,995	98,249
27	,473	1,751	100,000

Extraction Method: Principal Component Analysis.

(Fortführung Tabelle 6.18)

ITEM	F 1	F2	F3	F4	F5	F6	F7
CDI 9 ** "will nicht zur Schule gehen"	,72						
CDI 21 School dislike	,66						
CDI 23 " School performance decrement	,61						
CDI 15 School-work difficulty	,56						
CDI 8 Self-blame		,64					
CDI 7 Self-hate		,48					
CDI 14 Negative body image		,47		,35			
CDI 24 Self deprecation (via peer comparison)		,45					
CDI 3 Self-deprecation	,35	,44					
CDI 26 Disobedience	,40	,43					
CDI 27 Fighting	,32	,34					
CDI 19 Somatic concerns			,69				
CDI 11 Irritability			,66				
CDI 17 Fatigue			,63				
CDI 22 Lack of friends				,69			
CDI 25 Feeling unloved				,55			
CDI 20 Loneliness			,38	,53			
CDI 10 Crying spells			,30	,37			
CDI 6 Pessimistic worrying					,73		
CDI 1 Sadness			,33		,50		
CDI 4 Anhedonia	,42				,41		
CDI 16 Sleep disturbance					,33		
CDI 2 Pessimism						,56	
CDI 13 Indecisiveness						,56	
CDI 12 Reduced social interest				,37		-,50	
CDI 18 Reduced appetite							,76
CDI 5 Misbehavior							-,44

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. a Rotation converged in 10 iterations*Die mit den CDI Items gestellten Fragen sind in dieser Tabelle durch die Begriffe ersetzt, die von Kovacs, 2004, S. 59, als inhaltlich kennzeichnend verwendet werden. ** Suicid-Item der Originalversion ersetzt durch griechisches Item 9.

Die Faktorenanalyse gilt unter anderem als ein hypothesengenerierendes Verfahren. So ist nach einer Faktorenanalyse zum einen zu prüfen inwieweit die aktuell gefundene Lösung mit vorangegangenen übereinstimmt. Zum anderen ist die Frage zu beantworten, was die bekannt werdenden Gemeinsamkeiten, ausgedrückt durch die Faktorenbildung und die Ladung der einzelnen Variablen auf den Faktor bedeuten könnten. Die Wahl des Konjunktivs ist zwingend. Ob diese Bedeutung empirisch verifizierbar ist, müsste sich in nachfolgenden Untersuchungen erweisen.

Die Faktoren werden hier als Konstrukte betrachtet, die durch das Verfahren selbst gebildet werden und deren Definition durch die Variablen und deren Faktorenladungen auf den Faktor erfolgt. Hypothesengenerierend ist der Versuch der Übersetzung dieser Definition in psychologische Begriffe.

Der in dieser Untersuchung auftretende Faktor F1 hat eine zunächst widersprüchliche erscheinende Struktur, wenn der Ort betrachtet wird, den die einzelnen Variablen betreffen. Der Ort der Aussage selbst ist die antwortende Persönlichkeit, aber sie antwortet in Bezug auf Orte,

die durch die Fragen angesprochen werden. Der Faktor F1 zieht die Antwortvariablen der Fragen 9, 21 und 23 mit Ladungen von 0,61 bis 0,72 auf sich. Die Aussage, nicht zur Schule gehen zu wollen ist die korrelativ gewichtigste Antwort einer großen Untergruppe der Befragten. Die Trennschärfe der Items liegt bei 0,50. Der mit den Fragen repräsentierte Ort ist jeweils die Schule. Sie stimulieren bei einem Teil der Befragten übereinstimmend Antworten, die ein negativ akzentuiertes Erleben der Schule bezeugen. Faktor F1 ist durch weitere Variablen konstituiert worden, die dann andere Orte als die Schule betreffen. Zum einen ist durch die Items 26 und 27 die Interaktion zwischen Persönlichkeit und sozialer Umwelt konfliktbesetzt akzentuiert. Zum anderen ist der angesprochene Ort das Selbst, das mit der Antwort zu Frage 3 auf Selbstwertminderungen verweist. Ein subjektiv belastetes Erleben der Schule bildet eine Schnittmenge mit dem Erleben konflikthafter sozialer Interaktionen und beide überlappen mit Selbstwertproblemen. Schulkinder, die mit ihren Antworten den Faktor F1 zur Existenz verhalten, erleben die Leistungsforderungen der Schule als emotional belastend und assoziieren das mit konflikthafter sozialer Interaktion und Erniedrigung. Da Schule vor allem ein sozial definierter Ort ist, verschmelzen die Grenzen zwischen dem explizit schulischen und dem sozialen Aspekt, die beide mit dem Erleben von Erniedrigung assoziiert sind. Dementsprechend wird hier das mit dem Faktor F1 konstituierte Konstrukt mit der inhaltlich orientierten Bezeichnung „Subjektiv erniedrigende und schulisch akzentuierte Sozialkonflikte“ belegt. Die darauf neu zu bildende Hypothese ist die folgende: „Kinder, die mit der Beantwortung der griechischen CDI Fragen 21, 23, 15, 3, 20,27 hohe Werte erreichen, werden sich auch bei Messungen mit anderen Verfahren als mit schulischen Leistungsanforderungen und sozialen Konflikten belastet und zugleich erniedrigt fühlen.“

Faktor F2 vereint die Variablen 8,7,14,24,3, 26 und 27 auf sich. Der psychologische Ort der über die erstgenannten drei Variablen angesprochen wird, ist das Selbst (Item 8, 7, 14), von dem negative Selbsteinschätzungen gegeben werden. Aber auch die nächsten beiden Variablen (Item 24 und 3) nehmen Selbsteinschätzungen auf, die eine setzt den eigenen Wert herab, die andere weist auf die Geringschätzung der eigenen Person durch andere hin. Der soziale Bezug wird über die beiden restlichen Items (26 und 27) ergänzt. Faktor F2 vereint Informationen über „massierte Selbstwertprobleme, die in Verbindung mit konflikthafter Sozialbeziehungen stehen“.

Sowohl an Faktor F1 als auch an F2 ist der Strukturteil der Erniedrigung in Verbindung mit konflikthafter Sozialbeziehungen beteiligt.

Der Faktor F3 wird über die Items 19, 11, 17, 20, 10, und 1 konstituiert. Damit repräsentiert er „somatische Beschwerden“ (Items 19, 17) „in Verbindung mit massierten Bekundungen emotionaler Niedergeschlagenheit (Items 11, 20, 10, und 1).

Faktor F4 vereint mit den Items 14, 22, 25, 20, 10 und 12 massierte negative Selbstaussagen. Die eigene Person wird nicht geliebt (Item 14), auch nicht von anderen (Item 25), sie verharrt in Einsamkeit (Items 20, 22), Traurigkeit (Item 10) und verliert die Freude am sozialen Leben (Item 12). Mit dem Faktor 4 repräsentieren sich Bekundungen von „Liebesverlust, Einsamkeit und Traurigkeit im Zusammenhang mit dem Verlust an sozialen Interessen“.

Faktor 5 umfasst die Items 6, 1, 4 und 16. Damit repräsentiert er „pessimistisches Grübeln, Traurigkeit und Schlaflosigkeit“ als ein Syndrom.

Zum Faktor 6 gehören die Variablen 2, 13 und 12. Mit ihm wird „Pessimismus, Unentschlossenheit und Verlust an sozialem Interesse“ bekundet.

Der Faktor 7 gründet auf den Items 18 und 15. Er repräsentiert „Bedürfnisverlust und Fehlverhalten“.

Die identifizierten sieben Faktoren der griechischen Untersuchung 2034GR stehen alle in inhaltlicher Korrespondenz mit den nach DSM-IV und ICD-10 bei depressiven Störungen zu erwartenden Symptomen. Die Übereinstimmung mit der von Kovacs beschriebenen Faktorenstruktur des CDI (Kovacs, 2004) ist eher gering.

6.3.3 Zur Vertrauenswürdigkeit der mit dem CDI in dieser griechischen Anwendung erhobenen Daten

Die vorangegangenen Darstellungen von Angaben zur Messqualität des CDI in der griechischen Untersuchung von Kampisiou, (2034GR) dienten ausschließlich einem Zweck. Geprüft werden soll, ob die Messqualitäten des CDI in dieser griechischen Untersuchung die der Originalversion von Kovacs (2004) erreichen oder verfehlen. Diese Frage dient keinem Selbstzweck, sie wird vielmehr notwendig, weil die Übersetzung des CDI in eine griechische Version nicht zwangsläufig auf den Transfer der Messqualitäten vertrauen darf. Indessen weisen die mitgeteilten Daten nach, dass die griechische Version des CDI, auch nach dem Austausch des von Kovacs benutzten Suizid-Items, die Messqualitäten der Originalversion zweifelsfrei erreicht. Dann darf aber auch der Aussagekraft der mit dem griechischen CDI gewonnenen Informationen über das Auftreten depressiver Symptome uneingeschränkt vertraut werden. Eine erste griechische Validitätskontrolle ermöglicht die gleichzeitige Verwendung des DTK von Rossmann (1993, 2005). Dessen ungeachtet wird hier ausdrücklich auf die Notwendigkeit von möglichst bald geführter empirischer Untersuchung zur spezifischen griechischen Validität des CDI verwiesen.

6.4 Messqualitäten des ESDM

Auch für den in der Untersuchung von Kampisiou ebenfalls eingesetzten ESDM (Domènech et al., 1985) ist die Frage aufzuwerfen, ob den mit diesem Test erhobenen Daten das erforderlichlich Vertrauen geschenkt werden darf. Das Kriterium für die Vertrauenswürdigkeit der Daten ist der Nachweis einer hinreichenden Messqualität dieses Verfahrens. Das soll als erwiesen gelten, wenn Objektivität, Reliabilität und Validität dieser Skala in deren griechischer Version dem entsprechen, was für die Skala ESDM aus voran gegangenen Untersuchungen bekannt geworden ist.

6.4.1 Die Skala ESDM und ihre Bewährung in voran gegangenen Untersuchungen

Die gilt als messtechnisch vertrauenswürdig und ist in Spanien durch Domènech und Mitarbeiter in relativ großen Stichproben als Mittel zur Fremdeinschätzung von Tausenden von Schulkindern erprobt worden (Domènech et al., 1985; 1991). Das Verfahren war als ein für das Erfassen von Beobachtungen über das Auftreten depressiver Symptome bei Schulkindern einzusetzendes Instrument entwickelt worden, dass durch die Lehrkräfte der Kinder ausgefüllt werden soll. Der Konstruktion ging die Überlegung voraus, dass Lehrer und Lehrerinnen diejenigen sind, die den Grossteil eines Jahres mit den Kindern in ihren Schulklassen zubringen. Dabei wirken sie auch als teilnehmende Beobachter. Dann könnten sie auch über Informationen zum Auftreten beobachtbarer depressiver Symptome bei ihren Kindern verfügen. Anders als Eltern haben Lehrkräfte zugleich mit den Beobachtungen des einzelnen Kindes auch Vergleichsdaten über Verhaltensweisen altersgleicher Kinder und können so relativ zuverlässige Maßstäbe über die Bedeutung ihrer Beobachtungen ausbilden. Domènech hoffte, dass die im ESDM mitgeteilten Beobachtungen der Lehrkräfte einerseits als Screening Verfahren und andererseits als Hilfsmittel bei der Diagnose von Schulkindern nutzbar und somit hilfreich sein könnten. Hilfreich wäre allein schon, wenn bei der großen nationalen Aufgabe der psychischen Gesunderhaltung spanischer Schulkinder der Einsatz von Lehrkräften die personellen Ressourcen verstärken würde. Domènech und Mitarbeiter folgten mit der Entwicklung des Verfahrens ESDM einem

internationalen Trend, der unter anderem auch mit dem in der Methodik erhobenen Forderungen nach simultaner Berücksichtigung von Daten aus unterschiedlichen Quellen, unterschiedlichen Verfahren und dabei unter Einbeziehung verschiedenartiger Konzepte in Verbindung zu bringen ist (Campbell und Fiske, 1959; Bortz und Döring, 2002, S. 202). In der internationalen Forschungsliteratur zum Auftreten von Depressionen bei Kindern und Jugendlichen finden sich mehrere Untersuchungen, die Fremdbeobachtungen der Lehrkräfte von Schulkindern und auch den Beobachtungen von Eltern nachgehen. Dieser Tendenz folgte beispielsweise auch Kovacs, die CDI-Versionen für Lehrer und für Eltern konstruierte (Kovacs, 2004, S. 51 ff.). Was insoweit als eine im Sinne von Campbell und Fiske angestrebte Erweiterung und Verbesserung der Informationsbasis verstanden werden kann, erweist sich in anderer Perspektive als ein Dilemma. Das Konstrukt einer depressiven psychischen Störung bei Kindern und Jugendlichen ist sowohl im DSM-IV als auch im ICD-10 normativ beschrieben und festgelegt worden. Unterschieden werden unterschiedliche Formen dieser Störung, so etwa die Major Depression und die Dysthymia. Danach gelten bis auf weiteres die auf solche Formen bezogenen Übereinkünfte. Wenn das Auftreten der ein Störungsbild operationalisierenden Symptome unter Einhaltung der festgelegten diagnostischen Regeln nachgewiesen sein sollte, ist im gegebenen Fall vom Auftreten der entsprechenden Form einer depressiven Störung auszugehen. Sie gilt als nachgewiesen. Mit der Bestimmung des jeweiligen Störungsbildes wird aber nicht verbindlich vereinbart, wie der Nachweis des Auftretens einzelner Symptome zu führen ist. Wenn beispielsweise die eindeutige Vorschrift eingeführt wird, Voraussetzung für das Erstellen einer Diagnose sei das Auftreten der fraglichen Symptome während eines vollen Jahres, dann steht dieser Eindeutigkeit in der Festlegung des Zeitintervalles die Frage unbeantwortet gegenüber, welche Information vorhanden sein müsse, um über das Erreichen des genannten Kriteriums verbindlich entscheiden zu dürfen. Eine empirisch korrekte Form des Nachweises wäre die fortlaufende Beobachtung des Klienten über den verlangten Zeitraum von einem Jahr. Dieser Nachweis dürfte in der diagnostischen Praxis kaum zu erbringen sein. Die Frage bleibt, was dann an dessen Stelle zu setzen sei. Eine Möglichkeit ist die Aufnahme der Selbsteinschätzung des Klienten. Das Kind oder der Jugendliche werden befragt, ob die Symptome schon ein Jahr bestünden. Die Einbußen an empirischer Verlässlichkeit sind gravierend. Weitere Möglichkeiten sind durch Aussagen auf der Grundlage von Fremdbeobachtungen denkbar. Dann wäre die entsprechende Frage den Müttern, den Vätern, Freunden, Verwandten, Lehrkräften und vielleicht auch weiteren Personen zu stellen. Das Resultat wäre zunächst nicht mehr als eine Gesamtmenge an Aussagen. Von keiner würde gewusst werden, ob sie der Realität entspricht. Darüber müsste nun unter Zuhilfenahme von Entscheidungsmodellen befunden werden. Ein bekanntes Entscheidungsmodell folgt der Regel, dass die gesuchte realistische Information das Kennzeichen der Übereinstimmung aller Zeugenaussagen tragen müsse. Wenn alle Zeugen sich geirrt haben sollten, wäre die Inanspruchnahme des Kriteriums der Übereinstimmung die Grundlage eines Fehlurteils. Andere Entscheidungsmodelle gehen von geminderten Forderungen nach Übereinstimmung aus. So könnten 90, 60, 75 oder auch 55% Übereinstimmung schon hinreichend sein, um eine Aussage als realitätskonform annehmen zu können. Durch die Aufnahme von Selbstaussagen über das Auftreten depressiver Symptome (Beispiel CDI), Aussage der Eltern (Beispiel CDI P) und Aussage von Lehrkräften (Beispiel CDI T) wird eine solche Sammlung von Zeugenaussagen vorgenommen, ohne dass ein externes Entscheidungskriterium für den Realitätsgehalt der Aussagen herangezogen werden könnte. Gerichte helfen sich in ähnlichen Fällen damit, dass sie die Glaubwürdigkeit der Zeugen einschätzen lassen. Die hier exemplarisch für den CDI angeführten Korrelationen zwischen den Gesamtpunktwerten der drei CDI Versionen für die betroffenen Kinder, deren Eltern und deren Lehrkräfte veranschaulichen das diagnostische Dilemma.

Tabelle 6.19: Korrelationen zwischen den Gesamtwerten dreier CDI Versionen (in Anlehnung an Kovacs, 2004, S. 57)

Verfahren	CDI	CDI P	CDI T
CDI	1,00	0,41*	0,46*
CDI P		1,00	0,64*
CDI T			1,00

Zeichenerklärung: CDI = CDI für Kinder und Jugendliche; CDI P = CDI für Eltern; CDI T = CDI für Lehrkräfte; $p < 0.05$

Aus der Übersicht ist zu ersehen, dass die Übereinstimmung zwischen Selbst- und Fremdurteilen den Wert von $r = 0,50$ nicht übersteigt. Geschätzte 25% Übereinstimmung reichten aus, um diesen Koeffizienten entstehen zu lassen. Die Übereinstimmung zwischen den beiden Fremdurteilen, dem der Eltern und dem der Lehrkräfte, ist höher.

Wenn zu entscheiden wäre, welcher Einschätzung gefolgt werden sollte, stünden zunächst keine externen Entscheidungskriterien zur Verfügung. Ein weiteres Urteil wäre dann vom Diagnostiker zu erwarten. Dessen Urteil würde sich wiederum auf Aussagen stützen müssen, die von den Kindern und Jugendlichen selbst, oder aber aus ihrer sozialen Umgebung gegeben werden.

Die geführte Erörterung soll an dieser Stelle verdeutlichen, dass die von Domènech mit der Konstruktion der Skala ESDM eingenommene methodische und diagnostische Position die sowohl für ein Screening Instrument als auch für ein diagnostisches Hilfsmittel gewünschte Entscheidung, ob die aufgenommenen Informationen eher für das Bestehen einer depressiven Störung sprechen oder nicht, keineswegs nur erleichtern.

Die von Domènech und anderen mitgeteilten Daten zu den Messqualitäten der Skala ESDM sprechen für die Objektivität des Verfahrens und für Reliabilitätswerte, die ein gutes Forschungsinstrument ausweisen (u.a. Domènech & Polaino, 1990; Ezpeleta, 1987; Mestre, Garcia, Frias & Moreno, 1993; Monreal, 1988). Eine depressionsspezifische Validität wird von Domènech und Mitarbeitern (1985) angenommen.

Daten über die Bewährung der Skala ESDM außerhalb der griechischen Untersuchung 2034GR werden in den nachstehenden Ausführungen zum Vergleich mit denen aus der griechischen Anwendung aufgeführt.

6.4.2 Messqualitäten des ESDM bei seiner griechischen Anwendung in der Untersuchung 2034GR

Die ESDM-Skala besteht aus 16 Aussagen, die, obschon in eigenen Formulierungen und dann weiterhin für die Beantwortung durch Lehrkräfte verbalisiert, mit den Items des Selbsteinschätzungsverfahrens CDI korrespondieren. Die partielle inhaltliche und auch formale Übereinstimmung der Instrumente erlaubt aber keine weiteren Vorannahmen.

Bei der Durchführung der ESDM ändert sich bestimmungsgemäß die Beobachterperspektive, in der nach Informationen über das Auftreten depressiver Symptome in der jeweils untersuchten Stichprobe gefahndet wird. Ob nun die mit der Skala ESDM erfassten Informationen aus der Sicht von Lehrkräften höhere Übereinstimmungen mit den Selbstberichten der Schulkinder ergeben als bei der Anwendung der entsprechenden CDI Skalen für Schulkinder und ihre Lehrkräfte wird nachfolgend an anderer Stelle dieser Ausführungen zu betrachten sein, wenn auf die Ergebnisse der Datenanalyse der Skala ESDM einzugehen ist. Die aus der Perspektive des Verfahrens ESDM zu gebenden Informationen können in dieser Untersuchung mit zwei aus der Perspektive der

Selbsteinschätzung erhobenen Daten konfrontiert werden, mit den Ergebnissen aus CDI und DTK.

Zunächst muss aber für das ins Griechische übersetzte Messinstrument ESDM nachgewiesen werden, dass es hinreichend verlässlich misst. Diesem Ziel dienen die nachfolgenden Analysen zu den psychometrischen Eigenschaften der griechischen Version des ESDM.

6.4.2.1 Itemanalyse

In der Tabelle 6.20 sind für jedes einzelnes Item der griechischen Skala ESDM der Mittelwert, die Varianz und die Standardabweichung angegeben.

Tabelle 6.20: Kennwerte der Fragen 1 bis 16 des ESDM in seiner griechischen Fassung (Kampisiou)

	N	Mittelwert	Standard- abweichung	Varianz
ESDM 2	1947	,70	,71	,50
ESDM 15	1943	,64	,73	,53
ESDM 14	1942	,58	,70	,50
ESDM 12	1913	,53	,59	,34
ESDM 5	1941	,50	,64	,41
ESDM 13	1914	,46	,55	,31
ESDM 1	1947	,45	,54	,30
ESDM 7	1940	,43	,55	,30
ESDM 9	1868	,42	,58	,33
ESDM 10	1945	,37	,58	,34
ESDM 6	1918	,27	,52	,27
ESDM 8	1880	,26	,51	,26
ESDM 4	1905	,20	,42	,17
ESDM 16	1881	,20	,44	,20
ESDM 3	1895	,19	,44	,20
ESDM 11	1873	,18	,40	,16
Valid N	1825			

Die meisten Items der griechischen ESDM-Skala erreichen einen Mittelwert im Bereich zwischen 0,26 und 0,53. Ein ineffizienter Informationsträger scheint das Item 11 („*neigt zu Selbstvorwürfen*“) zu sein, das nur einen Mittelwert von 0,18 aufweist. Aber auch das Item 3 „*ist in der Pause allein*“) zeigt nur einen Mittelwert von 0,19. Für diese zwei Items wurde eine Häufigkeitstabelle erstellt.

Tabelle 6.21: Häufigkeitsauszählungen für Items mit niedrigem Mittelwerten

ESDM 3	Frq.	Percent	Valid Percent	Cum. Percent	ESDM 11	Frq.	Percent	Valid Percent	Cum. Percent
0 (fast nie)	1568	77,1	82,7	82,7	0 (fast nie)	1560	76,7	83,3	83,3
1 (manchmal)	289	14,2	15,3	98,0	1: (manchmal)	297	14,6	15,9	99,1
2 (immer)	38	1,9	2,0	100,0	2 (immer)	16	,8	,9	100,0
Missing Value	139	6,8			Missing Value	161	7,9		
Total	2034	100,0			Total	2034	100,0		

*** Frq. = Frequency, Cum. Percent = Cumulative Percent

Der Tabelle ist zunächst die Information zu entnehmen, die über die zuvor festgestellten Mittelwerte angekündigt worden ist. Die befragten Lehrkräfte scheinen selten Kindern zu begegnen, die zu Selbstvorwürfen neigen und ebenso selten haben sie Kinder beobachtet, die man oft allein sieht. Für diese beiden Items wäre nun wiederum die Diskussion zu führen, die bereits bei der Kenntnisnahme von Mittelwerten für Items aus DTK und CDI zu führen war. In Verkürzung der Erörterungen wird hier gefolgert, dass die Gesamtstichprobe wenig Kinder enthält, die dem Item 3 und dem Item 11 zuzuordnen sind. Damit ist aber nicht darüber entschieden, ob nicht unter den Kindern, die das Verfahren zu identifizieren sucht, also unter den Kindern mit hohen Frequenzen an depressiven Symptomen und entsprechenden Gesamtwerten, nicht auch höhere Frequenzen für Item 3 und 11 auftreten. Wenn der Beitrag dieser Items zur Profilierung höher als nach dem Mittelwert erwartbar wäre, würde sich das in weiteren Kennwerten widerspiegeln. Deswegen sind dann auch die nachfolgenden Aufschlüsse entsprechend aussagekräftig.

Die Korrelationen der Einzelitems mit dem Gesamtwert des ESDM weisen im Allgemeinen zufrieden stellende Trennschärfen auf. Die niedrigste Trennschärfe liegt nun jedoch nicht bei den Items 3 und 11, den niedrigsten Wert ($r = 0,28$) zeigt Item 10 („*widerspricht und streitet*“). Die nachstehende Tabelle stellt zum einen die Trennschärfe der ESDM-Items aus der griechischen Stichprobe (Kampisiou, 2003) dar. Zum anderen zitiert sie die entsprechenden Werte aus einer vorangegangenen deutschsprachigen Anwendung. Das ist die Stichprobe von Casanovas (2003), die das Verfahren ESDM bei Lehrkräften von spanischen Kindern in Deutschland angewendet hat.

Tabelle 6.22: Trennschärfe der ESDM-Items in zwei Untersuchungen

Item Nr.	Kampisiou (2003)	Casanovas (2003)
ESDM 1	0,60	0,55
ESDM 2	0,74	0,66
ESDM 3	0,56	0,50
ESDM 4	0,31	0,39
ESDM 5	0,72	0,69
ESDM 6	0,57	0,66
ESDM 7	0,56	0,45
ESDM 8	0,54	0,52
ESDM 9	0,70	0,72
ESDM 10	0,28	0,47
ESDM 11	0,44	0,54
ESDM 12	0,64	0,57
ESDM 13	0,59	0,54
ESDM 14	0,76	0,70
ESDM 15	0,75	0,64
ESDM 16	0,59	0,54

Das Signifikanzniveau für alle Items ist $p=0.001$

Die Trennschärfen aus den zwei in Tabelle 6.22 dargestellten Stichproben sind für die meisten Items vergleichbar. Ausnahme ist das Item 10, das in der spanischen Stichprobe eine mittlere Trennschärfe aufweist. Der Abfall dieses Items ist auffällig. Aber zu mehr als zu einem Hinweis darauf reicht der Informationsstand über die Ursachen dafür nicht aus.

6.4.2.2 Interne Konsistenz der ESDM-Skala

Die Reliabilität der ESDM wurde wiederum mittels Cronbach's Alpha gerechnet.

Tabelle 6.23: Reliabilität der ESDM-Skala

	Cronbach's Alpha	N of Items
	,876	16

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
es1	5,87178	25,180	,526	,868
es2	5,63397	23,346	,665	,861
es3	6,12658	25,932	,497	,870
es4	6,11452	27,130	,241	,878
es5	5,82904	23,925	,656	,862
es6	6,04767	25,460	,495	,869
es7	5,89096	25,433	,478	,870
es8	6,05534	25,663	,468	,870
es9	5,89589	24,387	,642	,863
es10	5,94904	26,964	,172	,883
es11	6,13973	26,619	,376	,874
es12	5,80329	24,773	,563	,866
es13	5,85973	25,157	,521	,868
es14	5,73589	23,198	,688	,860
es15	5,69096	23,078	,684	,860
es16	6,12274	25,808	,527	,869

Die voran gegangene Tabelle weist mit den Alpha Werten nach, dass die hier zuvor erörterten Kennwerte der einzelnen Items jeweils für sich betrachtet, als problematisch betrachtet werden konnten. Dessen ungeachtet ist ihre Funktionalität innerhalb des Verfahrens unübersehbar. Der mit der griechischen Übersetzung der Skala ESDM erreichte Reliabilitätswert von insgesamt 0,88 erlaubt, dem griechischen Verfahren als Forschungsinstrument durchaus gute Messeigenschaften zuzuschreiben.

Der erreichte Alpha Wert erreicht, wie der nachstehende Vergleich erweist, die entsprechenden Werte aus Studien, die mit spanischen Lehrkräften spanischer Kinder in Spanien und in Deutschland durchgeführt worden sind.

Die innere Konsistenz der ESDM Skala erreichte beispielsweise in der Studie von Casanova (2003) den Alpha-Wert 0,87. Weitere Werte zeigt die nachstehende Tabelle, die auf Angaben von Casanovas (2003) zurückgeht.

Tabelle 6.24: Reliabilitätskennwerte aus Studien mit dem ESDM (Informationen aus Casanovas, 2003)

Autoren	Stichprobe	Alpha-Wert der ESDM-Skala
Ezpeleta (1987)	1221 Kinder aus Spanien	0.91
Monreal (1988)	223 Kinder aus Spanien	0.89
Domènech & Polaino (1990)	1046 Kinder aus Spanien	0.88
Mestre, Garcia, Frias & Moreno (1993)	754 Jugendliche aus Spanien	0.87
Domènech & Allaga (1993)	961 Kinder aus Spanien	0.87
Casanova (2003)	167 spanische Kinder in Deutschland	0.87
Kampisiou (2003)	2034 Kinder und Jugendlichen in Griechenland	0.88

Über diesen Vergleich ist nachgewiesen, dass die ins Griechische übersetzte Version der Skala ESDM mit dem für sie gemessenen Alpha Wert das Niveau der in anderen Untersuchungen erreichten Werte erreicht. Danach wäre den mit der griechischen Version erhobenen Daten messtechnisch ebenso zu vertrauen wie den Daten aus den anderen Untersuchungen.

6.4.2.3 Faktorielle Struktur des ESDM

Die Faktorenanalyse wird unter anderem als hypothesengenerierendes Verfahren geschätzt. In der von Casanovas (2003) unter Einsatz einer deutschsprachigen Version der Skala ESDM geführten Untersuchung, wird diese hypothesengenerierende Funktion der Faktorenanalyse intensiv und entsprechend erfolgreich genutzt. Demnach ist die Vermutung gerechtfertigt, dass Lehrkräfte ihre eigenen und die Kinder der eigenen Schulklasse betreffenden Wahrnehmungen weniger in Abhängigkeit von klinischen als viel mehr in berufsabhängig entwickelten Urteilkategorien interpretieren. Die nachfolgend zu betrachtenden Ergebnisse einer Faktorenanalyse der mit der Skala ESDM gewonnenen Daten aus der griechischen Untersuchung von (Kampisiou, 2003) geben Gelegenheit, die ermittelte Struktur zur Kenntnis zu nehmen, sie mit der von Casanovas gefundenen zu vergleichen und bieten auch Gelegenheit, die Hypothesen über schulrelevanten Bezüge einer solchen Struktur zu überdenken.

Zur Ermittlung der Faktorenstruktur des ESDM wurde eine Hauptkomponentenanalyse mit anschließender Varimax-Rotation durchgeführt.

Tabelle 6.25: Auf Items bezogene Ergebnisse aus der Faktorenanalyse des ESDM in seiner griechischen Fassung (Kampisiou, 2003)

Communalities			Total Variance Explained			
	Initial	Extraction	Component	Initial Eigenvalues Total	% of Variance	Cumulative %
ESDM 1	1,000	,589	1	5,794	36,212	36,212
ESDM 2	1,000	,752	2	1,839	11,494	47,706
ESDM 3	1,000	,645	3	1,424	8,902	56,607
ESDM 4	1,000	,471	4	1,094	6,835	63,443
ESDM 5	1,000	,744	5	,789	4,928	68,371
ESDM 6	1,000	,610	6	,734	4,587	72,958
ESDM 7	1,000	,469	7	,659	4,118	77,076
ESDM 8	1,000	,548	8	,600	3,748	80,823
ESDM 9	1,000	,610	9	,531	3,318	84,141
ESDM 10	1,000	,847	10	,489	3,057	87,198
ESDM 11	1,000	,487	11	,459	2,869	90,067
ESDM 12	1,000	,513	12	,426	2,663	92,730
ESDM 13	1,000	,593	13	,388	2,424	95,154
ESDM 14	1,000	,768	14	,305	1,907	97,061
ESDM 15	1,000	,835	15	,287	1,791	98,853
ESDM 16	1,000	,670	16	,184	1,147	100,000

Extraction Method: Principal Component Analysis.

Die Anwendung des Scree-Tests führte bei Anlegung eines Eigenwertkriteriums (Eigenwert > 1) zu einer 4-Faktoren-Lösung, die insgesamt 63,44% der Varianz aufklärt, wobei allein der erste Faktor 36,2% der Gesamtvarianz erfasst. Inhaltlich sind die Items harmonisch gestellt, und so können die sich ergebenden Faktoren sinnvoll interpretiert werden.

Die Fragen des ersten Faktors beziehen sich fast ausschließlich auf leistungsbezogenen Äußerungen der Kinder. Das scheint indessen nicht auf die neunte Frage anwendbar, nach der das betreffende Kind durch die Lehrkraft wie folgt eingestuft wird. Es: „*fühlt sich weniger Wert*“. Wenn dies in inhaltlicher Übereinstimmung mit Items steht, die ausschließlich die schulische Leistung thematisieren, ist eine entsprechende Hypothese gerechtfertigt. „Die Lehrkräfte interpretieren das Selbstwertgefühl eines Kindes ihrer Schulklasse als vor allem mit seinen schulischen Leistungen variierende.“ Diese Hypothese scheint durch Alltagserfahrungen aus dem Schul- und dem Familienleben gerechtfertigt.

Der zweite Faktor steht mit Fragen in Verbindung, die Aspekte des Verhaltens der zu beurteilenden Kinder betreffen. Alle sechs Items dieses Faktors stehen in Beziehung zur Dysthymia. Schulisch sind diese Items auf Verhaltensweisen bezogen, die dem leistungsorientierten Ablauf des Unterrichts störend entgegen stehen. Solche Kinder sind pädagogisch eher „schwierige Kinder“, die anscheinend grundlos von dem pädagogischen Ideal eines fröhlichen, aktiven und lernbegierigen Kindes abweichen. Zugleich stellen die entsprechenden Aussagen über das gehäufte Auftreten dieser Aspekte in Verbindung mit Behinderungen oder Störungen des pädagogisch angestrebten Unterrichtsablaufes. Traurigkeit (Item 1), schwankende emotionale Befindlichkeit (13), Weinen (7), Unentschlossenheit (4), Müdigkeit (12) und Schuldgefühle (11) stehen dem zielgerichteten Interaktionsfluss schulischen Unterrichts entgegen. Treten sie vermehrt auf, stören diese Eigenheiten.

Die Ergebnisse der Faktorenanalyse dienen als Begründungen für Hypothesen. An anderer Stelle der Ausführungen über die Untersuchung 2034GR, im Kapitel über die Analyse der Daten aus dem ESDM, wird dem Erklärungswert diesen Hypothesen nachgegangen.

Der dritte Faktor bezieht sich deutlich auf das soziale Verhalten der Kinder in der Schule und speziell auf Interaktionen mit den Gleichjährigen.

Interessanterweise bildet die zehnte Frage des ESDM („*widerspricht und streitet*“) einen alleinigen Faktor, den vierten dieser Skala. Das ist zum einen bemerkenswert, gibt zum anderen aber auch Anlass zu der Nachfrage, ob denn diese Eigenheit von Kindern nicht auch schon einem der voran gegangenen drei Faktoren hätte assoziiert sein können. Der zweite Faktor wäre inhaltlich nicht ungeeignet gewesen, Der dritte Faktor ließe sich als noch eher geeignet denken. Dem ließe sich das Widersprechen und Streiten zuordnen. Das Antwortverhalten der Lehrkräfte stützte solche Faktorenzusammenordnungen nicht. Augenscheinlich ist für die Lehrkräfte das „Widersprechen und Streiten“ eine ganz spezielle Eigenheit von Schulkindern, die spezifisch, also ganz eigen zu interpretieren ist. Dieses Verhalten steht im schulischen Interaktionsgeschehen als eigens hervor zu hebende Verhaltenstendenz. Der kommt ein eigener Störungscharakter zu. Sie etabliert sich als eigener Faktor.

Die nachstehende Tabelle gibt eine Übersicht über die ermittelte Faktorenstruktur.

Tabelle 6.26: Die Faktorenstruktur der griechischen ESDM

ESDM-Item	Faktor 1	Faktor 2	Faktor 3	Faktor 4
15: Leistung	.88			
5: Verständnis des Unterrichts	.83			
14: aktive Teilnahme in der Klasse	.82			
2: Aufmerksamkeit	.82			
9: Selbstwert	.56			
1: Traurigkeit		.70		
13: schwankende Stimmung		.66		
11: Schuldgefühle		.65		
7: Müdigkeit		.64		
4: Weinen		.62		
12: Unentschlossenheit	.44	.55		
16: gemeinsames Spielen			.76	
3: Alleinsein in der Pause			.75	
6: Akzeptanz in der Klasse			.71	
8: Amüsieren in der Schule			.70	
10: Streiten				.91

Ladungen kleiner als .40 werden nicht angeführt

Die Durchführung von Faktorenanalysen in unterschiedlichen Teilgruppen der griechischen Stichprobe (Geschlecht der Lehrkräfte, Geschlecht der Kinder und Kombinationen des Geschlechts der Lehrkräfte mit dem Geschlecht der Kinder) bestätigten im Allgemeinen die für die Gesamtstichprobe gefundene Faktorenlösung. Einige Fragen wurden in dieser Lösung geschlechtsspezifisch anders positioniert.

Zu einer 4-Faktoren-Lösung, die inhaltlich mehr Gemeinsamkeiten als Unterschiede in den ermittelten Strukturen aufweist, kamen auch Domènech (1985) und Casanovas (2003). Die jeweilige Zuordnung der ESDM-Items zu den gefundenen Faktoren für die Untersuchungen von Domènech, von Casanovas und von Kampisiou stellt Tabelle 6.27 dar.

Tabelle 6.27: Faktorenlösungen des ESDM in den Studien von Domènech (1985), Casanovas (2003) und Kampisiou (2003)

Domènech		Casanovas		Kampisiou	
Faktoren	Items	Faktoren	Items	Faktoren	Items
schulische Leistung	2, 5, 14, 15	Leistungsorientierung	2, 5, 9, 14, 15	Schulleistung	2, 5, 9, 14, 15
soziale Interaktion	3, 6, 8, 16	soziale Interaktion	3, 4, 6, 11, 16	soziales Verhalten	3, 6, 8, 16
„inhibierte“ Depression	7, 9, 11, 12	Dysphorische Stimmung	1, 7, 8, 12	Störendes Verhalten	1, 4, 7, 11, 12, 13
„anxiety“Depression	4, 10, 13	Störanfälligkeit	10, 13	Widerspruch und Streit	10

Die in Tabelle 6.27 dargestellten ESDM-Faktoren haben sich ungeachtet der Unterschiede in den Stichproben aus der Anwendung des Verfahrens in drei unterschiedlichen Lehrerpulationen ergeben, mit der dann weiterhin in unterschiedlichen Kulturen lebende Kinder zu beurteilen waren. Die schulischen Leistungen und das soziale Verhalten der Kinder in ihren Schulen, wie es vom ESDM erfasst wird, erscheinen begrifflich als kulturunabhängige Kriterien in der Wahrnehmung der Lehrkräfte etabliert zu sein. Dessen ungeachtet könnten die Inhalte der schulisch vertretenen Forderungen durchaus variieren. Die in der hier zuvor dargestellten Tabelle zu beobachtenden Übereinstimmungen und Abweichungen der ESDM-Items in ihrer faktoriellen Zuordnung könnten jeweils Anlass für spezielle Untersuchungen werden, die aber innerhalb der hier geführten Erörterung nicht aufgenommen wird.

6.4.3 Zur Vertrauenswürdigkeit der mit dem ESDM in der griechischen Anwendung von Kampisiou erhobenen Daten

Die Übersetzung des ursprünglich durch Domènech und Mitarbeiter in spanischer Sprache konstruierten Skala ESDM (Domènech, 1985) ins Griechische verlangt, ebenso wie die Übersetzungen der ebenfalls benutzten Verfahren DTK und CDI, die Überprüfung der Messqualitäten dieser Version. Die griechische Fassung sollte die mit der spanischen Version und der deutschen Version von Casanovas (2003) vorgegebenen Standards zur Objektivität, Reliabilität und Validität einstellen. Wenn dieser Nachweis geführt sein sollte, erlaubte das eine entsprechende Feststellung zur Vertrauenswürdigkeit der mit der griechischen Version gewonnenen Daten.

Die hier zuvor mitgeteilten Daten lassen keinen Zweifel an der Reliabilität der griechischen Version dieser Skala. Die Validität steht insoweit nicht in Zweifel, als die ESDM Daten mit Sicherheit die Wahrnehmungsurteile der befragten Lehrkräfte wiedergeben. Inwieweit diese Urteile für die griechische Version nachgewiesene Beziehungen zu einem diagnostizierten klinischen Status von Depression bei Schulkindern haben sollten, ist empirisch zu prüfen und verbleibt zunächst als Aufgabe.

6.5 Messqualitäten des SDQ

Der SDQ ist ein von Goodmann (1997) konstruiertes Verfahren, mit dessen Hilfe spezielle Aspekte psychischer Besonderheiten bei Kindern und Jugendlichen erfasst werden sollen. Diese Zielsetzungen stellen sich selbst durch die unterschiedlichen Skalen des Verfahrens dar. Der SDQ hat eine (1.) Prosoziale Skala (1. „Prosocial Scale“), dann (2.) eine Hyperaktivitätsskala („Hyperactivity Scale“) und weiterhin (3.) die Skala emotionaler Symptome (3. „Emotional Symptoms Scale“), (4.) eine Skala Verhaltensführung (4. „Conduct Problem Scale“) und (5.) die Skala „Probleme unter Gleichaltrigen“ (5. „Peer Problems Scale“). Jede dieser fünf Skalen weist nur fünf Items auf. Diese Konstruktion verweist auf spezielle Besonderheiten des Verfahrens. Die Heterogenität der mit den Skalen angesprochenen Verhaltensaspekte und die Kurzskalen weisen den SDQ als ein Instrument aus, das der schnellen und ressourcenfreundlichen Orientierung über die in Persönlichkeiten und dann aber auch in Stichproben dominierenden Verhaltensaspekte. Sowohl die Kurzskalen als auch die Heterogenität des Verfahrens werden dessen Messqualitäten belasten. Die kurzen Skalen zielen auf kleine Stichproben an Selbstzuordnungen. Diese relativ kleinen Stichproben begünstigen das Auftreten extremer Abweichungen vom Mittelwert der Population.

Die Anwendung des SDQ in der Untersuchung 2034GR folgt anderen Zielsetzungen als die Einbeziehung von DTK, CDI und ESDM. Bereits die Ausgangssituation vor der Durchführung der Untersuchung 2034GR ist eine andere als die bei den genannten weiteren Verfahren. Der SDQ ist im Gegensatz zu DTK, CDI und ESDM in Griechenland bereits durch Bibou-Nakou, Kioseoglou und Stogiannidou (2001) und Kypraiou (2002) eingesetzt worden.

Insoweit hat die erneute Anwendung dieser bereits vorhandenen griechischen Version zum einen das Ziel der Überprüfung der bekannt gewordenen Messqualitäten des Verfahrens durch erneute Verwendung in einer größeren Stichprobe. Dann ist mit dieser Anwendung weiterhin die Absicht verbunden, die möglicherweise vorhandene Übereinstimmung der einzelnen Skalen mit den Daten aus DTK, CDI und ESDM zu prüfen. Weiterhin sollen die Daten aus dem SDQ geprüft werden, ob die Skalen messtechnisch verlässliche Hinweise auf etwa bestehende Komorbidität mit hohen Gesamtwerten in DTK, CDI oder ESDM zu geben.

Diesen Zwecken kann der SDQ in der Untersuchung HE1034 nur dienen, wenn er auch in dieser Untersuchung Messqualitäten erreicht, die den theoretischen Anforderungen nachkommen. Ob dem so ist, wird nachfolgend überprüft.

6.5.1 Messqualitäten des SDQ in voran gegangenen Untersuchungen

Der SDQ ist in zahlreichen englischsprachigen Untersuchungen verwendet worden. Ausführlicher Darstellungen zu den Messqualitäten des SDQ finden sich in Kapitel 11.

Wie schon zuvor mitgeteilt, ist der SDQ in Untersuchungen von Bibou-Nakou, Kioseoglou und Stogiannidou (2001) und Kypraiou (2002) in griechischen Versionen und in Stichproben mit griechischen Schulkindern eingesetzt worden. Dabei kam eine Übersetzung von Bibou, Stogiannidou, Kioseoglou und Papageorgiou zur Anwendung. Erfasst wurden bei diesen Untersuchungen sowohl der Gesamtpunktwert des Verfahrens als auch die jeweiligen Summenwerte aus den Skalen 1 bis 5. Die Stichproben umfassten Schulkinder ab 12 Jahren. Auf die aus diesen Untersuchungen vorliegenden Informationen zu den Messqualitäten des Verfahrens wird bei der nachfolgenden Prüfung der Reliabilität des SDQ in der Untersuchung 2034GR eingegangen werden.

6.5.2 Messqualitäten des SDQ bei seiner griechischen Anwendung in der Untersuchung 2034GR

Die mit dem Einsatz des SDQ in der Untersuchung 2034GR verfolgten Zielsetzungen sind in den vorausgegangenen Ausführungen (s. zuvor unter 8.5) genannt worden. Die Objektivität des Verfahrens sollte der englischsprachigen Version entsprechen. Hinweise, die daran zweifeln ließen, sind nicht bekannt geworden. In ähnlicher Weise stützen sich die Validitätserwartungen auf die Bewährung der englischsprachigen Versionen. Im folgen ist zu prüfen, mit welcher Genauigkeit die in der Untersuchung 2034GR verwendete griechische Version misst. Dem dient die nachfolgende Inspektion der für die Reliabilität heranzuziehenden Werte.

6.5.2.1 Itemanalyse des SDQ

In der nachstehenden Tabelle 6.28 sind die Mittelwerte, die Standardabweichung und die Varianz jedes einzelnen Items des SDQ, geordnet nach den Skalen des Verfahrens, aufgeführt. Diese Werte erlauben eine erste Einschätzung und kritische Wichtung der Itemqualitäten.

Tabelle 6.28: Kennwerte der griechischen SDQ-Items nach Subskalen geordnet (Kampisiou, 2003)

1. „Prosocial Scale“	N	Mean	Std. Deviation	Variance
SDQ 9	1033	1,67	,55	,30
SDQ 20	1033	1,62	,55	,30
SDQ 17	1033	1,60	,63	,40
SDQ 1	1033	1,50	,63	,39
SDQ 4	1033	1,40	,69	,48
Valid N (listwise)	1033			
2. „Hyperactivity Scale“	N	Mean	Std. Deviation	Variance
SDQ 15	1033	,82	,72	,52
SDQ 2	1033	,78	,75	,60
SDQ 10	1032	,64	,74	,54
SDQ 25	1033	,63	,63	,40
SDQ 21	1033	,44	,59	,34
Valid N (listwise)	1032			
3. „Emotional Symptoms Scale“	N	Mean	Std. Deviation	Variance
SDQ 8	1032	,85	,76	,58
SDQ 16	1033	,77	,75	,56
SDQ 13	1033	,53	,70	,50
SDQ 24	1033	,50	,69	,47
SDQ 3	1033	,44	,64	,41
Valid N (listwise)	1032			

(Fortführung Tabelle 6.28)

4. „Conduct Problem Scale“	N	Mean	Std. Deviation	Variance
SDQ 7	1031	,93	,66	,43
SDQ 5	1032	,89	,76	,57
SDQ 12	1033	,43	,65	,42
SDQ 18	1033	,36	,62	,38
SDQ 22	1032	,15	,43	,19
Valid N (listwise)	1030			

5. „Peer Problems Scale“	N	Mean	Std. Deviation	Variance
SDQ 23	1032	,58	,70	,49
SDQ 14	1033	,51	,61	,37
SDQ 6	1032	,35	,62	,39
SDQ 19	1033	,34	,61	,38
SDQ 11	1033	,28	,54	,29
Valid N (listwise)	1031			

Aus dieser Itemanalyse ergibt sich, dass Item 22 („*Ich nehme Dinge, die mir nicht gehören*“ („von zu Hause, in der Schule oder anderswo“), das zu der Conduct Problem Scale gehört, einen niedrigeren Mittelwert als der theoretisch mindestens erwartete von .20 aufweist. Umgekehrt haben zwei Items derselben Subskala (Item 5 „Ich werde leicht wütend; ich verliere oft meine Beherrschung“ und Item 7 „Normalerweise tue ich, was man mir sagt“) einen Mittelwert größer als .80. Letzteres gilt auch für Item 8 („Ich mache mir häufig Sorgen“) auf der Emotional Symptoms Scale und für Item 15 („Ich lasse mich leicht ablenken; ich finde es schwer, mich zu konzentrieren“), das der Hyperactivity Scale zugeordnet ist. An anderer Stelle war allerdings mehrfach nachzuweisen gewesen, dass auch solche Items, nachgewiesen über weitere Indikatoren, durchaus ihre Funktionalität innerhalb der Skala haben können. So wird dieser Wert zunächst nur zur Kenntnis genommen.

Für alle diese Items, die einen Mittelwert außer des erwünschten Wertbereiches aufweisen, und deshalb weniger informationsreich zu sein erscheinen, sind die nachstehenden Häufigkeitsauszählungen angefertigt worden.

Tabelle 6.29: Häufigkeitsauszählungen für Items mit niedrigem oder hohem Mittelwert

SDQ 5	Frq.	Percent	Valid Percent	Cumul. Percent	SDQ7	Frq.	Percent	Valid Percent	Cumul. Percent
0	358	17,6	34,7	34,7	0	264	13,0	25,6	25,6
1	428	21,0	41,5	76,2	1	580	28,5	56,3	81,9
2	246	12,1	23,8	100,0	2	187	9,2	18,1	100,0
Total	1032	50,7	100,0		Total	1031	50,7	100,0	
Missing System	1002	49,3			Missing System	1003	49,3		
Total	2034	100,0			Total	2034	100,0		

(Fortführung Tabelle 6.29)

SDQ 8	Frq.	Percent	Valid Percent	Cumul. Percent
0	385	18,9	37,3	37,3
1	415	20,4	40,2	77,5
2	232	11,4	22,5	100,0
Total	1032	50,7	100,0	
Missing System	1002	49,3		
Total	2034	100,0		

SDQ22	Frq.	Percent	Valid Percent	Cumul. Percent
0	914	44,9	88,6	88,6
1	86	4,2	8,3	96,9
2	32	1,6	3,1	100,0
Total	1032	50,7	100,0	
Missing System	1002	49,3		
Total	2034	100,0		

SDQ15	Frq.	Percent	Valid Percent	Cumul. Percent
0	379	18,6	36,7	36,7
1	463	22,8	44,8	81,5
2	191	9,4	18,5	100,0
Total	1033	50,8	100,0	
Missing System	1001	49,2		
Total	2034	100,0		

*** Frq. = Frequency, Cum. Percent = Cumulative Percent

In der folgenden Tabelle 6.30 ist die Trennschärfe jedes einzelnen Items mit dem Score der dazugehörigen Skala wie auch mit dem Score der gesamten Skala aufgeführt worden. Letzteres betrifft nur die Items, die den vier problembezogenen Skalen entsprechen, und nicht die Items der „Prosocial Scale“ (weil sie, als eine Skala, die „positive“ Eigenschaften der Kinder zu erfassen versucht, nicht in den gesamten Score, der eine Messung von Problemen und Schwierigkeiten repräsentiert, mitgerechnet wird).

Tabelle 6.30: Trennschärfe der SDQ-Items

Hyperactivity Scale	r1	r2
SDQ 2	,43	,68
SDQ 10	,48	,63
SDQ 15	,52	,66
SDQ 21	,40	,56
SDQ 25	,36	,55

Emotional Problems Scale	r1	r2
SDQ 3	,42	,62
SDQ 8	,48	,71
SDQ 13	,60	,73
SDQ 16	,53	,64
SDQ 24	,50	,67

Conduct Problems Scale	r1	r2
SDQ 5	,51	,67
SDQ 7	,17	,44
SDQ 12	,40	,62
SDQ 18	,46	,57
SDQ 22	,33	,43

Peer Problems Scale	r1	r2
SDQ 6	,49	,62
SDQ 11	,37	,58
SDQ 14	,40	,64
SDQ 19	,50	,64
SDQ 23	,40	,61

Prosocial Scale	r1	r2
SDQ 1	-	,58
SDQ 4	-	,67
SDQ 9	-	,65
SDQ 17	-	,58
SDQ 20	-	,69

Alle Korrelationen sind signifikant auf $p < 0,05$;

r1: Korrelation mit dem Gesamtscore des SDQ,

r2: Korrelation mit dem Gesamtscore der jeweiligen Subskala

Die mitgeteilten Korrelationen lassen erkennen, dass die einzelnen Items durchaus unterschiedliche Beiträge zu ihren Skalenwerten leisten. Ebenso unterschiedlich sind die Anteile an dem Gesamtwert des Verfahrens. Er wäre als Indikator für psychische Auffälligkeiten zu verstehen, deren Beziehungen zu DSM-IV oder ICD-10 noch in Frage stehen.

6.5.2.2 Interne Konsistenz

Wie sich die einzelnen Items in ihrer Messqualität bewähren, ist an den nachstehend aufgeführten Werten abzusehen.

Die erste der nachstehenden Tabellen führt auf, welche Reliabilitätswerte für die aus vier Subskalen gebildete Skala der Gesamtproblematik („*Total Difficulties*“) ergibt.

**Tabelle 6.31: Reliabilität der gesamten SDQ-Skala „Total Difficulties“
(Summwert aus vier Skalen)**

	Cronbach's Alpha	N of Items
	,777	20

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
good2	10,39359	30,373	,314	,770
good3	10,73081	30,856	,319	,769
good5	10,28377	29,626	,406	,763
good6	10,82799	30,693	,356	,767
good7	10,24976	32,680	,058	,786
good8	10,32459	29,877	,373	,766
good10	10,53936	30,023	,372	,766
good11	10,89699	31,511	,286	,771
good12	10,74927	30,993	,296	,771
good13	10,65112	29,177	,514	,755
good14	10,67055	31,108	,306	,770
good15	10,35763	29,722	,421	,762
good16	10,40816	29,516	,429	,761
good18	10,81341	30,652	,368	,766
good19	10,83868	30,398	,414	,763
good21	10,73761	31,194	,307	,770
good22	11,02915	32,069	,264	,773
good23	10,59281	30,827	,285	,772
good24	10,67930	30,068	,399	,764
good25	10,54908	31,304	,264	,773

Der erreichte Reliabilitätswert von .78 für die Gesamtskala (*Total Difficulties Scale*) ist hinreichend für die Verwendung des Instrumentes in empirischen Untersuchungen.

Im Weiteren wurden die Höhe der Reliabilität jeder einzelner Subskala des Verfahrens für die Gesamtstichprobe und im Hinblick auf das Geschlecht der Kinder berechnet. Diese Informationen sind in Tabelle 6.32 dargestellt.

Tabelle 6.32: Alpha-Werte für die 5 Subskalen und die Gesamtskala des SDQ in der Gesamtstichprobe und nach Geschlecht der Kinder (Kampisiou, 2003)

	Gesamtstichprobe	N	Knaben	N	Mädchen	N
Prosocial Scale	.62	1033	.60	528	.58	505
Hyperactivity Scale	.60	1032	.54	528	.64	504
Emotional Problems Scale	.70	1032	.64	527	.71	505
Conduct Problems Scale	.43	1030	.42	526	.42	504
Peer Problems Scale	.60	1031	.58	527	.61	504
Total Difficulties Scale	.78	1029	.77	526	.79	503

Die in der vorangegangenen Tabelle aufgeführten Alpha-Werte liegen sowohl unter als auch über der allgemein erhobenen Mindestanforderung von 0,70 für ein Forschungsinstrument. Dieser Normwert wird zum einen für die Gesamtskala überschritten (0,78). Zum anderen erreicht die Subskala „Emotionale Probleme“ diesen Wert, dies allerdings nur durch Einstellung des Limits von 0,70.

Die Reliabilität des SDQ erreicht für die Gesamtskala, die mit ihren Gesamtwerten als Indikator für das Bestehen psychischer Probleme gelten darf, mit einem Alpha Wert von 0,78 das für Forschungsinstrumente verlangte Messniveau.

Von den vier Subskalen für psychische Probleme erreicht nur die Skala für emotionale Probleme noch eben das Ausschlusskriterium von Alpha 0,70. Dieser Wert wurde auch nur durch die Messungen in der Teilstichprobe der Mädchen stabilisiert. Für die Teilstichprobe der Jungen sinkt der Alpha Wert bereits unter das geforderte Niveau.

Die drei anderen Skalen haben die für ein Forschungsinstrument gesetzte Norm von Alpha 0,70 insgesamt verfehlt.

Nach den Daten aus der Untersuchung von Kampisiou (2003) ist die Verlässlichkeit der Daten aus dem SDQ nur für die Gesamtskala und nicht für die Subskalen anzunehmen. Das gilt auch für die Skala der emotionalen Probleme, weil deren höchster Alpha Wert Stichprobenänderungen variiert und unter 0,70 fällt.

Die *Emotional Problems Scale* hat von den Subskalen in der Gesamtstichprobe die höchste Reliabilität (0,70), dennoch sind eventuell auffindbare Beziehungen zu Werten aus dem DTK, CDI oder ESDM wegen der deutlichen Alpha Schwäche kritisch zu beurteilen.

Die *Conduct Problems Scale*, mit dem außerordentlich niedrigen Reliabilitätswert von 0,42 muss als messtheoretisch eingestuft werden.

Die Magisterarbeit von Kypraiou (2002) über die Einschätzung der Schwierigkeiten und Möglichkeiten von Kindern beim Übergang zur Pubertät aus der Perspektive von Jugendlichen, Eltern und Lehrer stellt die einzige veröffentlichte Quelle dar, wo begrenzte Informationen über die Reliabilität der griechischen SDQ-Self Report Version zu finden sind. Aus dieser Studie bei 861 Kindern (291 Kindern mit dem Altersmittelwert von 10 Jahren und 5 Monaten und 570 Kindern mit dem Mittelwert von 13 Jahren und 8 Monaten, beide Stichproben sind aus Thessaloniki, haben sich folgende Alpha-Werte ergeben: für die *Prosocial Scala* einer Alpha-Wert von 0,52, für die *Hyperactivity Scale* 0,61, *Emotional Problems Scale* 0,69, *Conduct Problems Scale* 0,42 und *Total Difficulties Scale* 0,69.

Vergleicht man die Alpha-Werte aus den zwei Studien, stellt man fest, dass in beiden Studien die *Hyperactivity Scale* und die *Emotional Problems Scale* die höchste Reliabilität aufweisen, wobei die *Conduct Problems Scale* die niedrigste bleibt. Aber insgesamt bestätigt sich das Ergebnis aus

der Untersuchung von Kampisiou (2003). Die Subskalen des SDQ haben keine hinreichenden Messqualitäten. Die Ursachen dafür können in den extrem kurzen Subskalen gesehen werden. Eine Skalenverlängerung würde Besserung in Aussicht stellen. Wenn in Interesse an der Steigerung der Messqualitäten des SDQ bestehen sollte, wäre zu empfehlen, die Subskalen empirisch kontrolliert zu verlängern.

Eine weitere Anwendung fand der SDQ in einer Untersuchung durch Bibou-Nakou, Kioseoglou & Stogiannidou (2001). An der Untersuchung waren $n = 622$ Schulkinder beteiligt. Die Alpha Werte für die Subskalen erreichten die folgenden Koeffizienten:

Prosocial scale = 0,693, Hyperaktivitätsskala = 0,570, Emotional problems = 0,701, Conduct problems = 0,448, Peer problems = 0,539. Der Vergleich mit den bisher mitgeteilten anderen Reliabilitätswerten bestätigt die zuvor getroffenen Feststellungen.

Tabelle 6.33: Reliabilitätswerte des SDQ in verschiedenen griechischen Untersuchungen

	Bibou-Nakou et al., (2001)	Kypraiou, 2002	Kampisiou, 2003
Prosocial Scale	.69	.52	.62
Hyperactivity Scale	.57	.61	.60
Emotional Problems Scale	.70	.69	.70
Conduct Problems Scale	.45	.42	.43
Peer Problems Scale	.54	f. A	.60
Total Difficulties Scale	f. A.	.69	.78

Zeichenerklärung: f. A.= fehlende Angaben

Die Bewertung der Reliabilität der griechischen Übersetzung des SDQ kann unter Kenntnisnahme der vorliegenden Daten zu keiner anderen als der zuvor bereits mitgeteilten Feststellung kommen.

Die Gesamtskala des SDQ erfüllt, ungeachtet der Antwort auf die Frage, was sie valide misst, die an ein Forschungsinstrument hinsichtlich der verlangten Reliabilitätswerte gestellten Mindestanforderungen. Den Daten aus der Anwendung der Gesamtskala darf insoweit Verlässlichkeit zugeschrieben werden.

Den Daten der Subskala emotionaler Probleme solle nur hypothesenstützende Verlässlichkeit zuerkannt werden. Deren Erklärungswert bliebe gesondert zu prüfen.

Die drei weiteren Subskalen sind messtechnisch unzuverlässig.

6.5.2.3 Faktorenanalyse aus der Anwendung des SDQ innerhalb dieser Untersuchung

Für die Gesamtskala des SDQ war, trotz gravierender Defizite für die Subskalen, eine hinreichende Messqualität festzustellen. Das legt, in Fortsetzung der Vorgehensweise bei den anderen Verfahren dieser Untersuchung, die Durchführung einer Faktorenanalyse auch für den SDQ nahe.

In einem ersten Schritt wurde eine Oblimin-Rotation (5 Faktoren) durchgeführt, um das ursprünglich dem Verfahren unterliegende hypothetische Konstrukt zu prüfen.

Tabelle 6.34: Structure Matrix für den SDQ nach der Anwendung in der Untersuchung Kampisiou (2003)

	<u>Component</u>				
	1	2	3	4	5
8 "Ich mache mir häufig Sorgen" (emotional)	,723				
13 "Ich bin oft unglücklich oder niedergeschlagen; ich muss häufig weinen" (emotional)	,700			-,365	
24 "Ich habe viele Ängste; ich fürchte mich leicht" (emotional)	,655				
3 "Ich habe oft Kopfschmerzen oder Bauchschmerzen, mir wird oft schlecht" (emotional)	,625				
16 "Neue Situationen machen mich nervös; ich verliere leicht das Selbstvertrauen" (emotional)	,578				
20 "Ich helfe anderen oft freiwillig (Eltern, Lehrern oder Gleichaltrigen)" (prosocial)		,723			
9 "Ich bin hilfsbereit, wenn andere verletzt, krank oder traurig sind" (prosocial)		,711			
4 "Ich teile normalerweise mit Anderen (z.B. Süßigkeiten, Spielzeug, Buntstifte)" (prosocial)		,632			
1 "Ich versuche, nett zu anderen Menschen zu sein, ihre Gefühle sind mir wichtig" (prosocial)		,527			
17 "Ich bin nett zu jüngeren Kindern" (prosocial)		,443			
2 "Ich bin oft unruhig, ich kann nicht lange stillsitzen" (hyperactivity)			,662		
10 "Ich bin dauernd in Bewegung und zappelig" (hyperactivity)			,643		
12 "Ich schlage mich häufig; ich kann Andere zwingen zu tun, was ich will" (conduct)			,561		
5 "Ich werde leicht wütend; ich verliere oft meine Beherrschung" (conduct)			,514		
15 "Ich lasse mich leicht ablenken; ich finde es schwer, mich zu konzentrieren" (hyperactivity)	,411		,422		,353
19 "Ich werde oft anderen gehänselt oder schikaniert" (peer)				-,679	
18 "Andere behaupten oft, dass ich lüge oder moegele" (conduct)			,397	-,552	
6 "Ich bin meistens für mich alleine; ich beschäftige mich lieber mit mir selbst" (peer)	,402			-,551	
11 "Ich habe einen oder mehrere gute Freunde oder Freundinnen" (peer)				-,543	,418
23 "Ich komme besser mit Erwachsenen aus als mit Gleichaltrigen" (peer)				-,512	
14 "Im allgemeinen bin ich bei Gleichaltrigen beliebt" (peer)				-,508	,408
22 "Ich nehme Dinge, die mir nicht gehören (von zu Hause, in der Schule oder anderswo)" (conduct)				-,480	
25 "Was ich angefangen habe, mache ich zu Ende; ich kann mich lange genug konzentrieren" (hyperactivity)					,704
21 "Ich denke nach, bevor ich handele" (hyperactivity)					,630
7 "Normalerweise tue ich, was man mir sagt" (conduct)					,500

Extraction Method: Principal Component Analysis. Rotation Method: Oblimin with Kaiser Normalization.

Für den SDQ war ursprünglich angenommen worden, dass die Skalen sich als eigene Faktoren manifestieren würden. Für die „Prosocial Scale“ und die „Emotional Problem Scale“ trifft das auch in dieser Anwendung zu. Die anderen Faktoren entsprechen nicht den ursprünglichen Skalenbildungen. Man merkt, dass die Items der Conduct Problems Scale eine Interpretation der sich einstellenden Struktur komplizieren.

Ein zweiter Schritt wäre, die 5 Items der Conduct Problems Scale zu eliminieren und eine neue Oblimin-Faktorenanalyse zu rechnen.

Tatsächlich ergeben sich dann, nach Eliminierung dieser Skala, nach einer neuen Oblimin-Rotation 4 Faktoren, die den 4 verbliebenen Subskalen des SDQ entsprechen.

Tabelle 6.35: Oblimin-Rotation ohne die Items der *Conduct Problems Scale*

Structure Matrix	Component			
	1	2	3	4
8 "Ich mache mir häufig Sorgen" (emotion)	,714			
24 "Ich habe viele Ängste; ich fürchte mich leicht" (emotional)	,671			
13 "Ich bin oft unglücklich oder niedergeschlagen; ich muss häufig weinen" (emotional)	,657			,413
3 "Ich habe oft Kopfschmerzen oder Bauchschmerzen, mir wird oft schlecht" (emotional)	,645			
16 "Neue Situationen machen mich nervös; ich verliere leicht das Selbstvertrauen" (emotional)	,554			
20 "Ich helfe anderen oft freiwillig (Eltern, Lehrern oder Gleichaltrigen)" (prosocial)		,740		
9 "Ich bin hilfsbereit, wenn andere verletzt, krank oder traurig sind" (prosocial)		,723		
4 "Ich teile normalerweise mit Anderen (z.B. Süßigkeiten, Spielzeug, Buntstifte)" (prosocial)		,630		
1 "Ich versuche, nett zu anderen Menschen zu sein, ihre Gefühle sind mir wichtig" (prosocial)		,515		
17 "Ich bin nett zu jüngeren Kindern" (prosocial)		,424	-,369	
2 "Ich bin oft unruhig, ich kann nicht lange stillsitzen" (hyperactivity)			,708	
10 "Ich bin dauernd in Bewegung und zappelig" (hyperactivity)			,618	
15 "Ich lasse mich leicht ablenken; ich finde es schwer, mich zu konzentrieren" (hyperactivity)	,369		,594	
21 "Ich denke nach, bevor ich handle" (hyperactivity)			,562	
25 "Was ich angefangen habe, mache ich zu Ende; ich kann mich lange genug konzentrieren" (hyperactivity)		-,357	,507	
14 "Im allgemeinen bin ich bei Gleichaltrigen beliebt" (peer)				,686
11 "Ich habe einen oder mehrere gute Freunde oder Freundinnen" (peer)				,665
19 "Ich werde oft anderen gehänselt oder schikaniert" (peer)				,584
6 "Ich bin meistens für mich alleine; ich beschäftige mich lieber mit mir selbst" (peer)	,387			,544
23 "Ich komme besser mit Erwachsenen aus als mit Gleichaltrigen" (peer)				,511

Extraction Method: Principal Component Analysis. Rotation Method: Oblimin with Kaiser Normalization.

Nach einer neuen Oblimin-Rotation ergeben sich 5 Faktoren. Drei von denen entsprechen auf akribische Weise den Subskalen „*Emotional Problems Scale*“, „*Prosocial Scale*“ und „*Peer Problems Scale*“.

Die weiteren zwei Faktoren bestehen aus Items der „*Hyperactivity Scale*“, sie bilden aber zwei getrennte Faktoren:

Der erste besteht aus den Fragen SDQ 2 "Ich bin oft unruhig, ich kann nicht lange stillsitzen", SDQ 10 "Ich bin dauernd in Bewegung und zappelig" und SDQ 15 "Ich lasse mich leicht ablenken; ich finde es schwer, mich zu konzentrieren".

Der zweite Faktor besteht aus den Fragen SDQ 25 "Was ich angefangen habe, mache ich zu Ende; ich kann mich lange genug konzentrieren" und SDQ 21 "Ich denke nach, bevor ich handele"

Eine denkbare Erklärung für diese Faktorenstruktur wäre, dass unterschiedliche Zuordnungsregeln befolgt werden. Die Items werden 1. nach unterschiedlichen Ausdrucksweisen (negativ-positiv) getrennt. Dann gilt weiterhin die Akzentuierung motorischer oder kognitiver Repräsentation von Hyperaktivität. Ausserhalb dieser Regel bleibt dann aber Item 15, das sich doch eher auf den kognitiven Ausdruck der Hyperaktivität bezieht.

Zu ähnlicher Ergebnis kommt man nach Anwendung einer Faktoranalyse mit Varimax-Rotation (wieder nur für die Items der 4 Subskalen).

Tabelle 6.36: Varimax- Rotation der SDQ Daten aus der Untersuchung Kampisiou (2003)

Rotated Component Matrix(a)	Component				
	1	2	3	4	5
8 "Ich mache mir häufig Sorgen" (emotion)	,714				
3 "Ich habe oft Kopfschmerzen oder Bauchschmerzen, mir wird oft schlecht" (emotional)	,670				
24 "Ich habe viele Ängste; ich fürchte mich leicht" (emotional)	,664				
13 "Ich bin oft unglücklich oder niedergeschlagen; ich muss häufig weinen" (emotional)	,617				
16 "Neue Situationen machen mich nervös; ich verliere leicht das Selbstvertrauen" (emotional)	,510				
9 "Ich bin hilfsbereit, wenn andere verletzt, krank oder traurig sind" (prosocial)		,739			
20 "Ich helfe anderen oft freiwillig (Eltern, Lehrern oder Gleichaltrigen)" (prosocial)		,720			
4 "Ich teile normalerweise mit Anderen (z.B. Süßigkeiten, Spielzeug, Buntstifte)" (prosocial)		,648			
1 "Ich versuche, nett zu anderen Menschen zu sein, ihre Gefühle sind mir wichtig" (prosocial)		,531			
17 "Ich bin nett zu jüngeren Kindern" (prosocial)		,374			
23 "Ich komme besser mit Erwachsenen aus als mit Gleichaltrigen" (peer)			,618		
11 "Ich habe einen oder mehrere gute Freunde oder Freundinnen" (peer)			,570		,364
6 "Ich bin meistens für mich alleine; ich beschäftige mich lieber mit mir selbst" (peer)			,568		
19 "Ich werde oft anderen gehänselt oder schikaniert" (peer)			,568		
14 "Im allgemeinen bin ich bei Gleichaltrigen beliebt" (peer)			,555		,392
2 "Ich bin oft unruhig, ich kann nicht lange stillsitzen" (hyperactivity)				,807	
10 "Ich bin dauernd in Bewegung und zappelig" (hyperactivity)				,682	
5 "Ich lasse mich leicht ablenken; ich finde es schwer, mich zu konzentrieren" (hyperactivity)				,505	
25 "Was ich angefangen habe, mache ich zu Ende; ich kann mich lange genug konzentrieren" (hyperactivity)					,731
21 "Ich denke nach, bevor ich handele" (hyperactivity)					,711

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.
a Rotation converged in 7 iterations.

Die Faktorenanalyse bestätigt die Skalenstruktur des SDQ nur eingeschränkt. Dieser Sachverhalt und im Übrigen dann auch die sehr heterogenen Reliabilitätswerte lassen den Schluss zu, dass, wenn Interesse am Erhalt des Instruments besteht, die griechische Version des SDQ einer Rekonstruktion unterzogen werden sollte.

6.5.3 Zur Vertrauenswürdigkeit der mit dem SDQ in dieser griechischen Anwendung erhobenen Daten

Im Vergleich zu der für die bisher vorgenommenen Prüfungen der Messqualität aller eingesetzter Verfahren vorgefundenen Datenlage ist die des SDQ außerordentlich instabil. Nur die Gesamtskala des SDQ erreicht Messqualitäten, die Vertrauen in die erhobenen Daten rechtfertigen. Die mit den Subskalen ermittelten Werte sind nicht verlässlich. Die Datenanalyse vermittelt Hinweise darauf, dass dieses Verfahren einer Überarbeitung bedarf. Teile des Verfahrens messen augenscheinlich noch relativ verlässlich. Einige Items differenzieren nicht mehr hinreichend. Ob die Übersetzung ins Griechische ursächlich beteiligt sein könnte, ist nicht entscheidbar. Das Defizit an Messqualitäten ist jedoch eindeutig nachweisbar.

Die Frage nach der Verlässlichkeit der mit dem SDQ erhobenen Daten ist, in Wiederholung der schon zuvor getroffenen Feststellungen, wie folgt zu beantworten:

Die Reliabilität des SDQ erreicht für die Gesamtskala, die mit ihren Gesamtwerten als Indikator für das Bestehen psychischer Probleme gelten darf, mit einem Alpha Wert von 0,78 das für Forschungsinstrumente verlangte Messniveau.

Von den vier Subskalen für psychische Probleme erreicht nur die Skala für emotionale Probleme noch eben das Ausschlusskriterium von Alpha 0,70. Dieser Wert wurde auch nur durch die Messungen in der Teilstichprobe der Mädchen stabilisiert. Für die Teilstichprobe der Jungen sinkt der Alpha Wert bereits unter das geforderte Niveau.

Die drei anderen Skalen haben die für ein Forschungsinstrument gesetzte Norm von Alpha 0,70 insgesamt verfehlt.

Nach den Daten aus der Untersuchung von Kampisiou (2003) ist die Verlässlichkeit der Daten aus dem SDQ nur für die Gesamtskala und nicht für die Subskalen anzunehmen. Das gilt auch für die Skala der emotionalen Probleme, weil deren höchster Alpha Wert Stichprobenänderungen variiert und unter 0,70 fällt.

6.6 Zu den Messqualitäten der eingesetzten Fragenliste FIL

Der Gebrauch des Begriffes Fragebogen ist (siehe Bortz und Döring, 2002, S. 678) in mehrfacher Hinsicht dem Begriff Test gleichgesetzt. Fragebögen werden demnach wie Tests konstruiert. Dem steht das recht alltägliche Befragen von Personen entgegen, bei dem nicht die psychometrischen Qualitäten einer Frage interessieren, sondern einzig die Antwort. Die Frage: „Wie heißt du?“ soll nicht zu Antworten führen, die einer Normalverteilung zuzuordnen sind. Die Frage: „Wo sind sie geboren?“ wird nicht so gestellt, dass ein mittlerer Schwierigkeitsgrad oder eine Mittelwertstreuung zwischen 0,20 und 0,80 gegeben sein sollte.

Für empirischen Untersuchungen muss auch die Möglichkeit eingeräumt werden, Fragen zu stellen, bei denen die Antwort interessiert. Bei einer Untersuchung in griechischen Schulen wäre eine solche Frage möglicherweise die, ob das Kind einer Flüchtlingsfamilie zugehört und woher diese Familie kommt.

Diese Möglichkeit besteht ja auch und ist methodisch noch nicht wirklich verdächtig. Hier wird darauf eingegangen, weil für eine solche Liste von Fragen ein Name gesucht wird. Finanzämter und Katasterämter pflegten den Sprachgebrauch der Erhebungsbögen. Das sind Bögen, auf den

Eigenheiten von Gebäuden, Betrieben oder Grundstücken systematisch abgefragt werden. Nicht die Differenzierungsleistung der Frage interessiert, sondern nur die Antwort auf die Frage. So kann man auch am Abfragen anderer Umstände interessiert sein, beispielsweise an der Frage, ob Vater arbeite oder arbeitslos sei, ob die Eltern zusammen oder getrennt lebten und was sonst noch von Interesse ist.

In dieser Untersuchung 2034GR sind 19 solcher Fragen gestellt worden. Da sie nicht über eine Itemkonstruktion generiert worden sind, sollten sie in ihrer Gesamtheit auch nicht als Fragebogen bezeichnet werden. An dieser Stelle wird für diese Liste von Fragen der Name Fragenliste verwendet. Das ist eine Liste von Fragen, die der Untersuchende oder die Untersuchende an die Menschen in der von ihm aufgesuchten Stichprobe richtet, weil er an der jeweiligen Antwort inhaltlich interessiert ist. Die Frage nach der Zufriedenheit mit einem Klassenlehrer oder einer Klassenlehrerin zielt nicht darauf ab, eine Gleichgewichtigkeit von Ja-Antworten und Nein-Antworten zu erreichen. Falls alle Befragten mit Ja antworteten, könnte das auch eine Untersuchungsleiterin erfreuen, obschon das Item nach klassischer Testtheorie, widersinnig genug, keinen wesentlichen Informationsgehalt transportierte.

Der Begriff Fragenliste steht hier für das, was der Name aussagt. Es ist eine Liste von Fragen, die gestellt werden, weil die jeweilige Antwort darauf von Interesse ist.

Kapitel 7

Analyse der mit dem DTK aufgenommenen Daten

Die vorangegangene Darstellung einiger Aspekte der Messqualitäten mehrerer in einer griechischen Stichprobe von 2034 Kindern eingesetzten psychologischen Messinstrumente war mit dem wissenschaftlich unvermeidlich zu setzenden Ziel verbunden, empirisch verlässlich zu klären, ob die aus messtheoretischer Sicht notwendigen Voraussetzungen für die weitere Beschäftigung mit den aus dem Einsatz der Messinstrumente gewonnenen Daten vorliegen. Die für den Fortgang der Arbeiten erfreulichen messtheoretischen Befunde erlauben nun, die Analyse der erarbeiteten Datensätze zu beginnen. Mit den Ergebnissen aus diesen Analysen kann versucht werden, eine möglichst zweifelsfreie Beantwortung der für diese Untersuchung eingesetzten speziellen Forschungsfragen vorzunehmen, die dann insgesamt einer schrittweisen Annäherung an ein übergeordnetes Forschungsziel dienen, dem einer epidemiologischen Populationsbeschreibung zum selbst berichteten Auftreten depressiver Symptome bei Kindern und Jugendlichen in Griechenland. Die wiederum sollte helfen, Aspekte einer die griechische Gesellschaft insgesamt betreffende Problematik realistisch zu erfassen, Aspekte der psychischen Gesundheit der in Griechenland aufwachsenden Kinder und Jugendlichen.

Bevor den mit jedem einzelnen Instrument gewonnenen Informationen nachgegangen wird, ist an dieser Stelle der Arbeit zunächst die Sichtweise zu betrachten und zu erläutern, mit der die aus der Durchführung dieser empirischen Studie gesammelten Daten betrachtet werden. Allgemein gelten Daten als der in seiner Aussagekraft nicht zu überbietende harte Kern der Realitätsdarstellung. Nach Popper sind Daten, allein für sich genommen, meinungslos, Menschen seien genötigt, Daten durch auf sie bezogene Theorien verständlich werden zu lassen.

Das empirisch begründete Vertrauen in die Messqualitäten des jeweils verwendeten Verfahrens ist zum einen unerlässlich und sein Vorliegen insoweit zu begrüßen, andererseits ist mit dem Vorliegen der gewünschten Messqualitäten allein noch keine einzige weiter reichende Aussage gewonnen worden. Nachdem die Messqualitäten gesichert sind, muss eine zielgerechte Datenanalyse angesetzt und durchgeführt werden, die aus den Datensätzen Informationen erschließt, mit denen die Realisierung der bis dahin nur antizipierten Zielvorstellungen versucht werden kann. Mit dem Blick auf diese Setzungen erweist sich das Ziel einer epidemiologischen Populationsbeschreibung als allgemein kennzeichnender Oberbegriff, der in dem speziellen Fall dieser Untersuchung zunächst nur eine gedachte Position beschreibt, die formal nur vage und inhaltlich überhaupt nicht gefüllt ist. Diese doppelte Beliebigkeit soll über die gewonnenen Datensätze aufgehoben werden, dazu bedarf der Begriff der epidemiologischen Populationsbeschreibung einer formalen Struktur und deren inhaltlicher Füllung mit Fakten und Aussagen.

Die formale Struktur einer epidemiologischen Populationsbeschreibung wird durch eine Theorie über sie vermittelt (Bortz und Döring, 2002, S. 397-488). Deren Anwendungsfall verlangt die situative Adaptation von Theorieinhalten zum speziellen Zweck der Untersuchung. Der Zweck der Untersuchung ist die Synthese empirisch gesicherter Fakten zu einer Aussagenstruktur, die als spezifische epidemiologische Populationsbeschreibung bezeichnet wird. Spezifisch ist sie, weil sie eine spezielle Population zu einem spezifischen Zweck beschreibt.

Die Bildung prüfbarer Hypothesen und die Auswahl für ihre Prüfung geeigneter statistischen Verfahren stellen beispielsweise einige Aspekte dar, die während des Analysenprozesses

berücksichtigen werden sollten. Zuvor ist aber notwendig, den Rahmen beschreiben zu lassen, in dem die Analyse der Daten aus jedem eingesetzten Verfahren durchgeführt werden darf. Dieser Rahmen wird zwangsläufig aus dem Objekt, den Subjekten und, um beide letztgenannten zu verbinden, mit den für die vorliegende Arbeit ausgewählten Methoden geschaffen.

Die Methode, die ausgewählt wurde, um sich den gesetzten Zielen annähern zu können, muss unmittelbar dem Ziel dienen können. Also ist das zuerst und vor jeder Methode zu nennen. In eben diesem Sinne nennt schon Aristoteles das antizipierte Ziel als Beginn einer jeden Handlung. In der Handlungspsychologie, so bei W. Hacker (1986), findet sich die gleiche Bestimmung. Primäres Ziel der Untersuchungstätigkeit ist das Erfassen oder Aufnehmen von Selbstberichten, Selbstaussagen über das Auftreten von Belastungen, die in Psychologie und Psychiatrie als depressive Symptome bezeichnet und auch inhaltlich bestimmt werden. Das so noch sehr allgemein definierte und dennoch gesuchte Objekt dieser Studie tritt in ungezählten Einzelereignissen auf, über das nachfolgend, wenn auch in unterschiedlicher Weise, von allen Kindern einer Stichprobe berichtet wird, vom Erleben solcher Belastungen, die psychologisch und psychiatrisch als depressive Symptome verstanden und eingeordnet werden. Das Vorhaben, über solch Erleben bei den hier untersuchten 2034 Kindern aus Griechenland – zugleich Subjekte und Objekte in dieser Untersuchung – verlässliche Informationen einzuholen, bedarf eines Erfassungsinstrumentes. In dieser Untersuchung sind das zwei depressionsspezifischen Fragebögen, der DTK und das CDI, die einander als wechselseitige Kontrollinstrumente eingesetzt sind. Beide Verfahren bestehen aus Aussagen, die, in Übereinstimmung mit dem DSM und ICD, nach depressiven Phänomenen und nach solchen auf unterschiedlichen Ebenen fragen. Ausdrücklich ist von den Autoren beider Verfahren betont, dass diese Instrumente keine klinische Diagnose erstellen. Das könnten sie auch nicht. Sie sammeln lediglich Antworten ein und ermöglichen den Forschern, diese Antworten zu analysieren. Alles was analysiert wird, gewinnt den Status von Kennzeichnet für das individuelle und das gruppenspezifische Antwortverhalten der Schulkinder. Diese Kennzeichen können Bedeutungen erlangen oder schon erlangt haben, wenn sie als Indikatoren für spezielle Sachverhalte bekannt werden.

Dem entsprechend darf diese Arbeit auch nicht als ein Vorhaben zum Aufsuchen psychischer Störungen interpretiert werden. Primär werden Selbstwahrnehmungen von Schulkindern eingesammelt und analysiert. Im Übrigen können auch diagnostische Prozeduren, was dann zu ihrem eigentlichen Problem wird, depressive Störungen von Schulkindern nicht aufsuchen, die diagnostizierenden Personen begegnen auch nur Kindern. Als Analyseergebnis könnten aber Indikatoren oder Indizien gefunden werden, die das Bestehen depressiver Störungen bei den befragten Schulkindern als wahrscheinlich erkennen lassen könnten. Ob die damit geäußerten Vermutungen zutreffen, müssten nachfolgende diagnostische Untersuchungen prüfen.

Nach all dem kann die durchgeführte Untersuchung nicht unmittelbar dem Auftreten von klinisch definierten depressiven Störungen nachgehen. Das in dieser empirischen Studie erfassbare Objekt wird über die ausgewählten Messmethoden bestimmt und erfährt durch sie ihre Operationalisierung. Wie in den Naturwissenschaften gilt auch für die Sozialwissenschaften, dass nicht die jeweils interessierenden Objekte selbst zum Untersuchungsgegenstand werden. Nicht das gemeinte Ganze wird erfasst, sondern nur dessen Eigenschaften, soweit sie oder Aspekte von ihnen messbar und beobachtbar sind. Jedes Objekt kann durch ein System von Eigenschaften gekennzeichnet sein (Bortz, 1999, 5. Auflage, S. 17).

Gesucht wird eigentlich nach depressiven Störungen oder Belastungen im Sinne der Bestimmungen im DSM oder ICD. Die Messinstrumente, in diesem Fall DTK und CDI, sind aber nur geeignet, Aussagen von potentiell betroffenen Schulkindern aufzunehmen, die einzelne, wenn auch spezifische Eigenheiten depressiver Störungen betreffen. Immer wieder neu erweist sich dabei, dass mit der Sammlung von depressionsrelevanten Selbstberichten das Antwortverhalten

der befragten Kinder auf die durch die Instrumente gestellten spezifischen Fragen bekannt wird. Das nun wird nach Indizien für das Vorliegen depressiver Belastungen abgesucht.

Die aufgenommenen Selbstaussagen, das Datenmaterial, erlangt erst die gewünschte Aussagekraft, wenn andere Untersuchungen den Kennwerten des Antwortverhaltens spezifische, also aussagekräftige Bedeutung verleihen.

Die Analysearbeit strebt demnach mit ihrem ersten allgemeinen Ziel auf die Beantwortung der Frage zu, wodurch Selbstaussagen zu Indizien oder Indikatoren werden. Das ist auch die Frage nach der Bedeutung erfasster Daten. Bedeutungen entstehen immer in gedachten Beziehungssystemen. Die Bedeutung eines vor einem Haus stehenden Autos wäre auf einer 1 Quadratkilometer großen Insel im Schärengebiet vor Stockholm eine andere als die eines solchen Gefährts in einem sibirischen Dorf. Die Bedeutung wechselte sehr wahrscheinlich aber auch mit jeder anderen denkbaren Verschiebung des Objektes im gedachten Bezugssystem. Für ein Kind hätte es sehr wahrscheinlich andere Bedeutungen als für die Eltern des Kindes und für einen Kranken andere als für einen Gesunden. Die Beispiele stehen zur Veranschaulichung der Aussage, dass Bedeutungen nicht absolut sondern relativ bestimmt werden und dies in Relationen zu dem für die Bewertung zuständigen Bezugssystemen. Den Antworten der an der Untersuchung 2034GR beteiligten Schulkinder werden ebenso und in Abhängigkeit von den jeweils heran gezogenen Bezugssystemen unterschiedliche Bedeutungen zugesprochen bekommen. Sie können als Indizien dafür gesehen werden, dass diese Kinder lesen können, mit Erfolg in der Bewältigung solcher Aufgaben trainiert sind, sich als sozial ansprechbar erweisen oder mit ihrem Antwortverhalten anderes belegen, was als bedeutungsvoll angesehen werden könnte. In der Untersuchung 2034GR gewinnen die Antworten ihre jeweilige Bedeutung in Abhängigkeit von den Untersuchungsfragen, die zu beantworten versucht werden. Die wiederum stehen in einem eigenen Bezugssystem, über das eine Untersuchungslogik hergestellt wird, nach der die jeweils durchgeführten Arbeitsschritte ihre spezielle Bedeutung gewinnen.

Für die Untersuchung 2034GR gewinnen die mit den durchgeführten Messverfahren erhobenen Daten über das Antwortverhalten der beteiligten Schulkinder ihre untersuchungsspezifische Bedeutung über die Beantwortung der Frage, was die jeweils bemerkten Aspekte dieses Antwortverhaltens über das Auftreten depressiver Symptome innerhalb der untersuchten Stichprobe aussagen.

Eine im Antwortverhalten speziell beachtete Qualität ist die der Zahl von Ja-Antworten auf die mit den Verfahren n-fach gestellten Fragen nach dem Auftreten depressiver Symptome. Die Antworten selbst sind Hinweise auf die von den Kindern in ihrem Selbsterleben bemerkten Symptome. Eine darauf zu beziehende und selbst umgangssprachlich hochfrequent benutzte Nachfrage ist die, ob die aufgenommenen Aussagen der Kinder realitätskonform sind, ob sie „für wahr“ gehalten werden dürften. Dieser nahe liegende Zweifel entspricht einer in das Bezugssystem der Untersuchung an vielen Stellen eingesetzten Prüffrage. Je nach der auf sie zu gebenden Antwort wechselt die Bedeutung der von teilnehmenden Schulkindern über Ja-Antworten gesammelten Gesamtpunktwerte von hoher Aussagekraft zu zunächst relativer Bedeutungslosigkeit.

Welche Antwort auf diese Frage wirklich zu geben wäre, müsste empirisch verlässlich entschieden werden. Eine dabei zu bedenkende Hypothese wäre, das die Antworten der Kinder bei der Durchführung des DTK im wesentlichen der von ihnen erlebten Wirklichkeit entsprechen und somit dann auch höhere Frequenzen von Nennungen, also höhere Gesamtpunktwerte, in Absetzung gegenüber geringeren Mengen und vor allem gegenüber dem Punktwert 0, beachtenswerte Hinweise auf höhere Belastung der Auskunft gebenden Kinder geben würden.

Auf kritische Weise ist damit die Frage verbunden, wie der Grad der selbstberichteten Belastung differentiell gewürdigt werden kann. Zu klären wäre, ob die Stärke der depressiven Belastung kontinuierlich mit der Frequenz der Selbstberichte steigt und wo etwa Grenzen zwischen leichteren, milderer und starken Belastungen gesetzt werden könnten.

Die Beantwortung dieser und ähnlicher Fragen könnte gelingen, wenn gewusst werden würde,

1. ob die von den Schulkindern über den DTK selbst berichteten Symptome wirklich auftreten, ob die über den DTK erfassten Gesamtpunktwerte als Ausprägungen in der Häufigkeit der Zustimmung zum Auftreten eines Symptoms, mit der Häufigkeit des Auftretens von Symptomen systematisch übereinstimmen,
2. welche Punktwerte im DTK mit welchen Belastungsgraden korrespondieren,
3. welche auftretenden Belastungen mit welcher Berechtigung als häufig auftretend und alltäglich,
4. welche in Absetzung davon als bereits subklinisch und welche sogar als klinisch bedeutsam aufgefasst werden sollten.

Die im letzten Punkt eingeführte Benennung dreier kategorial bestimmter Bereiche an Gesamtpunktwerten bedürfte der genaueren inhaltlichen Bestimmung. Die wird hier zunächst ausgespart.

Die aufgeführten Punkte 1 bis 4 sind alle Teil- und Nebenaspekte der jeder Anwendung des DTK, wie aber auch der des CDI, allgemein unterliegenden Validitätsproblematik. Die Selbstausskünfte der Schulkinder sind zweifelsfrei das, was sie sind, aber das gibt keinen Aufschluss über den Realitätscharakter des behaupteten Auftretens der Symptome. Zu prüfen wäre auch, ob eine systematische Korrespondenz zwischen der Skala der Häufigkeit von Itemwahlen im DTK und einer bislang nur erdachten Skala der Frequenz real auftretender Symptome besteht. Für die Beurteilung der klinischen Bedeutung der Gesamtpunktwerte müsste, ungeachtet aller auch dann noch ungelöst verbleibender Probleme, bekannt sein, welche persönlichkeitswirksame Belastung mit dem behaupteten Auftreten von Symptomen verbunden ist.

Versuche zur Beantwortung dieser und anderer mit dem Bekanntwerden von Gesamtpunktwerten aus dem DTK verbundenen Fragen sind ausnahmslos auf dafür erforderliche empirische Daten angewiesen und speziell auf Ergebnisse aus griechischen Untersuchungen.

Die liegen nicht vor. Ungeachtet dessen verfolgt die hier darzustellende Untersuchung das Ziel, möglichst realitätsnahe Aufschlüsse über die mutmaßlich bestehende Prävalenz depressiver Störungen bei griechischen Schulkindern zu gewinnen. Erhofft wird das Auffinden von Indizien, die mit Wahrscheinlichkeiten für das Auftreten depressiver Störungen verbunden werden können. Die Voraussetzungen für das zweifelsfreie Auffinden von depressiven Untersuchungen sind Da eine zweifelsfreie Ermittlung der Prävalenzen für das Auftreten depressiver Störungen unter griechischen Schulkindern sind für diese Untersuchung, wie für viele andere auch, nicht gegeben. Realisierbar ist für diese Untersuchung aber der Versuch einer unter gegebenen Bedingungen größtmöglichen Annäherung an die mitgeteilte Zielsetzung, Aufschlüsse über das wahrscheinliche Auftreten depressiver Störungen zu finden.

Der aktuell bestehende Informationsstand über das Auftreten depressiver Belastungen unter griechischen Schulkindern würde auch so bemerkenswert erweitert werden können.

Das für die bislang unbeantworteten Fragen zu den Testgütekriterien des DTK erforderliche empirische Wissen wäre über eine entsprechend große klinische Stichprobe von griechischen Schulkindern zu gewinnen. Bei der Planung der Untersuchung, über die hier berichtet wird, stand eine solche hinreichend umfängliche und verlässlich diagnostizierte Stichprobe nicht zur Verfügung. Das damit weiter bestehende Wissensdefizit ist erkannt, benannt und ebenso die Notwendigkeit, es möglichst schnell zu reduzieren. Angemerkt wird dazu, dass die Verwendung

von Diagnostischen Interviews an Stelle von DTK und CDI die bestehende Grundproblematik nicht hätte aufheben können. Diese Einschätzung ist an anderer Textstelle (Kap. 6) angesprochen und begründet worden. Die meisten dieser Verfahren heben die gravierenden Validitätsprobleme nicht auf.

Da der Forschungsfrage nach dem Auffinden von Indizien für das Bestehen depressiver Störungen bei griechischen Schulkindern mit dem zusätzlichen Ziel einer möglichst realitätsnahen Annäherung an die Vorgabe weiterhin gefolgt wird, ist darzustellen, wie eine solche größtmögliche Annäherung an die angestrebte empirisch fundierte Beantwortung der Forschungsfrage vollzogen werden soll. Der Erörterung der damit akzentuierten wissenschaftstheoretischen, methodologischen oder methodischen Probleme wird in mancher vergleichbaren Untersuchung nur wenig Platz eingeräumt. Entsprechende deutschsprachige Lehrbücher geben Beispiele dafür (Essau, 2002; Groen und Petermann, 2002; Reicher, 1998). Vorzugsweise werden die Prozentsätze der mit Fragebögen ermittelten Anteile an Kinder mit einer Vielzahl selbst zugewiesener depressiver Symptome als das mitgeteilt, was sie sind, ohne auf die Validitätsproblematik einzugehen. Insofern müssten die Konstruktion und die Anwendung von diagnostischen Interviews als methodischer Fortschritt verstanden werden. Allerdings treten mit den diagnostischen Interviews auch für sie typische Validitätsprobleme auf.

Der erste Arbeitsschritt zur möglichst weitgehenden Annäherung an das mit der Forschungsfrage nach Indizien für das Bestehen depressiver Störungen bei griechischen Schulkindern gesetzte Ziel, ist die explizite Rückbesinnung auf den Realitätscharakter der DTK Daten. Die Analyse kann zunächst nicht mehr, als das erfassen, was sich bei der Durchführung der Untersuchung mit der Bearbeitung des DTK durch die Schulkinder und mit dem Einsammeln der Bögen vollzogen hat. Die Tätigkeit der Schulkinder beim Ausfüllen der DTK Bögen konstituiert den Datencharakter:

1. Die mit dem DTK erhobenen Daten sind von Schulkindern gegebene Auskünfte über das von ihnen an sich selbst erlebte Auftreten von Symptomen, die nach DSM und partiell auch nach ICD als Symptome depressiver Störungen betrachtet werden.
2. Die Arbeit auf einer erste Analyseebene kann dann die Analyse des Antwortverhaltens sein.
3. Die auf dieser Analyseebene erreichbare Gewissheit ist, dass die beteiligten Schulkinder diese Selbstzuweisungen vorgenommen haben. Die entsprechenden Daten sind uneingeschränkt inhaltsvalide.
4. Alle weiteren Analyseschritte beruhen bereits auf empirisch mehr oder weniger gesicherten Annahmen.
5. Die erste Annahme ist, dass die Gesamtpunktwerte aus dem DTK mit ihrem Ansteigen von 0 bis zu ihrem Maximalwert systematisch mit dem Schweregrad der depressiven Störung des jeweiligen Schulkindes auf Rangskalenniveau gesichert, korrespondieren. Diese Annahme unterliegt implizit oder explizit jeder auf Gesamtpunktwerte zurückgreifenden Aussage über die im Individualfall vorliegende Ausprägung einer depressiven Störung.
6. Das angenommene Bestehen der angesprochenen Beziehung zwischen der Variablen der Gesamtpunktwerte und einer latenten Variablen der Ausprägung oder des Schweregrades einer auftretenden Belastung durch depressive Symptome oder depressiven Störung erlaubt dann auch eine vereinfachende kategoriale Unterscheidung der angenommenen individuellen Belastung von Schulkinder durch depressive Symptome als eher alltäglich, subklinisch oder klinisch bedeutsam ausgeprägt. Diese Unterscheidung hat zunächst nur deskriptiven Charakter.

7. Als wahrscheinlich klinisch belastet sollen Schulkinder eingestuft werden, deren Gesamtpunktwerte im DTK für eine klinische Ausprägung ihrer Störung sprechen. Der Punkt auf der Skala der Gesamtpunktwerte, der wahrscheinlich klinisch belastete Kinder von subklinisch belasteten trennt ist der Cut-Off Punkt der DTK Skala.
8. Der Autor des DTK (Rossmann, 2005) hat keinen solchen Trennwert angegeben, verweist aber auf Untersuchungen dazu (Pötz, 20002; Zimmermann, in Rossmann, 2005).
9. Die nach solchen Hinweisen verwendeten Trennwerte helfen, zwischen Teilgruppen von Schulkindern aus der Stichprobe zu unterscheiden, für die angenommen wird, dass sie sich im Ausprägungsgrad ihrer Belastungen durch depressive Symptome, intergruppal variieren.

Inwieweit dann nachfolgend den von Punkt 1 bis 9 mitgeteilten Eigenheiten des aus der Durchführung des DTK gewonnenen Datensatzes widerspruchsfrei gefolgt werden kann, wird sich mit den zu vollziehenden Arbeitsschritten erweisen müssen. Modifizierungen der eingenommenen Perspektive auf den Datensatz können notwendig werden.

Die individuell mit dem Ausfüllen der DTK Formulare erreichte Punktzahl ist innerhalb der statistisch zu beachtenden Regeln für jedes der Schulkinder in der untersuchten Stichprobe zweifelsfrei zu bestimmen.

Damit sind aber nur die unter Punkt 1 bis 3 eben zuvor aufgeführten Sachverhalte abgesichert.

Für eine empirisch eindeutige Ableitung von Beziehungsaussagen zwischen DTK Gesamtpunktwerten einerseits und bei griechischen Schulkindern akut bestehenden depressiven Störungen fehlen, wie mehrfach angeführt, die notwendigen empirischen Daten. Die griechische Version des DTK wird, nach ihrer Erprobung, erstmals in einer Untersuchung eingesetzt. Die Erarbeitung von insbesondere klinischen Daten zur differentiellen Validität, wird hier nochmals als für eine jede DTK Anwendung in Griechenland unerlässliche Absicherung erklärt. Bis eine solche verbesserte empirische Grundlegung geschaffen ist, wird hier in dieser Untersuchung 2034GR auf die empirischen Grundlegungen zurückgegriffen, die für die deutsche Version des DTK gelten. Dieses Vorgehen ersetzt nicht die zu fordernde griechische Grundlegung, aber sie stellt eine Annäherung an die gewünschte Absicherung her. Diese Annäherung ist insoweit abgesichert, als der nachstehende Satz empirisch zutreffend sein sollte:

1. Ein in Deutschland bewährtes Testinstrument, der DTK, das sich bei der Erfassung depressiver Störungen in Deutschland bewährt, wird sich in einer griechischen Version beim Erreichen vergleichbarer Messqualitäten in ähnlichem Masse bewähren wie die Ursprungsversion.
2. Diese Bewährung in einer mit zumindest gleichen Messqualitäten arbeitenden fremdsprachigen Version wird umso wahrscheinlicher, als für die fremdsprachige Version Übereinstimmungen mit anderen Messinstrumenten empirisch verlässlich nachgewiesen werden könnten, wie sie für die deutschsprachige Ursprungsversion bekannt geworden sind.

Die vorangegangene zweite Aussage meint, dass den mit dem griechischen DTK erarbeiteten Ergebnissen um so eher zu vertrauen wäre, als sich nach der parallelen Durchführung eines griechischen CDI zwischen beiden griechischen Versionen Korrelationen ergeben sollten, die bereits für die deutschsprachigen Versionen beider Verfahren zuvor bekannt gewesen sind.

Die für den DTK in dieser griechischen Untersuchung auftretenden Validitätsprobleme können durch die Anlehnung an die für die deutschsprachige Anwendung in Anspruch genommenen empirischen Validitätsbelege nicht aufgehoben, wohl aber reduziert werden.

Die Messqualitäten, die den deutschsprachigen DTK auszeichnen, werden von seiner griechischen Fassung in gleicher Weise erreicht (siehe Kapitel 6). Die Übereinstimmung des Koeffizienten Alpha und die der Itemladungen mit der originalen Version ist sehr hoch. Diese Beobachtung führt einerseits zur Stärkung der Annahme von der relativen Übereinstimmung der Verfahren. Andererseits erlauben sie auch, neuen Fragen nachzugehen. Eine solche Frage wäre beispielsweise, ob man nach dieser Übereinstimmung annehmen dürfte, dass die griechische und die deutsche Version trotz der unterschiedlichen Sprachen jeweils Fragen identischer Bedeutung verwenden, die von Kindern aus beiden Ländern in fast übereinstimmender Weise verstanden werden. Die Ergebnisse der Faktorenanalysen widersprechen dem nicht. Die beobachtete faktorielle Übereinstimmung legt die Suche nach Erklärungen nahe.

Wenn so aber die griechische Version und die deutsche Version des DTK in ihren Messqualitäten einander entsprechen, dann wäre insoweit auch die Erwartung zu begründen, dass die griechische Version ebenso der für den deutschsprachigen DTK erwiesenen Validität nahe kommen wird. Das sollte als wahrscheinlich angesehen werden dürfen, weil sowohl der DTK als auch die mit ihm vergleichbaren Tests eine innerhalb einer gewissen Varianz international übereinstimmende Validität erreichen. So etwa erweist sich der CDI mit seinen Anwendungen in allen mitteleuropäischen, westeuropäischen und nordamerikanischen Staaten als ein für klinisch relevante Hinweise auf das Auftreten depressiver Störungen bei Schulkindern valides Instrument. (s. auch Kovacs, 2004). Sehr wahrscheinlich gilt das auch für andere Staaten. Erwartet werden dürfte, dass sich der CDI auch in Griechenland als ein solches Messinstrument bewähren würde. Für den DTK ist in diesem Sinne auf seine Bewährung in Österreich, Deutschland und der Schweiz hinzuweisen. Im Übrigen korrelieren aber Ergebnisse aus Messungen mit dem DTK in Anwendungen innerhalb identischer Stichproben mit Messergebnissen aus dem CDI. Falls sich erweisen sollte, dass Messungen aus der Anwendung der griechischen Version des DTK in einer identischer Stichprobe mit Messergebnissen aus einer griechischen Version des CDI übereinstimmen, wäre die Hypothese gerechtfertigt, die griechischen Versionen beider Verfahren könnten die depressionsspezifische Validität der international eingesetzten Varianten von DTK und CDI erreichen. Ob sich eine solche Korrelation zwischen den griechischen Versionen von DTK und CDI ergeben wird, ist bereits nach der Analyse der Daten aus der Anwendung beider griechischer Versionen in dieser als 2034GR (2034 für die Stichprobengröße und GR für Griechenland) bezeichneten Untersuchung zu überprüfen.

Den nach all dem dennoch zu verlangenden direkten Nachweis für die Validität des griechischen DTK müssen weitere empirische Untersuchungen an klinischen Stichproben von Schulkindern mit diagnostizierten depressiven Störungen erbringen. Die Untersuchung 2034GR liefert weder solche Daten noch können sie aus der psychologischen griechischen Literatur entnommen werden.

Die Untersuchung, über die hier zu berichten ist und für die in diesen Ausführungen die ihre sprachliche Repräsentation erleichternde Kennung 2034GR Verwendung findet, geht von der Annahme aus, dass der griechische DTK eine depressionsspezifische Validität hat, die der für die deutsche Version entspricht. Auszuschließen wäre die empirische Widerlegung der mitgeteilten Grundannahmen aber keineswegs. Sie wird hier eingeführt, um temporär eine Arbeitsbasis nutzen zu können. Sie ist nachfolgend auf ihre Haltbarkeit zu prüfen, spätestens mit dem Vorliegen der Korrelationen zwischen den griechischen Versionen von DTK und CDI in deren Anwendung in 2034GR.

Die zielgerechte Validität der deutschsprachigen Version des DTK ist belegt, wenn auch noch nicht in dem zu wünschenden Ausmaß (Rossmann, 2005). CDI und DTK sind zwei

depressionsspezifische Verfahren, die beachtenswerte Quoten bei der Identifizierung depressiver Kinder in entsprechenden Stichproben erreichen. Sowohl für den deutschsprachigen DTK als auch für den CDI werden Identifizierungsleistungen bei der Aufgabe erreicht, zwischen nachweislich von depressiven Störungen mit klinischer Ausprägung betroffenen Kindern und anderen ohne Wissen um die jeweilige Diagnose verlässlich zu unterscheiden. Für beide Verfahren werden sensitive Höchstwerte von fast 82% ermittelt. (Rossmann, 2005; Kovacs, 2004). Die insgesamt mitgeteilten Werte variieren für beide Verfahren mit den Untersuchungen, über die sie ermittelt wurden. Die Datenbasis ist beim CDI wesentlich breiter als beim DTK. Den genannten Höchstwerten stehen, so beim CDI, auch Trefferquoten von etwas mehr als 60% gegenüber. Bortz und Döring (2002) nennen für diagnostisch einsetzbare Messverfahren, wie andere Autoren auch, eine Erkennensleistung von 95% als untere Grenze. Beide Verfahren, DTK und CDI, bleiben unter diesem Grenzwert.

Da weiterhin beide Verfahren in ihrer bisherigen Anwendung beachtenswert hoch mit anderen depressionsspezifischen Tests korrelieren, scheint deren eigene depressionsspezifische Aussagekraft insgesamt belegt. Sicher ist aber auch, dass mit beiden Verfahren die Aufgabe der korrekten Trennung von durch klinische Ausprägungen einer depressiven Störung belasteten Schulkindern und anderen, weniger betroffenen Kindern, korrekt zu entscheiden, nicht nachgekommen werden kann. Mit jedem der beiden Verfahren, also auch mit dem DTK, werden sowohl Zuweisungsfehler der einen als auch der anderen Art auftreten. Für den DTK wird mit Rossmann (2005) das Erreichen einer Gesamtquote an korrekten Zuweisungen von etwa 75% zunächst für wahrscheinlich gehalten. Immerhin stünden jeweils drei richtigen Zuordnungen eine falsche Zuweisung gegenüber. Dabei würden die mit depressiven Symptomen hoch belasteten Schulkinder bei einer sensitiven Validität von 82% treffsicherer erkannt als die anderen. Die Trefferquote läge um 4:1. Dieser Sachverhalt ist, wegen der Verwendung eines Trennwertes im oberen Wertebereich, ein Indiz für die vermutete zunehmend höhere Sensitivität der zunehmend höheren Gesamtpunktwerte des DTK.

Alle solche Überlegungen, auf die dann nachfolgend noch ausführlicher einzugehen sein wird, ersetzen nicht die bislang fehlenden Validitätsprüfungen der griechischen Version des DTK in klinischen Stichproben, aber sie werden notwendig, um die Möglichkeiten für die angestrebten Annäherung an korrekte Zuweisungen, das wären die mit den geringsten Fehlerquoten, abschätzen zu können. Eine Trefferquote von 80% für die Benennung der nach DTK Gesamtpunktwerten als mutmaßlich von depressiven Symptomen in einem klinischen Ausprägungsgrad belasteten Schulkinder stellte sich nach dem derzeitigen Kenntnisstand als Grenze der maximale Annäherung dar. Indem diese Grenze genannt wird, entsteht keine Gewähr dafür, sie auch wirklich zu erreichen, weil ungewiss ist und für 2034GR zunächst auch unbekannt bleiben wird, welche stichproben- und verfahrensspezifische Höhe der benutzten Kennwerte für differentielle Validität mit dieser Untersuchung eingestellt wurde.

Die Anwendung der griechischen Version des DTK als Hilfsmittel zum Erreichen der mitgeteilten Zielsetzungen nötigt, wie dargestellt, zur Erörterung der Möglichkeiten, die sich für die Lösung der durch das Fehlen hinreichender Validitätsaussagen entstehenden Probleme anbieten. Das als allgemein bestehend bezeichnete Defizit erweist sich bei näherer Betrachtung als in unterschiedlichen Perspektiven auch unterschiedliche ausgeprägt. Sowohl CDI als auch DTK sind im Hinblick auf eine griechische Bewährung zunächst, soweit die Validität gemeint ist, ungeprüft. In einer Perspektive auf die internationale Bewährung der beiden Verfahren fehlen Antworten auf die Validitätsfrage nur für den DTK, weil die anzustrebende Dichte der Überprüfungen zur Validität noch nicht erreicht werden konnte. Für den CDI kann demgegenüber der empirisch nachgewiesene Grad der erreichbaren Validität je nach verwendeten Standard der Anforderungen unterschiedlich bewertet werden. Gegenüber den für ein diagnostisches Verfahren zu erhebenden

Forderungen wäre der CDI als defizitär einzustufen, als Verfahren zur schnellen und ökonomischen Trennung einer größeren Stichprobe in zwei Teilstichproben von wahrscheinlich nicht von einer depressiven Störung betroffenen und einer Teilgruppe von doch damit belasteten Angehörigen der untersuchten Stichprobe ist das Verfahren hinreichend valide. In beiden Teilgruppen wären abschätzbare Prozentsätze der gemeinten Personen erfasst worden. In jeder Teilgruppe befänden sich aber auch fehlerhaft eingeordnete Teilnehmerinnen und Teilnehmer. Die Fehlerquoten an falsch platzierten Personen könnten beispielsweise um 25% aller Angehörigen der gesamten Stichprobe liegen. Daraus ergibt sich für den Fall, dass diese Quoten nicht hinnehmbar wären, die Nötigung zu mindestens einem nachfolgendem Untersuchungsgang mit dem einem nächsten Untersuchungsziel dienenden Verfahren.

Beiden Verfahren unterliegt, wie vielen anderen auch, eine dimensionale Anwendungstheorie. Anwendungstheorien, explizit verbalisiert oder implizit unterlegt, erklären, warum die betrachtete Anwendung als Methode zum Erreichen der jeweils angegebenen Zielsetzungen nützlich sein soll, sollte oder auch ist. Letzteres wäre empirisch zu belegen. Sowohl DTK als auch CDI folgen der Annahme, die mit ihnen individuell zu erreichenden Punktwerte stünden in einem darstellbaren Zusammenhang mit der Ausprägung der Betroffenheit von depressiven Symptomen. Diese Betroffenheit, welcher Begriff auch dafür verwendet werden sollte, sei dimensional zu erfassen, würde bei einem Minimum beginnen und einem Maximum zustreben. Diese Dimension der Ausprägung des Auftretens einer von 0 bis hin zu maximalen Werten variierenden Betroffenheit durch depressive Symptome ist ein Konstrukt und kann selbst nicht als existent nachgewiesen werden, präsentiert sich aber durch die ihm zugeordneten Korrelate und wird durch seine Nützlichkeit legitimiert.

Der vermutete Zusammenhang zwischen höheren, weil hinter dem Cut-Off liegenden Ausprägungen auf der Skala der Gesamtpunktwerte aus der Anwendung des DTK einerseits (Rossmann, 2005) und in den Personen manifestierten Ausprägungen im klinischen Bereich einer dazu gedachten Skala der Ausprägungsgrade einer individuellen Betroffenheit von depressiven Symptomen andererseits, wäre als Beispiel für eine solche Beziehung zu betrachten. Das für den DTK von Rossmann berichtete Erreichen einer sensitiven Validität von 82% bei der verlangten 100% Identifizierung der inmitten einer gemischten Stichprobe befindlichen und als Personen zunächst unbekannt aber depressiv im klinischen Ausmaß belasteten Teilnehmer, gibt die zunächst nur erhoffte und nun bestätigte Legitimation für die Anwendung des DTK. Wer dieses Verfahren einsetzte, könnte bei Anwendung des definierten Cut-Off Wertes darauf vertrauen, mit einer um 18% variierenden Fehlerquote und demzufolge mit einer um 82% liegenden Quote der Gewissheit zutreffender Wahlen, die Kinder erfasst zu haben, die als depressiv belastet gelten dürften. Unbekannt ist zum einen, welches der Mittelwert einer solchen Variation der von Versuch zu Versuch wechselnden, aber erreichbaren Trefferquote ist. Der von Rossmann mitgeteilte Wert könnte extrem niedrig oder extrem hoch sein. Um das entscheiden zu können, sollte eine größere Zahl erreichter Trefferquoten bekannt geworden sein. Zum anderen fehlen Daten über die sensitive Validität der griechischen Versionen des DTK.

In Ermanglung verlässlicherer Daten, wird die von Rossmann (s.o.) berichtete Quote der sensitiven Validität auch als die zur Zeit bestmögliche Schätzung auch einer griechischen sensitiven Validität betrachtet. Sollten sich unter Anwendung des DTK für eine Stichprobe von 1000 Schulkindern unter Anwendung des mitgeteilten Cut-Off Wertes dann 100 Kinder als identifiziert erweisen, wären in dieser Zahl vermutlich etwa $18+x$ oder $18-x$ Kinder zu viel oder zu wenig enthalten. Das verdeutlicht die Notwendigkeit des Aufschlusses der Größe von x und der Feststellung des Mittelpunktes der Verteilung auftretender Abweichungen.

Ein von der Anwendungstheorie in diesem Beispiel genutztes methodisches Konzept ist das der Verteilung der Daten entlang einer Dimension. Von einem definierten Merkmal wird

angenommen, dass es zwischen den Endpunkten in beschreibbarer Weise variiere. So könnte beispielsweise Angst mit niedrigeren, mittleren und höheren Ausprägungsgraden auftreten. Die Skala der Gesamtpunktwerte aus dem DTK variieren ebenfalls entlang einer Dimension und dies aus sehr nahe liegenden Gründen. Die Zahl der individuell erreichten Gesamtpunktwerte kann zwischen dem Minimum 0 und einem Maximum variieren, das der Zahl der DTK Fragen entspricht. Eine Dimension ist damit beschrieben, weil zwischen einem Pol mit minimaler Zahl von Ja-Antworten und einem Maximum an Ja-Antworten die definierte Qualität sich quantitativ verändert. Ein höherer Wert entspricht einem höheren Quantum der definierten Qualität. Eine Skala ist beschrieben, weil die individuellen Werte jeweils eine Position in einer durch identische Abstände voneinander getrennten und somit kontinuierlich auf- oder absteigenden Mengenreihe einnehmen. Diese Dimension ist eine Mengenreihe, die als solche keinerlei Beziehungen zu depressiven Symptomen hätte, aber die Fragen sind Erkundigungen nach dem individuellen Auftreten depressiver Symptome. Damit haben die individuell erreichten Gesamtpunktwerte auch noch keine zwangsläufig auftretende Beziehung zur Ausprägung einer depressiven Störung. Wohl aber ist ein entlang der eingeführten Dimension variierendes Antwortverhalten erfasst worden. Die Anwendungstheorie hat dann zu erklären, wie das entlang dieser Dimension variierende und nach Ja-Antworten quantifizierte Antwortverhalten zu einer depressionsspezifischen Bedeutung kommt.

Die Anwendungstheorie unterstellt, alle befragbaren Kinder oder Jugendliche hätten Erfahrungen mit depressiven Symptomen. Sollte das nicht so sein, könnten einige Kinder nicht inhaltsbezogen antworten. Angenommen werden muss auch, dass Kinder oder Jugendliche mit umfangreicheren Erfahrungen auch mehr Ja-Antworten abgeben würden. Die Qualität der Anwendungstheorie, die nicht identisch mit der Testtheorie des Verfahrens sein muss, könnte an den Darstellungen geprüft werden, mit denen zu erklären versucht wird, warum die befragten Kinder und Jugendlichen sich alle zu den Fragen gehörenden Phänomene gemerkt haben sollten, die nun als Symptome abgefragt werden und wie zustande kommen soll, dass alle in übereinstimmender Weise die psychischen Qualitäten oder die jeweils erlebte Stärke der Symptome adäquat berücksichtigen würden.

Eine Annahme, die von den empirisch kaum erfassbaren oder schwer einschätzbaren Details absehen dürfte, wäre die nachfolgende Variante einer Anwendungstheorie. Sie bildet ein Konstrukt, das inhaltlich als Leidensdruck beschreibbar wäre, der durch das häufige Erleben depressiver Zustände erfahren wird. Diese Erfahrung steht in Dissonanz zu sorgenfreieren und emotional fröhlicher getönten Situationen. Das vor diesem Hintergrund zu bildende Konstrukt ist das einer Bereitschaft, über erlittene Dissonanzen zwischen depressiven Zuständen und erwünschtem fröhlicheren Erleben zu kommunizieren. Von dieser Variablen wird angenommen, dass sie dimensional gerichtet ist. An dem einen Pol der Dimension ist die hohe Bereitschaft zur Negierung des Kommunikationsbedarfes über depressive Zustände fixiert. Die Wahrscheinlichkeit des Auftretens einer spezifischen Kommunikationsbereitschaft ist auf diesem Punkt gleich Null. Das typische Antwortverhalten für diesen Punkt ist „n x nein“. Der Gesamtpunktwert GPW wird folglich auch = 0. Je stärker die erlebte Dissonanz ist, desto stärker ist die Bereitschaft Ja-Antworten zu geben. Mit anwachsender Bereitschaft zur Bejahung aller zum thematisierten Inhalt als zugehörig eingeschätzten Fragen, sinkt die Bereitschaft zur differentiellen Abwägung der unterschiedlichen Frageninhalte. Angenommen wird, dass die beschriebene erlebte Dissonanz ihren Höhepunkt noch vor dem Ende des Poles hat, der dem Erleben hoher Belastungen durch depressive Störung entspricht. Die Abflachung des Anstiegs wird erwartet, weil eine starke depressive Störung auch die Bereitschaft sich aktivieren zu lassen, stark reduziert. Dann könnten unter den Spitzenwerten auf der Skala der GPW im DTK relativ viele Kinder oder Jugendliche sein, die keine depressiven Störungen erlebten, wohl aber

inadäquates Antwortverhalten zeigten und beispielsweise, aus welchen Gründen auch immer, Ja-Antworten für besser als Verneinungen hielten.

Die Regulation des Antwortverhaltens im DTK durch individuelle Werte, die auf einer subjektiv erlebten Dimension der Dissonanz zwischen dem Erleben depressiv getönter Perioden und der Sehnsucht nach fröhlicheren Phasen des Eigenerlebens positioniert sind, ist als intendierte demonstrative Projektion des Erlebens depressiver Perioden zu interpretieren und hat gleichen Appellcharakter wie die Bekundung des Erlebens von Schmerz. Die Skala der GPW im DTK misst nicht die depressive Störung sondern den aus dem Erleben depressiver Episoden erwachsenden Appell der Betroffenen.

Die beschriebene Variante einer Anwendungstheorie für den DTK ist eine Spekulation. Sie dient hier zum Verweis auf mehrere nicht erklärte Aspekte der Anwendung des DTK. Zum einen wird verdeutlicht, dass die Testtheorie der einzelnen Verfahren, hier die des DTK, wesentliche Fragen zur Regulation des Antwortverhaltens unerwähnt und unbeantwortet lässt. Erkennbar ist weiterhin, wie die Erklärung dieses regulativen Prozesses zu Unterschieden in der Bedeutung der GPW im DTK nötigt. Exemplarisch verdeutlicht wird auch, welche Fehlerquellen dem Erreichen eines Höchstmaßes an Validität entgegen stehen. Eine Konsequenz daraus betrifft die Beurteilung der Bemühungen um Erhöhung der Validitätswerte für Screeningverfahren. Die Einschätzung liegt nahe, dass die erreichten Validitätswerte erstaunlich hoch liegen und weitere Steigerungen bei Anwendung vergleichbar dimensionalisierten Skalen nicht zu erwarten wären. Schließlich gibt die exemplarisch dargestellte Variante der Anwendungstheorie auch Gelegenheit, auf den Appell hinzuweisen, der in der durchgeführten Untersuchung von Hunderten griechischen Schulkindern an die untersuchenden Psychologen gerichtet worden ist. Die haben ihn übergangen und benutzen die Testtheorie zur Nivellierung des aufgenommenen Appells bis hin zur Bedeutungslosigkeit. Das geschieht in guter Absicht, sollen die Messdaten doch helfen, psychischen Störungen entgegen wirken zu können. Eine nicht beantwortbare Frage ist die nach dem Zahlenverhältnis zwischen denen, die in epidemiologischen Untersuchungen ihren Appell zu Protokoll gegeben haben, ohne dass ihnen geholfen wurde und jenen, die Hilfe empfangen durften.

Eine Erweiterung der Messproblematik bei der Erfassung depressiver Störungen wird mit Ansätzen vollzogen, die eine Anwendung mehrdimensionaler Messverfahren für notwendig halten. Selbst CDI und DTK basieren auf der Annahme mehrerer mit dem Störungsbild verbundener Verhaltensdimensionen, sie bilden aber dennoch eindimensionale Skalen für einen Gesamtpunktwert. Eysenck beispielsweise beschrieb Varianten der Funktionsweise des Zentralen Nervensystems über die Koordinaten von Extraversion und Neurotizismus. Mehrdimensionale oder multifunktionale Messverfahren zum Erfassen depressiver Zustände sind denkbar und wurden partiell mit den diagnostischen Interviews auch angewendet. Zu vermuten ist, dass alle solche Bemühungen der Intention folgen, validere Diagnosen zu ermöglichen, was wiederum auf das Erleben von Defiziten in der gegebenen Praxis verweisen würde.

Die Erwartung, mit dem griechischen DTK ähnliche Validitätswerte zu erreichen, wie das für die deutsche Version nachgewiesen ist, gibt keine Auskunft darüber, ob, wie hier zuvor schon angesprochen, die mit dem DTK errechneten Gesamtpunktwerte zumindest eine auf Ordinalskalenniveau zu behauptende Übereinstimmung mit dem jeweils individuell bestehenden Grad der depressiven Störung erreichen. Das ist allein schon deshalb zu bezweifeln, als bislang kein verlässliches System zur Bestimmung des Schweregrades einer depressiven Störung entwickelt und empirisch fundiert worden ist. Dieser Schwierigkeit könnte dadurch begegnet werden, dass die Gesamtpunktwerte aus einer auf Rangskalenniveau relativ verlässlichen Einschätzung des Ausprägungsgrades einer individuell bestehenden Störung benutzt wird, um das

jeweils beachtete Kind einer von mehreren und noch zu definierenden Kategorien des Ausprägungsgrades der depressiven Störung zuzuweisen. Vorgeschlagen wird eine, hier schon zuvor erwähnte kategoriale Dreiteilung.

In dieser Praxis wird nicht von einem kontinuierlichen Anwachsen des Ausprägungsgrades der depressiven Störung mit dem Anwachsen des GPW im Messverfahren ausgegangen, sondern nur von ähnlichen Wahrscheinlichkeiten für das Bestehen eines kategorial definierten Ausprägungsgrades unter den Mitgliedern der gebildeten Teilstichprobe.

Die dritte Kategorie, die einer Höchstbelastung, sollte dabei wie folgt definiert werden: „Für dieses Schulkind liegen aus dem DTK herrührende Hinweise auf das Bestehen einer im klinischen Sinne ausgeprägten depressiven Störung nach DSM und ICD vor.“

Der auf der Skala der Gesamtpunktwerte empirisch zu bestimmende Trennpunkt zwischen dieser Kategorie 3 mit relevanten Hinweisen auf das Bestehen einer klinisch ausgeprägten Störung und den beiden anderen Kategorien (1 und 2), sollte, allein schon aus Mangel an Alternativen, der empirisch belegte Messwert sein, an dem das Verfahren seine höchste Sensitivität in der Identifizierung klinisch depressiver Schulkinder erreicht. Dieser Messwert trennt die Kategorie 3 von der in Richtung auf den Wert 0 folgenden Kategorie 2 und 1.

Die Kategorien 2 und 1 repräsentieren dann die Gesamtpunktwerte aus dem Verfahren, die unterhalb des eben definierten empirischen belegten Cut-off Punktes beginnen und dann an dem Gesamtpunktwert 0 enden. Die Trennstelle zwischen den Kategorie 2 und 1 kann, solange nur der Skalenbereich mit der höchsten Sensitivität bekannt ist, nicht wirklich empirisch fundiert bestimmt werden. Für diesen Bereich wird dessen Mitte als Trennpunkt zwischen Kategorie 2 und Kategorie 1 vorgeschlagen. Diese Setzung ist willkürlich und kann sich als solche bewähren oder eben auch als nicht empfehlenswert beweisen.

Unter der Annahme einer auf Rangskalenniveau über die Gesamtpunktwerte im DTK zu vollziehenden verlässlichen Unterscheidung von Schweregraden bestehender Belastungen mit depressiven Symptomen ist dann gesichert, dass die Angehörigen einer Kategorie 2 sehr wahrscheinlich stärker belastet sind als die der Kategorie 1. In dieser Aussage sind die dann zu thematisierenden Vertrauensbereiche in den Grenzbereichen zwischen den Kategorien nicht thematisiert (siehe etwa Klauer, 2001).

Dem allen folgend kann für die Kategorie 2 die nachstehende Bestimmung eingeführt werden: „Für die Schulkinder der Kategorie 2 liegen mit den Gesamtpunktwerten aus dem DTK gegebene Hinweise auf das Bestehen einer unterhalb der klinisch ausgeprägten depressiven Störung bleibenden subklinischen Belastung mit depressiven Symptomen nach DSM und ICD vor.“

Für die auf der Skala der Gesamtpunktwerte im unteren Bereich platzierte Kategorie 1 wird die nächste Bestimmung benutzt: „Für die Schulkinder der Kategorie 1 liegen aus dem DTK herrührende Hinweise auf das Bestehen einer in diesem Ausmaß alltäglich möglichen leichteren Belastung mit depressiven Symptomen nach DSM und ICD vor.“

Für alle drei Kategorien ist der nachstehende Vorbehalt anzuwenden: „Die Zuweisung von Kindern zu den kategorial unterschiedenen Teilgruppen 1, 2 und 3 ist mit einem Zuweisungsfehler behaftet. Der wird für die Teilgruppe 3 auf etwa 20% ihrer Angehörigen geschätzt (nach Rossmann, 2005).“

Mit der Einführung der Kategorie 1 und 2 und 3 nach Gesamtpunktwerten aus dem DTK wird angenommen, dass jeweils drei Teilgruppen von Schulkindern identifiziert werden, die eine relativ sensitive Trennung von Hinweisen auf unterschiedliche Ausprägungen an Belastungen mit depressiven Symptomen nach DSM und ICD erreicht wird.

Die bis hier voran gebrachte Erörterung der methodischen Ausgangsposition bei der Analyse der Daten aus der Anwendung des griechischsprachigen DTK ist hinreichend für die Begründung der nächsten Analyseschritte. Der methodischen Positionierung der nachfolgend durchzuführenden Analysearbeit wird eine zwar nicht zwingend erforderliche, aber dennoch nützliche Feststellung vorangestellt.

Auf das Fehlen griechischer Untersuchungen zur Validität von DTK und CDI ist in diesen Ausführungen mehrfach und eingehend hingewiesen worden. Diese Problematik betrifft in dieser Untersuchung die Anwendung von DTK und CDI als Tests, die Hinweise auf das Bestehen depressiver Störungen erfassen und messen sollen. In einer anderen und zusätzlich zu beachtenden Perspektive ist zu erkennen, dass die beteiligten Schulkinder, wahrscheinlich frei von jeder Theorie über depressive Störungen, mit ihren Antworten auf Phänomene in ihrem Erleben verweisen. Sie berichten über Erfahrungen, die in ihrem Leben und vielleicht auch in ihrem täglichen Leben auftreten und es möglicherweise qualitativ verändern. Auch Kovacs verwies in einem ihrer frühen Artikel über depressive Störungen bei Kindern und Jugendlichen im Zusammenhang mit Erörterung zum Cut-off-Problem darauf, dass sie persönlich, unabhängig von allen wissenschaftlich zu betrachtenden Validitätsproblemen, ein klinisch beachtenswertes Problem sehe, wenn ein Kind beispielsweise über sich aussagt, immer traurig zu sein. Schon die in einem einzigen Item dieser Art gegebene Selbstdarstellung sollte nicht ignoriert werden. Mit einer solchen Aussage verschiebt sich das Validitätsproblem. In diesem Sinne klinisch beachtenswert ist bereits ein einziges depressionsspezifisch beantwortetes Item, wenn es gegen die alltäglich geltende Verhaltensnorm für Kinder einer Altersklasse zu verstoßen scheint. Sowohl aus der Perspektive eines Erwachsenen auf solche Kinder, als auch aus einer pädagogisch-psychologischen Sichtweise erscheinen die depressionsspezifischen Selbsteinschätzungen der Kinder, deren gesammelte Antworten, noch vor allen und unabhängig von wissenschaftlichen Zuordnungen als Bekundungen des Erlebens psychischer Belastungen von Kindern in ihren speziellen Lebenssituationen. Wer diese Äußerungen aufnimmt, täte gut daran, sich auch zu fragen, wie diese Bekundungen pädagogisch oder psychologisch im Hinblick auf die diese Kinder jeweils umschließenden Gesellschaft zu bewerten sind. Wenn Kinder berichten, dass sie sich einsam, wertlos und unglücklich fühlen, dass sie sich um die Zukunft sorgen und das Leben als traurig empfinden, unter Schlafstörungen und Kopfschmerzen leiden, dann stehen die Bekundungen einerseits in Relation zu Begriffen wie „Major Depression“, „depressive Episode“ und „Dysthyme Störung“, sie sind andererseits Indizien für Probleme in der Interaktion mit der Umwelt. Als solche fordern sie psychologisch und pädagogisch begründet dazu auf, die jeweilige Lebenssituation zu analysieren, um den Handlungsbedarf erkennbar werden zu lassen, der möglicherweise in diesem Lebenszusammenhang besteht. Die Forschertätigkeit gibt eine akademische Rechtfertigung für die Distanzierung von den Problemen der Kinder einer Stichprobe. Die Untersucher in aller Welt haben bis jetzt neben Selbsteinschätzungen von Stimmungslagen auch ungezählte Leidensbekundungen von Kindern aufgenommen. Der Prozentsatz jener Kinder, der durch solche Untersuchungen Hilfe erfahren hat, dürfte außerordentlich gering sein.

Eine spezieller Weise der akademischen Distanzierung entsteht mit der fehlerhaften Zuweisung von Kindern in die Teilgruppe der mutmaßlich von depressiven Störungen betroffenen Kindern oder Jugendlichen. Indem sich die falsche Zuordnung herausstellt, verliert diese Person den irrtümlich zugewiesenen Status, ist definitionsgemäß kein depressiv gestörtes Kind und somit

auch kein klinischer Fall. Ob solche Persönlichkeiten möglicherweise nicht allein aus messtheoretischen Gründen und dem Spiel von zufälligen Ereignissen, sondern wegen zwar nicht depressionsspezifischer aber doch bestimmbarer anderer extremer Merkmale falsch zugeordnet worden sein könnte, untersuchte eine Studie unter dem Titel „Symptoms versus Diagnosis: Differences in Psychosocial Functioning“, die von Gotlib, Lewinsohn und Seeley (1995) durchgeführt wurde. Beteiligt waren insgesamt 1709 Jugendliche, die an zwei Zeitpunkten mit klinischen Interviews und Fragebögen untersucht wurden. Dabei fand sich, dass Kinder, die mit Fragebogen fälschlich als depressiv eingestuft worden waren, sogenannte „false positive“ Zuordnungen, beachtenswert hohe psychopathologische Symptome aufwiesen, die sie in ihrem psychosozialen Alltag ernsthaft beeinträchtigten. Die Autoren beenden ihre Zusammenfassung mit dem Satz: „False positive, therefore, is not a benign condition.“ (Gotlib, Lewinsohn und Seeley, 1995, S. 90). Die zitierte Untersuchung wird hier als Beleg für den diagnostischen Eigenwert einer jeden über Gesamtpunktwerte bekannt gewordenen Selbsteinschätzung zur eigenen Befindlichkeit gewertet. Auch ein im DTK zu findender GPW =0 kann bei entsprechender Analyse eine andere Bedeutung gewinnen, als der Wert zu selbst zu belegen schien. Die unter Verwendung des von Rossmann (2005) mitgeteilten Cut-Offs beabsichtigte Bildung von drei Bereichen von GPW definierten Teilgruppen aus der Gesamtstichprobe dient einem bestimmten Zweck, dem der Erfassung von unterschiedlichen Wahrscheinlichkeiten für das Bestehen depressiver Störungen. Darüber hinaus tragen die durch Selbsteinschätzungen über den DTK vorgenommenen Positionierungen der Kinder und Jugendlichen auch andere Informationen, die durch weitere Analysen erfasst werden könnten. Auch mit der hier darzustellenden Untersuchung an griechischen Schulkindern werden wiederum Hunderte von Schulkindern erfasst, denen, unter großem Aufwand zur Anonymisierung der aufgenommenen Daten, in ihrer Leidenssituation nachweislich nicht geholfen wird. Sie bleiben in ihren Schulen, in ihren Wohnorten, auf ihren Inseln oder in ihren Großstadtquartieren zurück. Die Wissenschaft nimmt Kenntnis von ihnen, ohne Anstrengungen aufzuwenden, ihnen wirksam zu helfen.

Um Beihilfe zur Reduzierung dieses gesellschaftlichen Missstandes ist zu werben. Insbesondere die klinische Psychologie hat Anlass zu solchen Anstrengungen. In eben diesem Sinne wird das Ermitteln der mit dieser Untersuchung zu erkundenden Hinweisen auf emotionale Störungen als ein bescheidener Beitrag zur Anbahnung und Begründung gesellschaftlicher Hilfe für betroffene griechische Schul Kinder verstanden.

Mit den bisherigen Erörterungen ist die formale Struktur der zur Analyse aller mit dem Verfahren DTK erhobenen Daten notwendigen Arbeitsschritte hinreichend zu begründen. Zunächst wird 1. das Messinstrument dargestellt, das Gegenstand der Analyse wird. Sein theoretischer Hintergrund, seine Zielsetzungen und sein Aufbau determinieren die Analysemöglichkeiten. „Die inhaltliche Interpretation der Ergebnisse nimmt auf die Theorie Bezug, aus der die Hypothese abgeleitet wurde“ (Bortz, S. 89). Eine implizite aber grundlegende Annahme ist, dass der DTK solche Hinweise zu vermitteln vermag, wie sie mit dieser Untersuchung aufgefunden werden sollen- Des weiteren sind 2. die aus der Durchführung des Verfahrens für die untersuchte Stichprobe bekannt werden Daten zu erfassen. Dem ist 3. die Feststellung der entsprechenden Werte für alle beachteten Teilstichproben nachzuordnen. Dann sind 4. alle ermittelten Werte auf ihre Aussagekraft zur Beantwortung der aufgenommenen Untersuchungshypothesen zu prüfen und die Ergebnisse aus diesen Überprüfungen zu sichern.

7.1 Der DTK als Messverfahren zur Bestimmung von Hinweisen auf unterschiedliche Ausprägungen depressiver Störungen

Die Abkürzung DTK steht für „DepressionsTest für Kinder“ (Rossmann, 1993, 2005). Der Test ist in seiner ursprünglichen Form deutschsprachig. Innerhalb der hier geführten Untersuchung wurde der DTK in einer griechischen Übersetzung eingesetzt. Bei der Suche nach einem depressionsspezifischen Fragebogen für Kinder und Jugendliche fiel die Wahl auf den DTK, weil er ein besonderes Merkmal aufweist. Anders als die meisten der gebräuchlichen Messverfahren zur Erfassung depressiver Symptome bei Kindern, so also auch anders als beispielsweise der CDI, ist er auf das, mit dem Erscheinen von DSM-III, um das Konstrukt der Agitiertheit erweiterte Störungsbild der Depression bei Kindern und Jugendlichen bezogen. Agitiertheit ist sprachlich eine Eindeutschung, die in ihrem Ursprung auf das transitive Verb „agitate“ zurückgeht. In den um das Jahr 2000 gedruckten Ausgaben des Duden wird das Wort nicht aufgeführt. Der Duden bestimmt seine Auswahl von Wörtern an Auszählungen von Sprachproben aus dem alltäglichen Sprachgebrauch. Agitiertheit ist demnach in dem zitierten Berichtszeitraum kein umgangssprachlicher Begriff gewesen. Aus seiner Bedeutungsnahe führen die zitierten Duden Ausgaben nur Wörter auf, die auf Agitation zurück gehen. Nach unterschiedlichen Ausgaben des Oxford Dictionary bezeichnet „agitate“ ein Verhalten, das unter anderem durch folgende Umschreibungen zu kennzeichnen ist: „...shake briskly, ...cause anxiety, ...stir up interest or concern...“ Rossmann erweiterte mit der Konstruktion des DTK, konform zum DSM-III, die Symptomlisten, so dass neben Niedergeschlagenheit, Antriebsminderungen und somatischen Beeinträchtigungen nun auch Symptome agitierten Verhalten erfasst werden konnten. Der DTK ist damit ein Verfahren, das sowohl die schon seit langem bekannten Symptome einer depressiven Störung als auch, ganz im Gegensatz zu anderen depressionsspezifischen Fragebögen, die nun nach dem erweiterten Verständnis des Störungsbildes bei Kindern und Jugendlichen als ebenfalls wesentlich betrachteten ausagierenden Komponenten depressiver Störungen erfasst. Wegen dieser hinzu gekommenen Komponente unterscheidet sich das Störungsbild depressiver Störungen bei Kindern und Jugendlichen wesentlich von dem der Depression bei Erwachsenen. Trotz der genannten konzeptuellen Erweiterung bestehen deutliche Übereinstimmungen in den Messergebnissen von DTK und CDI (Reicher und Rossmann, 1991; Rossmann, 1993).

Der Depressionstest für Kinder (DTK) wurde 1993 in seiner bis heute gültigen Form von Peter Rossmann veröffentlicht. Das Verfahren ist als Resultat der Bemühungen zu sehen, ein damals noch fehlendes genuin deutschsprachiges Verfahren zu entwickeln, das auf die aktuelle Erweiterung des Störungsbildes durch das DSM-III bezogen wird. Mit dem Erscheinen des DTK wurden die im deutschsprachigen Raum gegebenen Messmöglichkeiten innerhalb der Depressionsforschung mit Kindern und Jugendlichen erweitert.

In der ersten Konstruktionsphase des Verfahren wurden insgesamt 300 Items gesammelt, von denen manche selbst entworfen und viele aus bis dahin verfügbaren depressionsspezifischen und zumeist englischsprachigen Fragebögen übernommen worden waren, wenn die in Übereinstimmung mit der klassischen psychologischen Testtheorie standen. Unter Verwendung eigens dafür eingesetzter Kriterien wurde aus dem ursprünglichen Bestand an Items eine aus 107 Items bestehende Rohfassung des DTK entwickelt. Beachtet wurden dabei insbesondere die Forderung nach einfacher syntaktischer Struktur, weiterhin die nach Übereinstimmung des jeweils benutzten Antwortformats und das Kriterium einer dem Leseverständnis jüngerer Schulkinder entsprechenden Formulierung und Länge der Fragen. Das Verfahren sollte ausdrücklich auch für eine Anwendung bei jüngeren Kindern geeignet sein.

Die so entwickelte Rohfassung wurde in einer Untersuchung mit 1450 Kindern aus 14 Schulen (dritte bis sechste Klasse in Volksschulen, Hauptschulen und Gymnasien) einer ersten testtheoretischen Bewährung ausgesetzt. Die angestrebte Struktur des DTK entstand aus der Bearbeitung der Daten der Analysetichprobe in zwei Analyseschritten mittels Hauptkomponentenanalyse (Oblimin-Rotation) und Scree-Test. Das führte zur Entfernung von Items mit niedriger Trennschärfe und zur Eliminierung inhaltlich redundanter Fragen. Die testtheoretische Überprüfung der verbliebenen 55 Fragen erfolgte in Wiederholung der schon zuvor benutzten faktoranalytischen Verfahren. Das wiederholt übereinstimmende Extrahieren dreier Faktoren verhalfen zur Bildung der im DTK verwendeten drei Skalen.

Durch seine Faktorenstruktur bietet der DTK die Möglichkeit, die selbstberichtete momentane depressogene Befindlichkeit von Kindern im Alter von 9-15 Jahren auf drei depressionsrelevanten Dimensionen abzubilden.

Der erste Faktor, im Fragebogen durch 25 Items repräsentiert, umfasst Fragen nach der Dysphorischen Stimmung und nach Selbstwertproblemen. Dieser Teil des DTK ähnelt dem CDI (Kovacs, s.u.).

Der zweite Faktor (16 Fragen) bezieht sich auf die vorher erwähnten Aspekte ausagierenden Verhaltens bei Kindern und Jugendlichen in Schule und häuslicher Umgebung.

Der dritte Faktor (14 Items) wird über Fragen nach Müdigkeit, Schlafstörungen und anderen autonomen Reaktionen, wie Kopfschmerzen und Bauchweh repräsentiert. Fragen dieser Art sind auch im CDI verwendet worden.

Die drei genannten Faktoren korrelieren untereinander und dies positiv.

Der DTK ist für die Anwendung in epidemiologischen Studien innerhalb der Depressionsforschung als Screeningsinstrument konzipiert, hat aber auch in klinischen Stichproben Treffsicherheit bewiesen. Reliabilität und Validität des DTK wurden in verschiedenen Studien überprüft und erwiesen sich als anforderungsgerecht ausgeprägt (siehe auch Kapitel 6.1.1 in dieser Arbeit).

Von spezieller Bedeutung, insbesondere dann auch für die vorliegende Arbeit, sind die Ergebnisse aus zwei mit einschlägigen Validitätsproblemen befassten Studien (Zimmermann, in Rossmann, 1993, und Pötz, 2003). Mit denen werden durch Anwendung des DTK in klinischen und nicht klinischen Stichproben von Kindern empirisch gesicherte Angaben über die diskriminante Validität des Verfahrens bekannt.

Die Daten der Untersuchung von Zimmermann, in einer Poliklinik in Berlin durchgeführt, wurden dem Testautor übermittelt, der dann die aus ihnen hervorgehenden Ergebnisse im DTK-Manual veröffentlichte (Rossmann, 1993, S. 20). Teilnehmer an dieser Studie waren 40 Kinder (Altersdurchschnitt 10 Jahre), für die nach ICD 9 erstellte Diagnosen über das Bestehen klinischer Störungsbilder vorlagen. Nachgewiesen worden waren 1. spezifische emotionale Störung des Kindes- und Jugendalters (ICD 313), 2. Anpassungsstörungen (ICD 309), 3. Störungen des Sozialverhaltens (ICD 312) und 4. das Auftreten des Hyperkinetischen Syndroms (ICD 314). Der intergrupale Vergleich der Mittelwerte für die aus der Anwendung des DTK hervorgegangenen Gesamtpunktwerte zwischen allen diagnostischen Gruppen führte für alle drei DTK-Skalen zu Ergebnissen, die den theoretischen Erwartungen entsprachen. Die Kinder mit der Diagnose einer depressiven Störung und die mit Anpassungsstörung weisen signifikant höhere Werte als die der beiden anderen klinischen Gruppen in der Skala „Dysphorie-Selbstwert“ auf. Ein solches Ergebnis war theoretisch zu erwarten, da sowohl depressive Störungen als auch Anpassungsstörungen mit ihrem Störungsbild auf Niedergeschlagenheit bezogen sind. Im Gegensatz dazu haben in der zweiten DTK-Skala „Agitiertes Verhalten“ jene Kinder signifikant

höhere Werte erbracht, deren Diagnosen auf eine Störung des Sozialverhaltens und auf das Bestehen eines Hyperkinetischen Syndroms hingewiesen hatten. Auch diese Störungsbilder zeigen inhaltliche Übereinstimmung, die das mitgeteilte Ergebnis aus der Studie von Zimmermann nicht nur erklären helfen, sondern es auch erwarten lassen.

Die Studie von Zimmermann spricht für eine skalenspezifische Validität von Skala 1 und Skala 2 im DTK. Für Skala 3 wurde das in Bezug auf die einbezogenen Störungsbilder nicht nachgewiesen. Das Ausbleiben signifikanter Unterschiede und die Beibehaltung der H_0 – Hypothese lassen zwar Erklärungsversuche zu, aber keine über die Mitteilung des Ergebnisses hinausreichenden Aussagen.

Eine zweite Untersuchung, die von Pötz (2003), bestätigt und erweitert das von Zimmermann erreichte Ergebnis. In dieser Studie analysierte Pötz die aus der Anwendung des DTK in unterschiedlichen Stichproben bekannt werdenden Gesamtpunktwerte auf deren diskriminante Validität. Der durchgeführte intergruppale Vergleich bezog sich auf drei unterschiedliche Stichproben. Eine erste Gruppe wurde von 33 Kindern mit der DSM-IV Diagnose einer „leichten depressiven Episode“ gebildet. Ihnen standen zwei weitere Gruppen gegenüber, die eine aus 33 Kindern mit der Diagnose „Aufmerksamkeitsdefizit/Hyperaktivität“ und die andere mit 33 zufällig aus der Normstichprobe von Rossmann ausgewählten Schulkindern. Forschungsziel war, empirische Aufschlüsse über die diskriminante Validität des DTK zu gewinnen. Gefragt war,

1. welchen Beitrag jede der drei DTK-Skalen zur Unterscheidung der genannten drei Kindergruppen leistet und
2. welche Kombination von Cutoff-Werten auf den drei DTK Skalen eine optimale Trennung zwischen den depressiven und den anderen Schulkindern ermöglichte.

Die drei Gruppen von Kindern waren im Hinblick auf ihr Alter vergleichbar. Erwartungsgemäß hatten die depressive Kinder im Vergleich mit der Gruppen der Schüler und Schülerinnen signifikant höhere Werte in den Skalen 1 (Dysphorie/Selbstwert) und 3 (Müdigkeit/autonome Reaktionen). Darüber hinaus erreichten hatten sie auch in der ersten Skala höhere Durchschnittswerte als die Gruppe der hyperaktiven Kinder. Die Durchführung einer Diskriminanzanalyse führte zum Ergebnis, dass die Gesamtpunktwerte auf der ersten und dritten DTK-Skala die als depressiv diagnostizierten Kinder am besten von den anderen Schülergruppen trennen konnten. Über die zweite DTK-Skala konnten die Angehörigen der Gruppe der Schulkinder und den als hyperaktiv diagnostizierten Kindern unterschieden werden. Alle drei Skalen haben skalenspezifische Validität. Die Ergebnisse sprechen dafür, dass die Skalen das messen, was sie messen sollten. Da sie, jede für sich und in anderer Weise, aus Items bestehen, die als für den jeweiligen Skalinhalt konstruktnahe Symptome zu bezeichnen sind, war eine entsprechende Validität begründet zu erwarten. Die Untersuchung von Pötz weist mittelbar nach, dass die Skalen messen, was sie messen sollen. Mittelbar ist der Nachweis, weil die Messergebnisse über Wenn-dann-Beziehungen validiert werden, in denen die Setzungen selbst auch nicht direkt beobachtbar sind.

Um den prädiktiven Gesamtpunktwert zu identifizieren, über den eine verlässliche Trennung zwischen der depressiven Gruppe und den Schulkindern erreicht werden kann, wurden durch Pötz alle auf den drei Skalen zu platzierenden Kombinationen von Trennwerten, von cut-off Werten, geprüft. Die bestmögliche Trennung würde sich dabei über die dafür relevanten Kennwerte, so über Sensitivität, Spezifität, Overall Accuracy, über prädiktiven Wert und Kappa Index, als effizient erweisen. Mit der Durchführung dieser Analysen konnten zwei Cutoff-Wert Kombinationen als leistungsfähig identifiziert werden. Mit deren Hilfe waren die als depressiv diagnostizierten Kinder von den Angehörigen der Gruppe der Schulkinder mit einem durch sie erreichten optimalen Wirkungsgrad zu trennen. Nach der einen der beiden gefundenen

Kombinationen, die über eine Addition des jeweils individuellen Gesamtpunktwertes aus Skala 1 und 2 zu vollziehen war, wurde ein Trennpunkt von $GPW \text{ Skala1} + \text{Skala 2} \geq 17$ ermittelt. Mit einer Trennung der oberhalb des Cut-Offs liegenden kleineren Teilstichprobe aus Teilnehmern mit den höchsten Gesamtpunktwerten von der zahlenmäßig weit größeren g Teilstichprobe mit Gesamtpunktwerten auf oder unterhalb des Trennwertes $X=17$ (cut-off) wurde gleichzeitig eine optimale Trennung der Kinder mit diagnostizierter depressiver Störung von einer unausgelesenen Gruppe mit Schulkindern erreicht. Die Güte der Trennleistung wird durch eine Sensitivität von 81,82 bei einer Spezifität von 66,77 und einem negativ prädiktiven Wert von 0,79 repräsentiert. Eine identische Trennleistung wird erreicht, wenn die Separierung über die Skala 3 des DTK unter Benutzung des Trennwertes Y als Gesamtpunktgleich oder grösser 7 auf der Skala 3 vollzogen wird. Damit ergibt sich die bei der Testkonstruktion möglicherweise nicht angestrebte oder aber doch vorhergesehene Situation, die gewünschte prädiktive Leistung erst über eine nachträglich kombinierte neue Skala aus DTK Skala 1 und Skala 2 erreichen zu können.

In der zweiten aufgefundenen Kombination gelten die gleichen Vorschriften mit leicht veränderten Gesamtpunktwerten als cut-off scores. Die Trennung der Gruppe der Kinder mit diagnostizierter depressiver Störung von der Gruppe der Schulkindern erfolgt über die Addition der Gesamtpunkte aus Skala 1 und der aus Skala 2 mit einem Trennwert X von ≥ 18 . Die Trennleistung entspricht dann einer Sensitivität von 75,76 und einer Spezifität mit 72,73 und dem negativ prädiktiven Wert 0,75. Die Trennleistung wird auch von einem Trennwert Y aus den Gesamtpunktwerten auf Skala 3 von ≥ 7 erreicht.

Die über die Skala 3 erreichbare Trennleistung wird hier abseits der Validitätsprobleme als ein Beispiel für die Funktion von Indizien bei dem Versuch, das wahrscheinliche Bestehen einer depressiven Störung zu erfassen, benutzt. Die Items der Skala 3 repräsentieren nicht die das Störungsbild kennzeichnenden Symptome. Dennoch geben die Antworten auf die dort gestellten Fragen nutzbare Hinweise auf das wahrscheinliche Bestehen einer depressiven Störung. Indizien sind in diesem Sinne einzelne, aber hochbedeutsame Stellvertreter für die Vielzahl der das Gesamtbild repräsentierenden Informationen.

Hervorgehoben wird hier, dass die Untersuchung von Zimmermann und auch die von Pötz auf drei validitätsspezifische Sachverhalte hinweisen:

1. empirisch belegt wird, dass die 3 unterschiedlichen Skalen des DTK eine jeweils skalenspezifische Validität erreichen, indem sie Ausprägungen des Konstruktes messen, dem sie zugeordnet worden sind (siehe Zimmermann bei Rossmann, 1993),
2. empirisch nachgewiesen wird, dass der DTK über Trennwerte (cut-offs) nach Gesamtpunktwerten aus und auf Skalen des DTK eine diskriminante Validität mit einer Sensitivität von etwa 81,5 gegenüber diagnostizierten depressiven Störungen bei Kindern erreichen kann, (Pötz, 2005),
3. der insgesamt erreichte negative prädiktive Wert von 0,75 lässt aber erwarten, dass in einer epidemiologischen Untersuchung immerhin $\frac{1}{4}$ aller untersuchten Personen fehlerhaft zugeordnet werden.

Mit beiden Untersuchungen erweist sich der DTK als Verfahren, dessen Validität empirisch nachweisbar ist. Mit der deutschsprachigen Version des DTK lassen sich nachweislich valide Unterscheidungen zwischen Kindern mit diagnostizierten depressiven Störungen und Kindern ohne diese Diagnose treffen. Bezogen auf die für die hier darzustellende Untersuchung mit griechischen Schulkindern vorgesehene Trennung der Teilgruppen 1 und 2 von der Teilgruppe 3, den Kindern mit den höchsten GPW in der erfassten Verteilung, lässt sich vorhersehen, dass Zuweisungsfehler im Umfang von etwa 25% aller Entscheidungen auftreten werden. Immerhin

bestünde für die Teilgruppe 3 die begründete Erwartung, sie werde zu 82% gesuchte, weil depressiv gestörte Kinder umfassen.

Indem ein Testverfahren aus einer gemeinsam gebildeten Stichprobe von Kindern mit und ohne Diagnose allein auf der Basis seiner Testwerte zwischen den diesen beiden Gruppen in signifikanter Weise zu unterscheiden vermag, erweist sich auch seine Nutzbarkeit. Über das Testergebnis wird bei entsprechend ausgebildeter diskriminanter Validität eine Unterscheidungsleistung hergestellt, die der alltäglich ausgeübten klinischen Praxis durchaus nahe kommt.

Die Anwendung des Verfahrens schließt dann aber auch das Auftreten fehlerhafter Zuweisungen als systembedingt ein.

Die Teilstichprobe klinisch bekannt gewordener Kinder wird fehlerhaft auch solche Kinder einschließen, für die das Bestehen einer depressiven Störung angenommen wurde, obschon diese Kinder nicht wirklich von einer depressiven Störung betroffen sind. Die Teilstichprobe der klinisch bekannt gewordenen Kinder ist nicht frei von irrtümlich vermuteten Belastungen durch depressive Störungen. Die mit den Testergebnissen gebildete zweite Teilstichprobe, die der mutmaßlich störungsfreien Kinder kann auch Kinder umfassen, die von einer depressiven Störung im klinischen Ausmaß betroffen sind, obschon ihre soziale Umgebung das nicht bemerkt hat.

Die Vorgehensweise schließt darüber hinaus eine weitere Fehlermöglichkeit ein. Jede Stichprobe ist ein Teil der ihr zugehörigen Population. Damit variieren bekanntermaßen auch die Mittelwerte der beachteten Merkmale von Stichprobe zu Stichprobe. Wenn eine für die diskriminante Validität des zu prüfenden Verhaltens günstige Stichprobe gezogen wird, darf ein Anwachsen der diskriminanten Leistungen erwartet werden. Dementsprechend ist bei sogenannten ungünstig gezogenen Stichproben mit geringeren Unterscheidungsleistungen zu rechnen. Darauf bezogen schreiben Verhulst und Koot: „.....sensitivity and specificity do not reflect an intrinsic quality of a test. They are not absolute values. Sensitivity and specificity will vary with the samples on which they were based and with the critical values chosen” (Verhulst und Koot, 1992, S. 49)

Der Schluss von einer in Stichproben nachgewiesenen diskriminanten Validität auf ein für die gesamte Population erbringbare Leistung unterliegt eben den Bedingungen, die allgemein für den Schluss von Parametern der Stichprobe auf Parameter der Population gelten (siehe auch Bortz und Döring, 2002).

Dessen ungeachtet beweist das Auftreten diskrimanter Validität des DTK in Stichproben, dass er, ungeachtet der Stichprobenproblematik, eine, sicherlich auch begrenzte, diskriminante Validität hat. Da in der Stichprobe die diskriminante Leistung über die Unterscheidung der vom jeweiligen Kind mit dem Verfahren erreichten hohen, mittleren oder niedrigen Summenwerte erreicht werden soll, ist die nachgewiesene Korrespondenz des DTK über seine Gesamtpunktwerte mit dem Vorliegen einer diagnostizierten depressiven Störung, sind die Voraussetzungen dafür in einem hinreichenden, aber begrenzten Maße gegeben. Die Begrenzung liegt in der Nötigung zur Hinnahme einer geschätzten Fehlerquote von 25% aller Zuweisungen zu den Teilgruppen. Die Nutzung der mit dieser Untersuchung erarbeiteten Ergebnisse muss dies, um die Verbreitung falscher Orientierungsdaten zu vermeiden, ausnahmslos beachten.

7.2 Die in der Untersuchung 2034GR verfügbare Datenbasis der Analysearbeit zum DTK

Nachdem der DTK in der griechischen Stichprobe 2034GR eingesetzt worden ist, die Auswertung erfolgte und die Datenbasis mit der so erarbeiteten Datenmatrix festliegt, können die aus der Anwendung des Verfahrens gewonnenen Informationen mitgeteilt werden.

Das Verfahren DTK verhilft über die Ja-Nein Entscheidungen der teilnehmenden 2034 Kinder bei insgesamt 55 mit den Items des DTK gegebenen Anlässen, zu Informationen über das Antwortverhalten der Mitglieder dieser Stichprobe. Die aufgenommenen Daten bilden mit ihrem Datensatz die Datenbasis des auf den DTK bezogenen Untersuchungsteils.

Die nachfolgend darzustellende Analyse dieser Daten soll die mit ihnen verbundenen Informationen erschließen. Trotz der in der zu Grunde liegenden Datenmatrix gesicherten Konstanz der Daten bieten die durch die in der Analysearbeit herstellbaren Variationen in der Datenorganisation unterschiedliche Informationen. Diese Vielfalt an Informationen ergibt sich aus der Möglichkeit zur unterschiedlichen Organisation in der Darstellung der von den Schulkindern mitgeteilten 2034 x 55 Entscheidungen. Einige der aus der vorhandenen Datenmatrix zum DTK abrufbaren Informationssätze sind nachfolgend aufgelistet. Berücksichtigt werden zunächst nur Strukturierungen von Daten, die das Antwortverhalten aller Teilnehmer an der Durchführung des DTK wiedergeben. Da jeweils Ja-Nein Entscheidungen zu treffen waren, dürfte deren Wiedergabe sowohl der einen als auch der anderen Entscheidung nachgehen. So ließen sich Verteilungen der Ja-Antworten ebenso wie solche von Nein-Antworten zusammenstellen. Die einnehmbaren Perspektiven auf das Antwortverhalten ergänzen einander numerisch, sind aber in Fragestellung, Antwort und Informationsgehalt keineswegs identisch. Im folgenden wird die Perspektive auf die Ja-Entscheidung gewählt, weil mit ihr das selbstbehauptete Auftreten von depressionsspezifischen Symptomen im eigenen Erleben erfasst werden kann und das behauptete Auftreten solcher Symptome auf das vom Forschungsinteresse thematisierte Bestehen depressiver Störungen hinweisen könnte.

In dieser Forschungsperspektive sind die nachstehenden Informationen aus der Datenbasis abrufbar:

1. Zahl der Ja-Entscheidungen für die Items 1 bis 55,
2. Zahl der Ja-Entscheidungen für jeden Teilnehmer (individueller Gesamtpunktwert) und
3. Informationen über die individuellen Antwortmuster der teilnehmenden Schulkinder.

Aus Informationen der ersten Art wäre beispielsweise und neben anderem zu erkennen, welche Items bevorzugt und welche gemieden worden sind. Diese Aufschlüsse korrespondierten mit Daten aus der für diese Stichprobe zum DTK durchgeführten Itemanalyse.

Informationen der zweiten Art lassen erkennbar werden, wie die teilnehmenden Kinder das Ausmaß der subjektiven Betroffenheit von depressiven Symptomen beurteilen. Zwar wird der DTK als ein Messinstrument betrachtet, das primär angesprochene Messinstrument sind aber die Schulkinder selbst. Sie sollen ihre persönliche Belastung von Symptom zu Symptom der Fragenliste im DTK erinnern, reflektieren und zu einer itemspezifischen Entscheidung bringen. Ob sie sich während der Durchführung des DTK so verhalten oder überhaupt verhalten können, ist eines der vielen Validitätsprobleme von Fragebogenuntersuchungen.

Informationen der dritten Art sehen vom Gesamtpunkt看 nicht ab, zeigen aber darüber hinaus, durch welche Konfiguration an Einzelantworten der individuelle Wert zustande kommt und ermöglichen den Vergleich zu den Konfigurationen anderer Teilnehmer.

Die erste Forschungsfrage der die hier geführte Untersuchung (siehe Kapitel 4) folgt, ist die nach Hinweisen und Indizien zum auch bei griechischen Schulkindern zu vermutenden Auftreten von depressiven Störungen. Solche Indizien lägen vor, wenn GPW oberhalb eingesetzter Cut-Offs auftreten sollten. Diese Forschungsfragen korrespondieren mit den Informationen, die hier eben der zweiten Art verfügbarer Daten zugerechnet worden waren. So wird in den nachfolgend darzustellenden Analyseschritten zunächst auf die individuellen Gesamtpunktwerte (GPW) eingegangen werden, die von den Teilnehmern, also von den Mitgliedern der untersuchten Stichprobe, mit der Beantwortung des DTK erreicht wurden.

Die Daten über die von den einzelnen Teilnehmern mit ihren Ja-Antworten erreichten Gesamtpunktwerte liegen, wegen der Verwendung von drei Skalen im DTK in mehreren Varianten vor. Das ergibt sich für den DTK zum einen wegen der mit seiner Konstruktion bestimmten Verwendung von drei inhaltlich unterschiedlichen Skalen. Zum anderen wird mit dem Manual des DTK (Rossmann, 2005) die Bildung einer zur Datenanalyse zu benutzenden vierten Skala durch die Kombination $Skala4 = Skala\ 1 + Skala\ 2$ empfohlen. Somit sind dann auch die folgenden Verteilungen von Gesamtpunktwerten zu berücksichtigen:

1. Gesamtpunktwerte nach Skala 1 (S1-GPW),
2. Gesamtpunktwerte nach Skala 2 (S2-GPW),
3. Gesamtpunktwerte nach Skala 3 (S3-GPW) und
4. Gesamtpunktwerte nach Skala 1+2 (S1+2-GPW).

Die Analyseschritte zu der Frage nach den auf den einzelnen Skalen von 2034 Teilnehmerinnen und Teilnehmern erreichten Gesamtpunktwerten folgen der hier zuvor dargestellten kategorialen Ordnung.

7.3 Analyse der über die Skala 1 des DTK in 2034GR erhobenen Daten

Mit der Durchführung des DTK in der Untersuchung 2034GR wurden Daten bekannt, die innerhalb der Datenmatrix gesichert, bis hier unerkannte Informationen tragen, von denen erwartet wird, dass sie zur Beantwortung der mit dieser Untersuchung verfolgten Forschungsfragen beitragen können. Die nachfolgende darzustellende Analysearbeit soll die Daten aus der Skala 1 des DTK aufschließen und so die von ihnen getragenen Informationen aufnehmbar werden lassen. Diese Daten tragen Informationen aus dem Antwortverhalten der an der Untersuchung beteiligten Schulkinder.

Das Antwortverhalten könnte durch die Aufnahme unterschiedlicher Beobachtungen und durch danach vorzunehmende Ableitungen verschiedener Kennwerte dokumentiert worden sein. Beispielsweise variiert das Antwortverhalten der Schulkinder bei der Bearbeitung des DTK unter anderem in dem von Item zu Item unterschiedlichen Zeitbedarf zur Beantwortung der jeweils gestellten Frage und als Folge davon auch in der Summe aller Bearbeitungszeiten und damit im individuellen Gesamtaufwand an Zeit zur Bearbeitung des DTK. Diese und auch andere Merkmale der interpersonal unterschiedlichen Bearbeitung des DTK werden in der standardisierten Durchführung des Tests nicht erhoben. Ausschließliche Basis der Informationen über die Bearbeitung des DTK sind die auf 55 Items und 3 Skalen bezogenen Ankreuzungen von Ja- oder Nein-Antworten auf dem Antwortbogen des DTK. Jede Ja-Antwort wird als ein Punkt

verrechnet und ihre Summe als Gesamtpunkt看 bezeichnet. Dann könnten bei diesem Test zumindest fünf Summenwerte gebildet werden, für jede der drei Skalen 1 bis 3 ein skalenspezifischer Gesamtwert und dann über die Summierung der drei Skalensummen ein so berechneter vollständiger Gesamtpunkt看 und der GPW aus der Summierung der Skalen 1+2.

7.3.1 Verteilung der über die Skala 1 des DTK bekannt werdenden skalenspezifischen Gesamtpunktwerte (S1-GPW)

Einige erste Informationen über das Antwortverhalten der mit der gezogenen Stichprobe untersuchten Kinder sind nachfolgend der Verteilung von Antworten auf die Fragen der DTK-Skala 1 „Dysphorie/Selbstwert“ zu entnehmen. Gefragt wird, welche Gesamtpunktwerte die Mitglieder der Stichprobe auf Skala 1 erreicht haben. Geantwortet wird mit der Darstellung der Verteilung der festgestellten Gesamtpunktwerte auf Skala 1. Dadurch werden auch die sich ergebende Verteilungsform und die sie beschreibenden Kennwerte bekannt.

Von den insgesamt 2034 befragten Kindern antworteten 2016 Kinder (99,1% der Stichprobe) auf alle 25 Fragen der DTK-Skala 1 „Dysphorie/Selbstwert“. Die nicht berücksichtigten 18 Schulkinder gaben jeweils für ein Item der Skala 1 oder auch für mehrere aus nicht bekannten Gründen keine Antwort. Da der Gesamtpunkt看 (GPW) als Summe der mit Ja beantworteten Items definiert wurde, kann nach dem Auslassen der Entscheidung bei einem Item oder bei mehreren auch kein GPW berechnet werden. Die auf Skala 1 bezogene individuelle Summe von Ja-Antworten dieser 18 Kinder ist nach den gewählten Auswertungsbedingungen – zunächst jedenfalls – nicht darstellbar.

Die 2016 Kinder, die alle Fragen der Skala 1 beantworteten, erreichen auf Skala 1 einen Gesamtpunkt看, der in Absetzung von ähnlichen Werten als „Skala 1-Gesamtpunkt看“ (S1-GPW) bezeichnet wird. Die von diesen Kindern erreichten S1-GPW sind in Tabelle 7.1 als Frequenzliste für die erreichten S1-GPW aufgeführt. Das dann nachfolgende Histogramm (Abbildung 7.1) veranschaulicht die für den S1-GPW der gesamten Stichprobe gefundene Verteilung.

Tabelle 7.1: Häufigkeits- und Prozentwertverteilung der Skala1-Gesamtpunktwerte für die Stichprobe von 2016 griechischen Kindern und Jugendlichen

S1-GPW	Frequenz	Prozent	S1-GPW	Frequenz	Prozent
0	63	3,1	13	81	4,0
1	90	4,5	14	57	2,8
2	143	7,1	15	52	2,6
3	158	7,8	16	44	2,2
4	156	7,7	17	33	1,6
5	173	8,6	18	29	1,4
6	139	6,9	19	23	1,1
7	150	7,4	20	14	,7
8	155	7,7	21	12	,6
9	132	6,5	22	9	,4
10	107	5,3	23	5	,2
11	117	5,8	24	3	,1
12	70	3,5	25	1	,0
			Total	2016	100,0

Zeichenerklärung: S1-GPW: Skala1-Gesamtpunktwerte

Zu dieser Verteilung gehören die nachstehenden Kennwerte.

Mean	7,760	Range	25,00	Skewness	,674
Std. Err. of Mean	0,112	Maximum	25,00	Std. Error of Skewness	,055
Median	7,000	Minimum	,00	Kurtosis	-,037
Mode	5,00	Sum	15645,00	Std. Error of Kurtosis	,109
Std. Dev.	5,01627	Variance	25,163		

Obschon ein solcher Vergleich an dieser Stelle nicht zwingend notwendig wäre, werden nachfolgend die Gesamtpunktwerte auf Skala 1 des DTK aus der hier erörterten griechischen Untersuchung mit den entsprechenden Daten der deutschen Normstichprobe zum DTK (Rossmann, 1993) verglichen.

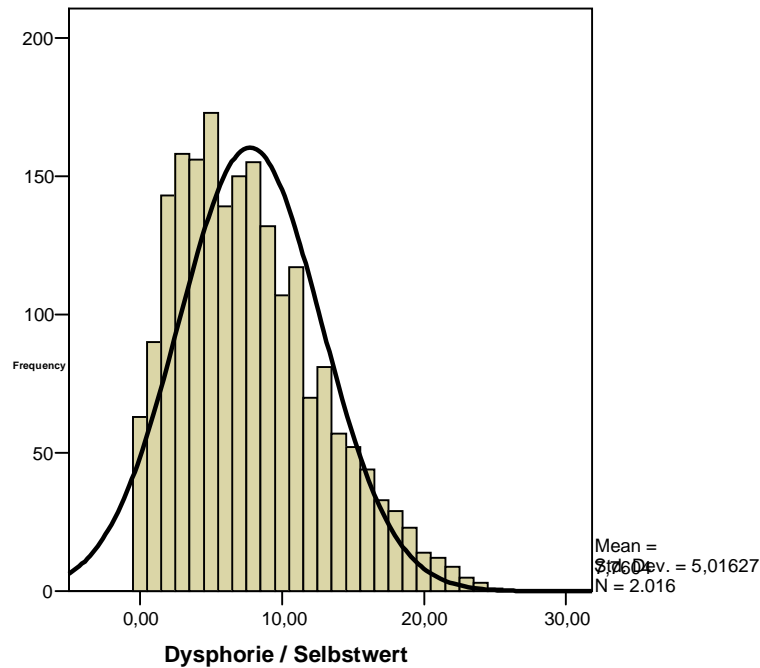
Tabelle 7.2: Auflistung der S1-GPW aus zwei Untersuchungen: Rossmann, deutschsprachig, (1993) und Kampisiou, griechisch, (2003)

Skala 1: Dysphorie/Selbstwert			
Punkte	Prozentrang (Rossmann, 1993)	Prozentrang (Kampisiou, 2003)	Unterschied*
0	4.2	3.1	1.1
1	10.5	7.6	2.9
2	18.3	14.7	3.6
3	26.2	22.5	3.7
4	33.3	30.3	3.0
5	41.2	38.8	2.4
6	48.3	45.7	2.6
7	55.6	53.2	2.4
8	62.1	60.9	1.2
9	67.3	67.4	-0.1
10	72.6	72.7	-0.1
11	77.0	78.5	-1.5
12	81.5	82.0	-0.5
13	85.0	86.0	-1.0
14	87.7	88.8	-1.1
15	91.0	91.4	-0.4
16	93.1	93.6	-0.5
17	95.3	95.2	0.1
18	96.6	96.7	-0.1
19	97.6	97.8	-0.2
20	98.3	98.5	-0.2
21	98.8	99.1	-0.3
22	99.3	99.6	-0.3
23	99.8	99.8	0.0
24	99.9	100.0	-0.1
25	100.0	100.0	0.0

Zeichenerklärung: Unterschied* = Prozentsatz bei Rossmann - Prozentsatz bei Kampisiou

In der nachstehenden Abbildung ist für $n = 2016$ griechische Schulkinder die Verteilung der Gesamtpunktwerte aus der Beantwortung der Fragen auf der Skala 1 des DTK in der Untersuchung Kampisiou (2003) dargestellt.

Abbildung 7.1: Verteilung der Gesamtpunktwerte in der DTK-Skala „Dysphorie/Selbstwert“



Die Nachfrage, ob die sich aus der Verteilung der S1-GPW ergebende Kurve einer Normalverteilung entspreche, korrespondiert einerseits mit üblichen statistischen Vorgehensweisen. Geprüft wird dann beispielsweise, ob die Voraussetzungen für die Anwendung bestimmter statistischer Verfahren, die Normalverteilung der Daten verlangen, gegeben sei. Die an dieser Stelle vorzunehmende Prüfung auf Normalverteilung gründet im Gegensatz zu dem angeführten Beispiel auf Erwartungen, die mit einem Testverfahren verbunden werden, dass sensibel gegenüber dem Auftreten psychischer Störungsbilder sein soll. In größeren Stichproben würden nur wenige Personen die als Hinweis auf das Bestehen einer solchen Störung anzusehenden Testpunktwerte erreichen. Eine vielbenutzte Suchstrategie solcher Tests ist die n-fach wiederholte Frage nach dem Auftreten spezifischer Symptome. Von einem, von zwei oder drei dieser Symptome kann so gut wie jeder aus der Gesamtbevölkerung und der sie repräsentierenden Stichprobe betroffen sein, kaum aber, so die der Vorgehensweise unterliegende Suchstrategie, von einer Vielzahl dieser Symptome. Da nun aber alle Fragen sich auf störungsspezifische Symptome beziehen, sinkt nach den ersten Fragen mit jeder weiteren Testfrage die Wahrscheinlichkeit, dass alle Befragten dieser n-fachen Häufung als für sie persönlich relevant noch zustimmen werden. Daraus ergibt sich, dass mit ansteigendem Gesamtpunktwert der Prozentsatz der Befragten, die einem n-fach gehäuften Auftreten von

Symptomen noch zustimmen, immer geringer wird. Anders als bei der Normalverteilung müssen bei der Verteilung von Gesamtpunktwerten aus der Durchführung solcher Tests, die hohen und höchsten Punktwerte deutlich niedrigere prozentuale Zustimmung finden als die entsprechenden Punktabschnitte im Anfangsteil der Skala. Die ersten 5 Gesamtpunktwerte sollten prozentual deutlich höher besetzt sein als die letzten 5 Gesamtpunktwerte. Das ist in der oben dargebotenen Frequenztafel für die S1-GPW deutlich erkennbar. Die Punkte 0, 1, 2, 3 und 4 zum Anfang der Skala werden von 30,2% der beteiligten Schulkinder besetzt. Die Punktwerte 21, 22, 23, 24 und 25 zum Ende der Skala vereinen insgesamt nur 1,3% der beteiligten Kinder auf sich. Da über die Tests dieser Art der Verdacht auf das Bestehen einer psychischen Störung an die vom Probanden jeweils zu gebende Bestätigung des Auftretens einer in der subklinischen Stichprobe nicht zu vermutenden Gesamtzahl n-fach gehäufte Symptome gebunden ist, darf die deutliche Abweichung der Kurve der Gesamtpunktwerte von Normalverteilungen gewissermaßen als Kennzeichen solcher Tests angesehen werden. Für die Verteilung der S1-GPW muss sich dementsprechend ergeben, dass die nachstehende H_1 – Hypothese zu bestätigen ist: „Die von der griechischen Stichprobe erreichten Skala1-Gesamtpunktwerte unterscheiden sich von denen aus einer Normalverteilung zu erwartenden signifikant, sie entstammen nicht einer normalverteilten Grundgesamtheit.“

Der durchgeführte Kolmogorov-Smirnov Test zeigt, dass die eingeführte H_1 -Hypothese zu verifizieren ist (Kolmogorov-Smirnov $Z = 4,370$, $p = 0,000$). Die Positionen des arithmetischen Mittels ($AM = 7,76$), des Modalwertes ($Mo = 7,0$) und des Medianwertes ($Md = 5,0$) verweisen auf die nach der Testkonstruktion des DTK zu erwartende linkssteile Verteilung, die durch den negativen Exzeß (Kurtosis = $-0,037$), wiederum theoriekonform (s.o.), im Vergleich zur entsprechenden Normalverteilung als flacher zu charakterisieren ist (Bortz, 1999, 5. Aufl., S. 34 und S. 46; Brosius und Brosius, 1998, S. 303).

Die gefundene Verteilungsform stimmt mit dem entsprechenden Ergebnis aus der Normstichprobe von Rossmann überein (Rossmann, 1993, S. 9). Sie entspricht damit (s.o.) den theoretischen Erwartungen, die man im Hinblick auf die Anwendung eines zur Erfassung psychopathologischer Symptome konstruierten Messverfahrens (Depressionsskala) in einer eher unauffälligen Stichprobe von Schulkindern haben durfte.

In Zusammenhang mit dieser von der Testkonstruktion gewollten und somit theoriekonformen linkssteilen Verteilungsform der Skala1-Gesamtpunktwerte stehen die bei der Itemanalyse auffallenden Items mit relativ geringen Lösungshäufigkeiten. Auch Rossmann sieht für die Werte in seiner Normstichprobe die relativ niedrigen Lösungshäufigkeiten der zur Skala1 gehörenden Items als im Zusammenhang mit den Zielen der Testkonstruktion stehend. Die Itemanalyse der Daten aus der griechischen Stichprobe bestätigt diesen Zusammenhang.

Die aus den mitgeteilten Daten entnommenen Informationen über das auf den Skala1-Gesamtpunkt bezogene Antwortverhalten der in der untersuchten Stichprobe vereinten griechischen Schulkinder können zu den nachstehend aufgeführten Arbeitsergebnissen zusammengefasst werden:

7.1 AS: Das Antwortverhalten der befragten 2034 griechischen Schulkinder zu den Items der Skala 1 des DTK erlaubt die Berechnung einer Variablen von individuellen Skala 1 - Gesamtpunktwerten (S1-GPW).

7.2 AS: Die S1-GPW des DTK erweisen sich in ihrem Auftreten als nicht normalverteilt. Ihre spezielle linkssteile Verteilungsform mit einem signifikant unter dem Verlauf der Normalverteilung abgeflachten Bereich hoher Punktwerte entspricht den gesetzten Vorgaben aus

der Testkonstruktion. Indem die griechische Version der Verteilungsform mit der deutschsprachigen Normierung des DTK übereinstimmt, erweist sich, dass auch nach der Übersetzung der Items ins Griechische das dem Test unterliegende Messprinzip des DTK in der Skala 1 funktional ungemindert erhalten geblieben ist.

7.3 AS: Die mit dem DTK befragten Schüler und Schülerinnen erreichen in der griechischen Version individuelle S1-GPW, die über die gesamte Breite der Skala der Gesamtpunktwerte von 0 bis 25 streuen.

7.4 AS: Auch in der griechischen Version des DTK streuen die S1-GPW so, dass eine Majorität von 82,1% aller Mitglieder der Stichprobe auf den ersten 50% der erreichbaren Gesamtpunktwerte platziert ist. So wird die zweite Hälfte der Skala der Gesamtpunktwerte, die mit den Werten über dem Mittelwert, einer Minderheit von 17,9% der Mitglieder dieser Stichprobe überlassen.

7.5 AS: In der griechischen Übersetzung erweist sich über die von der Normalverteilung abweichende Verteilungsform der individuellen Gesamtpunktwerte auf Skala 1, dass die Konstruktionsprinzipien der deutschen Testform ungemindert erhalten sind und sich, nachgewiesen durch die Messdaten, auch in der griechischen Anwendung als sensibel gegenüber dem Auftreten der relativ seltenen Ereignisse einer individuellen Bestätigung von subjektiv gehäuft auftretenden Skala 1-Symptomen erweisen. Das rechtfertigt den Rückgriff auf die für den deutschsprachigen DTK bekannt gewordenen Validitäten.

7.6 AS: Die Items der Skala 1 des DTK sind dem Konstrukt „Dysphorie und Selbstwert“ depressionsspezifisch zugeordnet. Die Verteilung der S1-GPW weist insoweit für die Kinder der untersuchten Stichprobe inhaltsvalide nach, dass 45,7% von ihnen die ihrem Erleben auftretenden unterschiedlichen Symptome auf Zahlen von 0 bis 6 einschließlich beschränken. Weitere 43,0% wählen Angaben zwischen 7 und 14 einschließlich. Etwa 11% geben Werte von 15 bis 25 einschließlich.

7.7 AS: Nach den Selbstzuweisungen der Schulkinder dieser Stichprobe sind etwa 89% von ihnen mit depressionsspezifischen Symptomen zum Konstrukt Dysphorie und Selbstwert mit einer individuell zutreffenden Anzahl belastet, die zu den unteren 66% der wählbaren Mengen gehört (0 bis 14 von 25 Items). Etwa 11% der Kinder geben eine Belastung an, die durch eine Zahl an Symptomen repräsentiert wird, die zu den 33% an höchsten wählbaren Werten gehört (15 bis 25).

7.8 AS: Nach den Selbstangaben der untersuchten 2016 griechischen Schulkinder sind etwa 11% von ihnen durch Zahlen depressionsspezifisch unterschiedlicher Symptome belastet, die aus den nur von diesen Kindern genutztem oberen Drittel der wählbaren Anzahlen bestehen.

Die aufgeführten acht Aussagen zur Verteilung der Gesamtpunktwerte auf der Skala 1 des DTK entsprechen den Daten. Als von den teilnehmenden Schulkindern selbst ausgewählte Aussagen haben sie, bezogen auf den Prozess der Selbstzuschreibung, uneingeschränkte Validität. Die angegebenen Zahlen von Schulkindern haben die über den S1-GPW bekannt werdenden Mengen an Ja-Entscheidungen mitgeteilt.

7.3.2 Der individuelle Gesamtpunkte aus der Skala 1 des DTK als Ausprägung des Skalenkonstruktes in der Persönlichkeit (S1-GPW)

Der DTK ist ein Depressionstest für Kinder und danach ein Instrument zur Erfassung von Informationen über das Auftreten dieses Störungsbildes. Die Skala 1 ist ein Teil des Gesamttestes. Aber als ein solcher Teil repräsentiert sie mit ihrer Anwendung eigenständig Informationen über eine von den Persönlichkeiten jeweils bezugte individuelle Ausprägung des Skalenkonstruktes.

Dem folgend, wird betrachtet, welche Informationen die GPW aus Skala 1 zu diesen individuellen Ausprägungen geben.

Mit der Analysearbeit an den Daten des DTK wird versucht, die so aufzuschließen, dass alle von ihnen getragenen Informationen erfassbar werden. Dem stünde entgegen, dass nur die Informationen gesucht werden, nach denen gefragt worden ist. Empirisch effizienter scheint die Suche nach allen vorhandenen Informationen zu sein. Erst wenn der Inhalt von Informationen bekannt geworden ist, kann versucht werden zu bestimmen, welche von ihnen in welcher Weise dem Erreichen des Untersuchungszieles dienlich sein können.

In diesem Zusammenhang wird hier zunächst zwischen expliziten und impliziten Informationen unterschieden. Explizit sollen Informationen heißen, die mit den Daten unmittelbar gegeben werden. Das sind die jeweils bekannt werdenden Gesamtpunktwerte. Die geben primär die Zahl der individuell abgegebenen Ja-Antworten bekannt. Weitere Informationen ergeben sich aus der Verteilung der GPW in der jeweils durchgeführten Untersuchung. Die S1-GPW vermitteln unmittelbar, wie viel Ja-Antworten ein teilnehmendes Schulkind bei der Bearbeitung der Items zur Skala 1 gegeben hatte. Unmittelbar zugänglich sind auch Informationen über das Antwortverhalten der untersuchten Stichprobe.

Implizite Informationen werden dem gegenüber hier alle die genannt, die sich nicht unmittelbar aus den Gesamtpunktwerten ergeben, aber erkennen scheinen, wenn externe Informationen hinzugezogen werden. Die in den vorangegangenen Ausführungen mehrfach erwähnte Untersuchung von Pötz bringt extern ermittelte Kenntnisse über den DTK, die auf die eigenen Untersuchungsdaten angewendet werden können. Den eigenen Daten lassen sich daraufhin Bedeutungen zusprechen, die sich aus dem extern erarbeiteten Wissen ableiten und übertragen lassen. Solche impliziten Informationen haben generell den Mangel, aus einer anderen als der eigenen Untersuchung zu kommen, was immer die Frage aufwirft, ob solch Wissen auf die eigene Untersuchung zutrifft. Das aus der Untersuchung von Pötz bekannt gewordene Wissen ist an einer österreichischen Stichprobe erarbeitet worden und eben nicht an einer griechischen. Im Übrigen ist der Stichprobenumfang jener Untersuchung so begrenzt, dass Stichprobenfehler bei der Generalisierung sehr wahrscheinlich scheinen. Dennoch sind Informationen aus externen Stichproben aus mehreren Gründen nicht nur unersetzbar, sondern auch methodisch wünschenswert. Nachfolgend wird nach Informationen gefragt, die sich aus den GPW der Skala 1 des DTK in der betrachteten Untersuchung unter Nutzung aller heranziehbaren Quellen ergeben.

Für die S1-GPW sind auch einige Informationen bekannt geworden, die auf alle Verteilungen von S1-GPW zu beziehen wären und ihnen über ihren numerischen Gesamtpunkt看wert hinausreichende Bedeutung geben oder geben könnten.

Eine solche außerhalb der in der Untersuchung 2034GR durchgeführten Anwendung des DTK liegende Information enthält die hier in diesen Ausführungen bereits zuvor zitierte Untersuchung von Zimmermann (s.o. Zimmermann, bei Rossmann, 2005). Die Untersuchung von Zimmermann gibt Informationen, die den Gesamtpunktwerten S1-GPW aus der Untersuchung 2034GR eine spezifische Bedeutung zusprechen. Die durch Zimmermann aufgefundenen Informationen sollten allein schon deswegen durch Daten aus weiteren und dann auch aus griechischen Quellen bestätigt werden. Die von ihm geführte Untersuchung weist nach, dass die GPW aus Skala 1 des DTK mit den von ihm beachteten Störungsbildern variieren. Dann aber kommt den Messwerten auch eine eigene skalenspezifische Bedeutung zu. Das Skalenkonstrukt hat eine eigene psychologische Bedeutung. Zimmermanns Daten sprechen dafür, dass die DTK Skalen 1 und 2 valide messen. Sie könnten als Indikatoren für die erlebte Ausprägung des Skalenkonstruktes in der Persönlichkeit angesehen werden.

Da sich nach Zimmermann die von ihm untersuchten Kinder mit unterschiedlichen Störungsbildern in den DTK Werten störungsspezifisch unterscheiden, ist zu erkennen, dass nicht nur das Erreichen von Cut-Off Werten spezifische Informationen bietet, sondern auch schon die absolute Höhe der individuellen Punktwerte. Darauf verweisen die Daten von Zimmermann, weil sich die von ihm nach Störungsbildern unterschiedenen Teilgruppen bereits unterhalb von definierten Trennwerten auf den Skalen des DTK unterscheiden. Zu vermuten wäre also, dass die mit den Punktwerten aus Skala 1 vermittelten Informationen insgesamt störungsspezifische Hinweise bieten. So könnte beispielsweise ein niedriger Punktwert auf Skala 1 die Wahrscheinlichkeit dafür senken, dass ein diesen Wert erreichendes Kind bestimmten Störungsbildern zuzurechnen wäre, dafür dann möglicherweise um so wahrscheinlicher einem anderen oder auch keinem zuzuordnen wäre.

Skala 1 ist mit ihren depressionsspezifischen Items dem Konstrukt „Dysphorie/Selbstwert“ zugeordnet. Die Skala misst unstrittig die Zahl der von den jeweiligen Teilnehmern nach ihrer eigenen Einschätzung als qualitativ und quantitativ für sie zutreffende Mengen an unterschiedlichen Symptomen. Gewissheit besteht über die Selbstzuschreibung, allerdings nicht über deren Validität. Die Ergebnisse aus der zitierten Untersuchung von Zimmermann sprechen dafür, dass nicht allein das Erreichen von Trennwerten (cut-offs) verwertbare Informationen trägt und vermittelt, sondern auch schon der individuell erreichte Punktwert. Die individuelle Position auf der Skala 1 des DTK wäre als klinisch relevante Information zu betrachten. In diesem Zusammenhang ist als klinisch bedeutsam nicht nur der Nachweis bestehender Störungsbilder zu verstehen, sondern ebenso ein etwa erreichter valider Beleg für die Abwesenheit von psychischen Störungen.

Die von Zimmermann mitgeteilten Daten bekräftigen die Annahme, dass alle auf der Skala 1 durch GPW einnehmbaren Positionen unterschiedliche Ausprägungen des Skalenkonstruktes repräsentieren.

Das wird bei Rossmann durch die Bezeichnung „Dysphorie/Selbstwert“ verbal repräsentiert. Eine präzisere Bestimmung des Konstruktes bieten die Items der Skala 1.

Die Persönlichkeiten mit hohen Punktwerten geben dann nicht nur an, ihnen seien aus eigenem Erleben eine Vielzahl unterschiedlicher Symptome dieser Art bekannt, sie wären auch durch sie von ihnen mit steigender Punktzahl in steigender Symptomvielfalt und in steigender Stärke betroffen. Wenn dem nicht so wäre, könnten die S1-GPW nicht in den von Zimmermann diagnostisch unterschiedenen Teilstichproben gruppenspezifisch unterschiedliche Werte auf den DTK Skalen erreichen. Die sich in der Untersuchung von Zimmermann theoriekonform präsentierenden Höchstwerte aus Skala 1 und Skala 2 sprechen für die Konstruktvalidität der Skalen 1 und 2 im DTK.

Für die Skala 3 bietet die Untersuchung von Zimmermann keine empirischen Argumente für deren skalenspezifische Konstruktvalidität an. Die von Zimmermann in die Untersuchung einbezogenen vier Gruppen mit unterschiedlichen Diagnosen nach ICD 9 unterscheiden sich mit ihren gruppenspezifischen Mittelwerten auf Skala 3 nicht signifikant voneinander. Zu vermuten ist, dass die von den Symptomen auf Skala 3 angesprochene oder definierte Konzeptualisierung körperlicher Beanspruchung ein für viele psychische Störungen zutreffendes Syndrom ist, wie vergleichsweise erhöhte Temperaturen oder Fieber auch Begleitumstände vieler Erkrankungen sind und in dieser allgemeinen Form nicht als differentielles Diagnostikum verwendbar sind. Vielleicht wäre hilfreich, ein Störungsbild auszuwählen und in eine entsprechende Untersuchung einzubeziehen, bei dem solche somatischen Auffälligkeiten nur in Ausnahmefällen auftreten. Die im Vergleich zu anderen Störungen signifikant unterschreitende Abweichung eines Mittelwertes

auf Skala 3 wäre das theoriekonform vorhersagbare Ereignis, das im Versuch zu falsifizieren oder zu verifizieren wäre.

Alle Vorbehalte gegen die Annahme der skalenspezifischen Konstruktvalidität bei Skala 1 und Skala 2 des DTK, die weitere und vor allem auch griechische Belege für diese Validitätsannahmen fordern, sind zu akzeptieren. Solange aber nicht empirische Gegenargumente vorliegen, darf das der Untersuchung von Zimmermann folgende Argument für die durchgängige Konstruktvalidität der Skalen 1 und 2 des DTK als aussagekräftig vermutet werden. Es spräche zumindest für die Beachtung der mit diesem Argument geöffneten Perspektive auf den Eigenwert von Daten der Skalen 1 und 2 des DTK.

Danach bestünde die Möglichkeit, die nachfolgende Aussage den vorangegangenen acht Aussagen zu den Messergebnissen auf Skala 1 des DTK hinzuzufügen.

7.9 AS: Unter Berufung auf Zimmermann (siehe Rossmann, 2005) wird empirisch begründet für die Gesamtpunktwerte aus Skala 1 (und denen aus Skala 2) eine skalenspezifische Konstruktvalidität angenommen. Danach wären die S1-GPW der Untersuchung 2034GR Indikatoren für das Ausmaß einer persönlichkeitspezifischen Ausprägung von depressionsspezifischer Dysphorie und depressionsspezifischen Selbstwertproblemen.

7.10 AS: Nach der 7. Aussage wird den S1-GPW für die Teilnehmer an der Untersuchung, ungeachtet eines Beitrages der Skala zum Erkennen von Hinweisen auf das Bestehen depressiver Störungen, ein skalenspezifischer Informationswert zuerkannt. Dann können auch aus den GPW auf Skala 1, unabhängig von Validitätsproblemen im Zusammenhang mit Depression, skalenspezifische psychologische oder pädagogische Folgerungen gezogen werden.

7.3.3 Die mit den Gesamtpunktwerten auf Skala 1 verbundenen skalenspezifischen Informationen

Das Hauptziel der Datenanalyse zum DTK ist das Auffinden von Hinweisen auf das Bestehen depressiver Störungen. Aus der Erörterung der Validitätsprobleme des DTK ist bekannt, dass die Skala 1 mit den über sie erhobenen Punktwerten nach Rossmann (2005) unter Hinweis auf Pötz (2003) einen Vorhersagewert vermittelt, der unter dem der Skalenkombination 1+2 oder dem der Skala 3 bleibt. Neben dem Beitrag der Skala 1 zu den prädiktiven Leistungen des DTK im Hinblick auf depressive Störungen bietet die Skala 1 auch eigene, skalenspezifische Informationen.

7.3.3.1. Der Gesamtpunkt看wert aus Skala 1 des DTK als Hinweis auf individuelle Ausprägungen von Symptomen nach dem Skalenkonstrukt

Sowohl nach der dem DTK unterliegenden theoretischen Begründung, als auch nach entsprechenden empirischen Belegen aus der Anwendung des Verfahrens, wie etwa in der zitierten Untersuchung von Zimmermann, leistet die Skala 1 einen nicht übergehbaren Beitrag zur Messleistung des DTK. Sie positioniert die an Messungen mit dem DTK teilnehmenden Personen in einer Rangreihe der von ihnen erreichten Gesamtpunktwerte. Diese Position trägt, wie unter anderem die weiter oben zitierten Untersuchungen von Zimmermann und Pötz nachweisen, störungsspezifische Informationen. Störungsspezifisch sind Informationen, wenn sie Hinweise zum mehr oder minder wahrscheinlichen Auftreten oder Nichtbestehen von Störungen geben.

Insoweit besteht auch Anlass, die erfasste Verteilung der GPW aus Skala 1 eingehender zu betrachten.

Die durch den DTK verlangte Selbstzuschreibungen von Symptomen aus den Itemlisten der jeweiligen Skala geben auf Skala 1 die Wahl zwischen 0 bis 25 auf den Symptomkomplex Dysthymie/Selbstwert bezogene Items als relevant für die eigene Person auszuwählen. Die Items sind so formuliert, dass Zustimmung über eine Ja-Antwort die Information weitergibt, man sei selbst vom Auftreten des jeweiligen Phänomens betroffen. Der jeweils erreichte Gesamtpunktwert gibt an, wie viele Items das jeweilige Schulkind aus der Liste auf Skala 1 ausgewählt hat. Die schon zuvor weiter oben dargestellte Verteilung der Gesamtpunktwerte zeigt eine Verteilung über die gesamte Breite der wählbaren Items an. Alle erreichbaren Gesamtpunktwerte haben Wahlhäufigkeiten von 1 bis 158. Mit jeder Wahl bekundet ein Schulkind, mit dem gewählten Symptom bekannt sein. Durchschnittlich wurden 7,71 Symptome gewählt.

Da die vorab überprüften Messqualitäten dessen Reliabilität bestätigten, ist davon auszugehen, dass der DTK das, was er misst, durchaus gut misst. Dann weist der Mittelwert der Gesamtpunkte auf Skala 1 nach, dass dysphorische Symptome oder Selbstwertprobleme so gut wie allen Schulkindern dieser Stichprobe bekannt sind. Immerhin haben 38,8% der Kinder lediglich 0 bis 5 Items als für sich zutreffend gewählt. Gut 1/3 der Kinder kommt mit einem Fünftel der über die Skala angebotenen Wahlmöglichkeiten aus. Fast $\frac{3}{4}$ der Kinder, nämlich 72,7% aller Mitglieder der Stichprobe wählen nicht mehr als $\frac{2}{5}$ der angebotenen Items, also nicht mehr als 10, als für sie selbst relevant. Wird betrachtet, wie viele Kinder mit 44% der wählbaren Items für ihre Wahlen auskommen, sind das 78,5% aller Kinder. Daraus gehen die folgenden Informationen hervor:

1. Phänomene, die dem Konstrukt „Dysphorie und Selbstwert“ nach Rossmann als spezifische Symptome zugerechnet werden, sind über 95% der mit dieser Stichprobe untersuchten Schulkindern bekannt.
2. Die Zahl der dem einzelnen Schulkind bekannten Symptome variiert sehr stark.
3. Über $\frac{3}{4}$ der Kinder, 78,5%, schreiben sich 0 bis 11 Symptome zu und nutzen so maximal 44% der von der Skala 1 angebotenen Items.
4. In Absetzung davon gibt 1/5 der Kinder an, mit mehr als der Hälfte der angebotenen Items persönliche Erfahrungen zu haben.

Um die Verteilung der Gesamtpunktwerte aus Skala 1 übersichtlicher darstellen zu können, werden die Werte drei Klassen zugewiesen. Das erleichtert den Überblick über die gesamte Verteilung und schärft den Unterschied in den prozentualen Anteilen an Schulkindern auf den Skalenabschnitten. Eine Kategorie 1 sammelt Kinder, die nur niedrige Mengen an unterschiedlichen Symptomen wählten. In Kategorie 2 sind dann Kinder versammelt, die zwar mehr Symptome nannten als die in Gruppe 1, sich dabei aber so verhielten, wie die relative Mehrheit der Kinder dieser Stichprobe. Gruppe 3 fasst demgegenüber Kinder zusammen, die, ganz im Gegensatz zu fast $\frac{3}{4}$ der teilnehmenden Kinder, größere Mengen an Symptomen ausgewählt hatten. Die Trennwerte zwischen den Gruppen sind am Kurvenlauf der GPW aus Skala 1 orientiert, dann aber, dem Zweck entsprechend, so gesetzt worden, wie das für die Übersicht über die Mengenverteilung günstig schien.

In der nachstehenden Tabelle sind die verwendeten Trennwerte angegeben. Die sind an Merkmale der zuvor bekannt gewordene Verteilung angelehnt.

Tabelle 7.3: Trennpunkte für drei Gruppen nach GPW aus Skala 1 DTK

Bezeichnung der Gruppen nach GPW aus Skala1 des DTK	Skala1-Gesamtpunktwerte	Selbstzuweisungen an Symptomen aus der Itemliste Skala 1
Gruppe 1 (Sk1)	0 - 6	Niedrige Frequenz an Selbstzuweisungen von 0% bis 24% der Symptome
Gruppe 2 (Sk1)	7 - 13	Mittlere Frequenz, bis zu 52% der Symptome
Gruppe 3 (Sk1)	14 - 25	Hohe Frequenz an Selbstzuweisungen von 53% bis 100% wählbarer Symptome

Werden die genannten Trennwerte genutzt, ergibt sich die nachstehende prozentuale Verteilung der Mitglieder der Stichprobe auf drei Gruppen von Schulkindern mit unterschiedlichen GPW aus Skalas 1 des DTK.

Tabelle 7.4: Verteilung der Schul Kinder der Stichprobe auf drei Gruppen mit unterschiedlichen GPW aus Skala 1 des DTK, n =2016

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Gruppe1	922	45,3	45,7	45,7
	Gruppe 2	812	39,9	40,3	86,0
	Gruppe 3	282	13,9	14,0	100,0
	Total	2016	99,1	100,0	
Missing	System	18	,9		
Total		2034	100,0		

Aus der Verteilung der Kinder mit niedrigen, mittleren und hohen Frequenzen an Selbstzuweisungen von Symptomen aus der Itemliste der Skala 1 (Gruppe1, Gruppe2, Gruppe 3) ergibt sich ein hochsignifikanter Unterschied in den Frequenzen zwischen den Gruppen 1 und 2 einerseits und der Gruppe 3 andererseits. Die Mitglieder von Gruppe 1 und 2 stellen zusammen einen Anteil von 85,2% der Mitglieder der Stichprobe und nutzen den Bereich von 0 bis 52% der wählbaren Items für ihre eigenen Selbstzuweisungen. Die darüber liegenden 48% der Gesamtpunktwerte werden von einer Minderheit der Kinder genutzt, die insgesamt nur einen Anteil von 14% an der Stichprobe halten. Danach sind den vorangegangenen zehn Aussagen zwei weitere hinzuzufügen.

7.11 AS: Durch die Verteilung der GPW aus Skala 1 des DTK erweist sich, dass 85,2% der Kinder dieser Stichprobe ihre Selbstzuweisungen auf die unteren 52% der insgesamt wählbaren Mengen an Symptomen der Itemliste von Skala 1 beschränken. Die restlichen 14% der Kinder nehmen die verbleibenden 48% an höheren Mengen an Selbstzuweisungen für sich in Anspruch.

7.12 AS: Die durchschnittliche Frequenz der Kinder, die einen Gesamtpunkt看wert auf unterschiedlichen Teilbereichen der Verteilung von GPW aus Skala 1 in Anspruch nehmen, ist ein informativer Wert über die Besetzung von Teilabschnitten der gesamten Verteilung an GPW. In dieser Untersuchung sind die Gesamtpunktwerte von 0 bis 6 einschließlich jeweils durchschnittlich von 131,72 Kindern frequentiert. Auf die Items von 7 bis 13 entfallen durchschnittlich pro GPW 116 Kinder. Für die GPW von 14 bis 25 einschließlich ist die durchschnittliche Frequenz pro GPW 23,5 Kinder.

7.13 AS: In der durchgeführten Untersuchung mit griechischen Schulkindern sind Selbstzuschreibungen von mehr als 13 Symptomen aus der Liste der Skala 1 seltene Ereignisse.

7.14 AS: Nach den in dieser Untersuchung vorliegenden Ergebnissen zur Skala 1 des DTK wird die Empfehlung ausgesprochen, Werte von mehr als 13 GPW in Skala 1 als kritische Ereignisse einzuordnen, die im Interesse der betroffenen Kinder als Anlässe zur psychologischen Beratung gesehen werden sollten.

Die Aufteilung der GPW aller Kinder auf Gruppe 1, 2 und 3 verdeutlicht die Merkmale der Verteilung dieser Werte. Die Kinder der Gruppe 3 schreiben sich selbst im Vergleich mit Gruppe 1 fast 4,9-fach höhere und im Vergleich zur Gruppe 2 noch fast 1,8-fache höhere Mengen an Symptomen zu. Für die Mitglieder der Gruppe 3 besteht mit hoher Wahrscheinlichkeit psychologischer und pädagogischer Beratungsbedarf. Die Möglichkeit besteht, dass für eine nicht bekannte Zahl dieser Kinder ein kritischer Notstand besteht, der Beziehungen auch zur Familiensituation, zur Schule und zur sozialen Umgebung am Wohnort haben könnte.

Die nachstehende Tabelle zeigt die Verteilung der gruppenspezifisch selbst zugewiesenen Items der Symptomliste aus Skala 1 und die durchschnittliche Zahl der Selbstzuweisungen von Symptomen in den gebildeten Gruppen.

Tabelle 7.5: Durchschnittliche Selbstzuweisungen an Items der Skala 1 pro Kind in den Gruppen 1, 2 und 3

	Gruppe1 (NFS1)	Gruppe2 (MFS2)	Gruppe3 (HSF3)	Summe
gew. It.	3173	7728	4744	15645
N	922	812	282	2016
It. pro Kind	3,44	9,52	16,82	7,76

Zeichenerklärung: gew. Items = gewählte Items; Items pro Kind = gew. Items / n;
n = Anzahl Kinder in der Teilgruppe

7.3.3.2 Modell einer verteilungsinternen Bestimmung von Trennwerten

Mit der Durchführung des DTK werden die Punktwerte auf den einzelnen Skalen und die daraus zu berechnenden Gesamtpunkte eines jeden teilnehmenden Kindes bekannt. Die unterschiedlichen Gesamtpunkte, GPW, informieren darüber, welche individuellen Mengen an Ja- und Nein-Antworten von jedem Kind gegeben worden sind. Die GPW und die in ihnen enthaltenen Konfigurationen von Items mit Ja-Antworten tragen Informationen, die nutzbar und zur Beantwortung verschiedener Forschungsfragen nützlich sind. Alle Forschungsfragen, zu deren Beantwortung eine Trennung der teilnehmenden Kinder in solche mit und ohne mutmaßlich klinisch relevanten Ausprägungen depressiver Störungen verlangt wird, benötigen dazu Informationen aus empirischen Untersuchungen zur Validitätsbestimmung der GPW des DTK. Die jeweils auftretenden Gesamtpunkte werden dann in ihrer klinischen Bedeutung nach den extern bestimmten Relationen zwischen GPW und dem Auftreten klinischer Störungen gewichtet. Daraus gehen die für die entsprechenden Skalen zu verwendenden Trennwerte hervor.

Eine verteilungsinterne Form der Bestimmung von Trennwerten, die zu verwenden wäre, wenn keine Validitätsuntersuchungen bekannt sein sollten, wird nachfolgend vorgeschlagen.

Die diesem skaleninternen Trennwert unterliegende Modellvorstellung ist auf den von der Normalverteilung abweichenden Kurvenlauf für Daten aus Messinstrumenten bezogen, die nach Hinweisen auf psychische Störungen suchen. Die Verfahren, die dazu die Selbstzuschreibung der

an einer Untersuchung beteiligten Persönlichkeiten benutzen, fragen im allgemeinen nach dem Auftreten von Phänomenen, die, mit extrem geringen Ausnahmen, in der Selbsterfahrung aller Menschen oberhalb und unterhalb bestimmter Altersgrenzen bekannt sind. Die Altersbegrenzungen nehmen Lebensabschnitte von der Gültigkeit dieser Zuschreibung aus, in denen die individuellen kognitiven Leistungen den mit dem Verlangen von Selbstzuschreibungen gesetzten Anforderungen noch nicht oder nicht mehr genügen können. Zu den als Beispiele für solche Phänomene anführbaren Erlebnisinhalten gehören etwa Angst, Vorlieben, Befürchtungen, Hoffnungen, Anlässe zur Freude und zu anderen Emotionen ebenso wie viele weitere Beziehungen zwischen Selbst und Umwelt.

Die mit dem jeweiligen Verfahren gesuchten Hinweise auf Störungen sind, wie unterstellt wird, allen teilnehmenden Personen bekannt, die dann auch individuelle Ausprägungen der für sie nach eigener Wahrnehmung geltenden Betroffenheit bekunden.

Für die nicht von solchen Störungen betroffenen Persönlichkeiten ist eine Normalverteilung dieser Selbstzuschreibungen zu erwarten. In hinreichend großen Stichproben sollte sich erweisen, dass die gesuchten Hinweise innerhalb der Stichprobe aus Persönlichkeiten ohne Störungen normal verteilen.

Die demgegenüber erwarteten Ausnahmen bilden Persönlichkeiten, bei denen die Frequenz der bekundeten Hinweise in außergewöhnlicher Weise von dem abweicht, was die überwiegende Mehrheit aller Beteiligten erlebt. Die gesuchte Störung ist immer als ein Übermaß des Auftretens der für sie spezifischen Symptome definiert. Die Störung ist psychologisch ein an der Mehrheit einer Population bestimmter Defekt. Hunger kann jeder verspüren, Störungen des Essverhaltens gehen einher mit einer nach unten oder oben abweichenden Frequenz des Erlebens von Hunger in der Population. Aus diesem zu unterlegenden Sachverhalt entspringt die Verteilung der Daten aus entsprechenden Messungen. Deren Kurve flacht in Richtung auf das Maximum der Messungen zunehmend ab. Unterstellt wird hier, dass die Abflachung ab dem Wert klinisch beachtenswert ist, der erstmals in seiner Besetzung die Besetzung des minimalen Wertes, häufig ist das der Wert 0, unterschreitet. Der Wert, der dieser Bestimmung entspricht, wird als verteilungsinterner Trennwert zwischen der in der Verteilung dominierenden Besetzungen von Messwerten und den von diesem Verteilungsgang extrem abweichenden Besetzungen von höheren Messwerten beschrieben.

Der abgetrennte Teilabschnitt hat eine durchschnittliche Besetzungsdichte (Frequenz), die signifikant unter der aller anderen in dieser Verteilung definierbaren und zum Anfang der Verteilung gerichteten Teilabschnitte liegt.

Dieser Bestimmung folgend könnte der in der betrachteten Verteilung der GPW von Skala 1 des durchgeführten DTK gesuchte Trennwert eindeutig berechnet werden. Die hier zuvor beschriebene Bestimmung, der gesuchte Trennwert sei identisch mit der Position, an der die Frequenz eines Messwertes die des Anfangspunktes der Verteilung erstmals unterschreitet, wird hier als nützliche Annäherung an die berechenbare Position eines verteilungsinternen Trennwertes aufgefasst. Nützlich ist dieser Trennwert, weil er beim Fehlen relevanter empirischer Aufschlüsse über die Validität der einzelnen Messwerte einen verfahrenslolgischen Ersatz für den gesuchten Trennwert bietet.

Der verteilungsintern bestimmte Trennwert auf Skala 1 in der hier zu analysierenden Stichprobe (2034GR von Kampisiou, 2003) ist der Gesamtpunktwert 14, der noch zur Gruppe der Kinder mit hoch ausgeprägter Belastung gehört. Demgegenüber steht GPW=13 am Anfang der in Richtung auf GPW=0 vorfindbaren ersten und größeren Gruppe, die mit ihren GPW von den extrem schwach besetzten GPW hinter GPW=13 getrennt worden ist.

Der Erklärungswert der Modellannahmen lässt sich prüfen. Der verteilungsintern bestimmte Trennwert müsste statistisch mit dem übereinstimmen, der in empirischen Validitätsprüfungen als

Trennwert für die maximal erreichbare Sensitivität der Messungen erkannt wird. Nach Pötz (s.o.) wäre dieser Wert in ihrer Verteilung der GPW 10.

Unter Hinweis auf die Untersuchung von Zimmermann (s.o.) ist daran zu erinnern, dass die durch Trennwerte bestimmte Teilgruppe von Kindern mit hohem GPW auf Skala 1 des DTK auch die umfassen kann, für die nicht keine depressive Störung wohl aber ADHS diagnostiziert worden war.

Der definierte verteilungsinterne Trennwert ist nochmals in den nachstehenden Tabellen aufgeführt, in der die Verteilung der Gesamtpunktwerte für Skala 1 auszugsweise wiedergegeben ist. Sie findet sich in voller Länge in dem vorangegangenen Abschnitt (7.3.3.1)

Tabelle 7.6: Teilbereich der Tabelle über die Verteilung der GPW aus Skala 1 des DTK in der Untersuchung 2034GR, Kampisiou (2003)

S1-GPW	Frequenz	Prozent	S1-GPW	Frequenz	Prozent
0	63	3,1	14	57	2,8
			15	52	2,6
			16	44	2,2

Zeichenerklärung: S1-GPW: Skala1-Gesamtpunktwerte

Auszugsweise wird auch die Tabelle über die Verteilung der GPW aus Skala 1 in der Untersuchung von Rossmann (1993), begleitet von den entsprechenden Daten nach Kampisiou (2003), wiedergegeben.

Tabelle 7.7: Teilbereich der Tabelle über die Verteilung der GPW aus Skala 1 des DTK in den Untersuchungen Rossmann, deutschsprachig, (1993) und Kampisiou, griechisch, (2003)

Punkte	Skala 1: Dysphorie/Selbstwert	
	Prozentrang (Rossmann, 1993)	Prozentrang (Kampisiou, 2003)
0	4.2	3.1
7	55.6	53.2
8	62.1	60.9
9	67.3	67.4
10	72.6	72.7
11	77.0	78.5
13	<i>81,5+ 3,5</i>	82.0
14	<i>85+2.7</i>	82+4,0
		86+2,8

Zeichenerklärung: kursiv = erste GPW der abgetrennten Teilgruppe von Kindern mit hohem GPW

Der Trennwert zwischen den verteilungsintern bestimmten Teilstichproben der Kinder mit hohem GPW einerseits und niedrigeren andererseits ist für Skala 1 des DTK in der Stichprobe von Rossmann (1993) der Gesamtpunkt看wert 13. Dem entspricht für diese Untersuchung ein Prozentsatz von 18,5% der Stichprobe.

Für die in griechischer Sprache geführte Untersuchung 2034GR von Kampisiou (2003) ergab sich der Prozentsatz von 14% als der damit korrespondierende Wert.

Beide Prozentsätze repräsentieren Teilstichproben von Kindern mit hohen GPW nach Skala 1, die wegen ihrer deutlichen Abweichungen von den GPW der anderen Kinder als Indikatoren für psychische Belastungen gesehen werden, die eine diesem Phänomen nachgehende psychosoziale Betreuung verlangen. Die hohe Belastung wird hier als ein hoch wahrscheinlich bestehender Risikofaktor für die aktuelle Situation dieser Kinder und für deren weitere Entwicklung gesehen.

7.3.3.3 Informationen aus den Gesamtpunktwerten auf Skala 1 im DTK über das Auftreten klinisch ausgeprägter depressiver Störungen

Für die mit dem DTK aufgenommenen Daten gilt, entsprechend den vorausgegangenen Erörterungen, allgemein die Einschätzung, sie würden Informationen tragen, die als Hinweise auf das Bestehen klinisch ausgeprägter depressiver Störungen zu verstehen wären.

Die Gesamtpunktwerte aus Skala 1 des DTK wurden bisher unter zwei verschiedenen Aspekten betrachtet. Zum einen wurde die Verteilung der Antworten der an der Untersuchung beteiligten Kinder über die Verteilung der Gesamtpunktwerte mitgeteilt. Zum anderen war die Verteilung der GPW aus Skala 1 des DTK über einen verteilungsinternen Trennwert in Teilbereiche mit hohen und weniger hohen GPW getrennt worden. In beiden Perspektiven auf die zu analysierenden Daten sind keine weiteren unmittelbaren oder mittelbaren Aufschlüsse über das Bestehen depressiver Störungen sichtbar geworden.

Zwar ist die Gefährdung einer Teilstichprobe von 14% der Teilnehmerzahl über auffallend hohe Quoten von Selbstzuschreibungen eindeutig operationalisiert und quantifiziert worden, ungewiss bleibt dennoch, wie groß die von dieser postulierten Gefährdung ausgehende Wirkung auf die betroffenen Kinder sein wird. Wenn die in den GPW der Skala 1 des DTK vermuteten Informationen über das Bestehen depressiver Störungen aufgeschlossen werden sollen, wird dazu ein empirisch gewonnener und validierter Algorithmus zur Umwandlung von GPW in quantifizierte Hinweisen auf das Bestehen von depressiven Störungen benötigt.

Da nur mit Hinweisen auf möglicherweise in der untersuchten Stichprobe bestehende depressive Störungen gerechnet wird, ist jedes erreichbare Ergebnis als Annäherung an die bestehenden Gegebenheiten einzuschätzen. Eine solche Annäherung ist die empirisch begründete Schätzung des als prozentualen Anteil an der Gesamtstichprobe vermuteten Umfangs der möglicherweise depressiv gestörten Teilstichprobe von Kindern.

Verglichen mit der internationalen Forschung ist die Vorgehensweise als eine durchaus übliche zu klassifizieren. Versucht wird, ein „Screening“ vorzunehmen. Eingesetzt wird ein depressionsspezifisches Messinstrument, das zwar keine auf Diagnosen beruhende Prävalenz ermitteln kann, wohl aber eine empirisch kontrollierte und in ihrer Qualität abschätzbare Annäherung an Prävalenzraten zu vermitteln vermag. Die Nützlichkeit dieser Vorgehensweise besteht darin, dass durch sie trotz aller methodischen Einschränkungen das Wissen über das Auftreten depressiver Störungen unter griechischen Schulkindern vermehrt wird.

Darzustellen ist nun, wie diese Schätzung vollzogen werden soll.

Der erste Schritt der Annäherung ist bereits durch die Auswahl und Anwendung eines depressionsspezifischen Messinstrumentes vollzogen worden, eben durch den Einsatz des DTK (Rossmann, s.o.).

Der zweite Schritt besteht aus der Analyse der von den griechischen Schulkindern erreichten Testwerte.

Der dritte Arbeitsschritt ist nun auf die Beantwortung der Frage nach der Bedeutung von im DTK erreichten Testwerten als Hinweise auf das Bestehen depressiver Störungen gerichtet.

Der Beantwortung kann man sich annähern, wenn man die Informationen über die Validität des DTK zur Kenntnis nimmt.

Die bisher verfügbaren Angaben über empirische Untersuchungen zur Validität des DTK sind nach Rossmann (2005) die Untersuchungen von Zimmermann (s.o.) und Pötz (2003). Wenn zu wissen gewünscht wird, welcher DTK Gesamtpunktwert mit welcher Wahrscheinlichkeit als Hinweis auf das Bestehen einer klinisch ausgeprägten depressiven Störung gelten darf, können Aufschlüsse über die nachgewiesene diskriminante Validität des DTK hilfreich werden. Informationen über die depressionsspezifische Diskriminanz von DTK Werten sind von Pötz gegeben worden.

In ihrer Untersuchung sollte versucht werden, eine Gruppe von Kindern, von denen für jedes eine Diagnose über das Bestehen einer depressiven Störung vorlag, von anderen Kindern, sogenannten Kontrollkindern, allein über die Kenntnisnahme von individuell erreichten Testwerten aus dem DTK korrekt voneinander zu trennen. Da vorab gewusst wird, dass der DTK kein Diagnoseinstrument ist, sollte realistischerweise erwartet werden, dass auch dieses Ziel nur näherungsweise erreicht werden konnte.

Da der DTK drei Skalen hat, dürfte testtheoretisch unterstellt werden, dass die eigentliche diskriminante Leistung des Verfahrens sich über die Gesamtpunktwerte als Gesamtsumme der individuell auf den drei Skalen erreichten Werte einstellen wird ($GPW_{1,2,3} = GPW_1 + GPW_2 + GPW_3$). Insoweit wäre die Nachfrage, ob Gesamtpunktwerte der Skala 1 auch schon beachtenswerte diskriminante Leistungen erreichen könnten, gewissermaßen als voreilig zu verstehen. Andererseits korrelieren Ergebnisse aus Skala 1 relativ hoch mit Werten aus dem CDI. Für dieses Messverfahren aber sind diskriminante Leistungen nachgewiesen (Kovacs, 2004).

Die Frage nach den diskriminanten Leistungen des DTK ist mit Erfolg und Misserfolg in den Bemühungen um das Auffinden entsprechender Trennwerte verbunden. Wer unter Verwendung von Gesamtpunktwerten erfolgreich trennen will, muss eine nutzbare Trennstelle auf zumindest einer empirisch gewonnenen Verteilung der Gesamtpunktwerte finden. Dieser Trennwert hat zumindest der Forderung nach einer gegenüber Zufallswerten ertragreichen Sensitivität nachzukommen. Nach Pötz (2003) lassen sich von den durch die Anwendung des DTK ermittelten individuellen Punktwerten einige GPW relativ erfolgreich als Trennwerte nutzen. Selbst der auf Skala 1 des DTK erreichte individuelle Gesamtpunktwert erreicht bei entsprechender Ausprägung diskriminante Potenzen. Dieser individuell erreichte S1-GPW ist zur Begründung von relativ treffsicheren Vorhersagen über das Auftreten von klinischen Ausprägungen depressiver Störungen einsetzbar.

Die nachstehende Tabelle gibt einen Überblick über die nach Pötz (s.o.) mit unterschiedlichen Gesamtpunktwerten von Skala 1 im DTK erreichbaren diskriminanten Trennleistungen von Kindern mit einer Diagnose über das Bestehen depressiver Störungen einerseits und Kontrollkindern ohne Diagnose andererseits.

Tabelle 7.8: Trennwerte zur diskriminanten Validität von GPW aus Skala 1 in der Untersuchung von Pötz (2003)

GPW aus Skala	Sensitivität
Skala1 > 10	75,76
Skala1 > 10	72,73
Skala1 > 11 oder Skala2 > 7	72,73
Skala1 > 17	72,73
Skala1 > 18 oder Skala2 > 5	72,73
Skala1 > 19	72,73
Skala1 > 20	72,73

Wenn der höchste Grad von Sensitivität gegenüber den Schulkindern mit dieser Diagnose erreicht werden soll, ist der Wert von 75,76 - wie er über den Gesamtpunktwert Skala 1 mit 10 erreicht wird - der für diese Skala optimale Wert. Er besagt, dass die über S1-GPW 10 von den übrigen Teilnehmern abgetrennte Teilgruppe mit entsprechender Wahrscheinlichkeit, also zu 75,76%, Kinder umfasst, die zu den in der untersuchten Stichprobe befindlichen Kindern mit diagnostizierten depressiven Störungen gehören. Pötz konnte das prüfen. Sie hatte Informationen über den diagnostischen Status der Kinder ihrer Untersuchung. Bei 24,24% war der Verdacht, wie die Nachprüfung zeigte, unberechtigt und somit nicht zu halten.

Immerhin beweist Pötz mit ihren Daten für den DTK eine nicht bezweifelbare depressionsspezifische Validität. Ob die sich in anderen Untersuchungen in gleicher diskriminativer Stärke wiederholen wird, kann über die von Pötz mitgeteilten Daten nicht vorhergesagt werden. Die von Pötz mitgeteilten Daten gelten für ihre Untersuchung. Wer ihre Ergebnisse für eine eigene empirische Untersuchung mit dem DTK nutzt, überträgt ihre Ergebnisse auf eine Stichprobe, nämlich die seine, in der andere Beziehungen zwischen Gesamtpunktwert und diagnostizierbaren Ausprägungen depressiver Störungen gegeben sein können. Jede Übernahme der Ergebnisse von Pötz ist empirisch als ein Risiko einzuschätzen. Und dennoch sollte dieses Risiko hinnehmbar sein, wenn der dadurch erreichbare Informationsstand über das mutmaßliche Auftreten depressiver Störungen bei Schulkindern größer als zuvor wäre.

Der vierte Schritt zur aktuell größtmöglichen Annäherung an das Auffinden einer Antwort auf die mit dieser Untersuchung primär verfolgten Forschungsfrage ist die Übernahme des dargestellten Risikos durch die Anwendung der Daten von Pötz auf die Ergebnisse der hier dargestellten Anwendung des DTK in der Stichprobe 2034GR aus griechischen Schulkindern. Niemand kann bis dahin die Übertragbarkeit der Trennwerte nach Pötz auf eine griechische Stichprobe nachweisen. Die Begründung für die Übernahme dieses Risikos ist ebenfalls die schon zuvor mitgeteilte. Der damit trotz aller bestehenden Ungewissheiten zu gewinnende neue Informationsstand ist größer als der zuvor gegebene.

Der von Pötz vorgeschlagene Trennwert des S1-GPW = > 10 erreicht auch in ihrer Untersuchung nur eine begrenzte Treffsicherheit. Seine Sensitivität beträgt 75,76. Mit höheren GPW fällt sie auf etwa 72. Insgesamt konnten so aber gut 75% der Kinder zutreffen eingeordnet. Sie wurden kriteriumsspezifisch erkannt. Zugleich waren aber auch fast 25% aller Zuweisungen fehlerhaft.

Das sind dann Kinder, für die ein Status angenommen wurde, weil der Testwert dafür sprach. Die Nachprüfung deckte die fehlerhafte Zuweisung auf.

Wenn hier die Daten von Pötz übernommen werden, ist auch deren Einschätzung der Sensitivität des deutschsprachigen DTK auf die griechischen Daten aus 2034GR zu übertragen. Somit ist bei einem Trennwert von $S1-GPW > 10$ mit einer Fehlerquote von etwa 25% aller Zuweisungen auszugehen.

Da in der einschlägigen Literatur für diagnostisch nutzbare Verfahren eine Treffsicherheit von mindestens 95% gefordert wird, erweist sich mit der Diskrepanz von 75% zu 95% der Abstand der über $S1-GPW$ erbrachten diskriminanten Leistung von einer nach diagnostischen Kriterien zu fordernden Treffsicherheit. Dennoch tragen die so zu gewinnenden Aufschlüsse zur Verringerung des Informationsdefizits über die zu vermutenden Prävalenzen für depressive Störungen bei griechischen Schulkindern bei.

Im Zusammenhang mit den übernommenen Risiken sind die zu gewinnenden Informationen explizit zu relativieren. Die zu erreichenden Aufschlüsse sind Informationen über das nachweisbare Auftreten von empirischen Daten, die auf die begründbare Wahrscheinlichkeit für das Auftreten einer entsprechenden Prävalenz depressiver Störungen unter griechischen Schulkindern schließen lassen. Das Auftreten der Störungen selbst ist nicht gemessen worden. In eben diesem Sinne wird hier und dann auch nachfolgend der Terminus von einer empirisch begründeten Schätzung der Prävalenz benutzt.

Mit der Übernahme des von Pötz mitgeteilten Trennwertes für die auf der Skala 1 des DTK erreichbaren individuellen Gesamtpunkte wird über den Cut-off Wert $S1-GPW = > 10$, entsprechend der weiter oben dargestellten Verteilung von GPW aus Skala 1, der empirisch begründete Hinweis auf eine mutmaßlich vorhandene Prävalenzrate von 21,5% bekannt. Dem entspricht eine statistisch quantifizierte Teilgruppe von 433,44 Kindern aus der Gesamtstichprobe mit $n = 2016$ teilnehmenden Schulkindern. Danach hat die auf der anderen Skalenseite vom Trennwert - gegenüber der „empirisch begründet geschätzten und somit mutmaßlich bestehenden Prävalenz“ – verbleibende größere Teilgruppe 1582,66 Schulkinder als Gruppenangehörige. Die mitgeteilten Werte beruhen auf Angaben aus einer externen Untersuchung (Pötz, s.o.). Deren Aussagekraft für die Untersuchung 2034GR ist nicht nachgewiesen, darf aber empirisch begründet angenommen werden. Da in der Untersuchung von Pötz über die mitgeteilten Trennwerte nur insgesamt etwa 75% der teilnehmenden Kinder richtig zugewiesen werden konnten, wäre möglich, dass auch die Angaben zur Untersuchung 2034GR mit dem Fehler von 25% falscher Zuweisungen belastet sind. Wenn diese Fehlerquote anteilmäßig auf die benannten Teilgruppen verteilt wird, entfällt auf die Teilgruppe der mutmaßlich klinisch belasteten Schulkinder etwa 1/5 dieser Quote. Demnach sollte die zutreffende Zahl der dieser Teilgruppe zugehörigen Kinder auch zwischen etwa 350 und 515 liegend vermutet werden.

Die geschätzte Prävalenz gibt den auf Grund von Hinweisen vermuteten Umfang der wahrscheinlich in klinischer Ausprägung depressiv belasteten Kinder der untersuchten Stichprobe wieder. Die Sensitivität der nach Gesamtpunkten auf Skala 1 vorgenommenen Zuweisungen an die Teilgruppe vermutlich depressiv belasteter Kinder beträgt nach Pötz 75,78 und zeigt damit an, dass etwa 25% der als im klinischen Umfang belastet eingeschätzten Kinder niedrigere $S1-GPW$ hatten als die ermittelten Trennwerte vermuten ließen. Die Spezifität der Messungen entspricht dem. Auch die als kaum belastet geltenden Kinder wurden durch den Trennwert nicht korrekt eingruppiert. Diese Kinder sollten sich mit ihren Gesamtpunkten aus Skala 1 des DTK im Zahlenraum von und mit 0 bis einschließlich 10 befinden. Der von Pötz mitgeteilte Wert für die Spezifität des Trennwertes ist aber nur 63,64 und lässt erwarten, dass 36,35% der gesuchten Kinder in diesem Zahlenraum nicht aufzufinden sein werden.

Tabelle 7.9: Sensitivität und Spezifität der diskriminanten Validität über Trennwerte auf Skala 1 des DTK nach Pötz (2003)

	Sensitivität	Spezifität
Skala1 > 10 oder skala2 > 9	75,76	63,64

Die bis hier beschriebene Analysearbeit an den die Skala 1 aus der Durchführung des DTK betreffenden Daten erbringt ihren eigenen, nachstehend zusammengefassten Beitrag zur angestrebten Beantwortung der zu Grunde liegenden Forschungsfrage nach der Verbreitung depressiver Störungen in der untersuchten Stichprobe griechischer Schulkinder.

7.3.3.4 Der nach Analyse der Gesamtpunktwerte aus Skala 1 des DTK vorliegende Informationsstand

Die bis hier vorangebrachte Analyse der die Skala 1 des DTK betreffenden Daten aus der Datenmatrix der von Kampisiou (2003) durchgeführten Untersuchung 2034GR führte zu Informationen, die im Hinblick auf die zu verfolgenden Forschungsfragen vor allem auf Hinweise zum Auftreten depressiver Störungen unter den Schulkindern der gezogenen Stichprobe geprüft worden sind.

Nach der Analyse der Daten aus der Skala 1 des DTK besteht kein Anlass für Zweifel am Auftreten von Hinweisen auf in klinischer Ausprägung bestehende depressive Störungen unter den Kindern der untersuchten Stichprobe. Die Feststellung ist zunächst offen gegenüber Varianten im Verständnis dessen, was klinische Ausprägungen sind. Die für die Setzung der verwendeten Trennwerte herangezogene Studie von Pötz (s.o.) hatte ihre Ergebnisse zur Validität des Verfahrens an diagnostizierten leichteren Ausprägungen depressiver Störungen gewonnen. Die Aussage über das Auftreten von depressiven Störungen in der untersuchten Stichprobe von griechischen Schulkindern ist weiterhin relativ gegenüber der Quantifizierung des durch Zahlen von Personen ausgedrückten Ausmaßes der Betroffenheit durch die Störung. Die aus Skala 1 aufgenommenen Hinweise sind einerseits wegen der ausschließlich subjektiven Datenquellen relativ ungeschützt gegenüber allen Formen der persönlichen Verkennung von Sachverhalten. Weiterhin ist die Quantifizierung der Betroffenheit durch den Hinweischarakter der aufgenommenen Daten behindert. Die mutmaßlich auftretenden depressiven Störungen konnten – wie bislang in allen Untersuchungen - nicht unmittelbar gemessen werden. Die versuchte Annäherung an die in der untersuchten Stichprobe bestehende und in Personenzahlen ausgedrückte Betroffenheit durch depressive Störungen muss unter Zuhilfenahme von Kennwerten für die in anderen Untersuchungen ermittelte diskriminante Validität des DTK zu berechnen versucht werden. Für diesen Vorgang wurde der Terminus von einer empirisch begründeten Schätzung der Prävalenz benutzt. Mit der Übernahme des von Pötz (s.o.) mitgeteilten Trennwertes für die auf der Skala 1 des DTK erreichbaren individuellen Gesamtpunktwerte wird über den Cut-off Wert S1-GPW=>10, der empirisch begründete Hinweis auf eine mutmaßlich vorhandene Prävalenzrate von 21,5% bekannt. Dem entspricht eine aus der Gesamtstichprobe zu entnehmende Teilgruppe von 433,44 Kindern.

Denen stehen 1582,66 Schulkinder als Gruppe gegenüber, die keine oder niedrigere Ausprägungen im Aufkommen depressiver Symptome angegeben hatten.

Da in der Untersuchung von Pötz über die mitgeteilten Trennwerte nur insgesamt etwa 75% der teilnehmenden Kinder richtig zugewiesen werden konnten, wird dieser Zuweisungsfehler auch auf

die Teilgruppen in der Untersuchung 2034GR proportional umgerechnet. Wenn diese Fehlerquote anteilmäßig auf die benannten Teilgruppen verteilt wird, entfällt auf die Teilgruppe der mutmaßlich klinisch belasteten Schulkinder etwa 1/5 dieser Quote. Demnach sollte die zutreffende Zahl der dieser Teilgruppe zugehörigen Kinder auch zwischen etwa 350 und 515 liegend vermutet werden. Die mit den über die GPW aus Skala 1 aufgenommenen Hinweise auf das Bestehen depressiver Störungen in der untersuchten Stichprobe lassen eine Quote von 17,36% bis 25,55% betroffener griechischer Schulkinder als hoch wahrscheinlich zutreffend vermuten. Dann könnten bei $n = 2016$ Kindern zwischen 1501 und 1666 von ihnen, das ist der Bereich von 74,75% bis 82,64% aller Kinder der Stichprobe, durch geringere Ausprägungen depressiver Episoden gekennzeichnet sein.

Unter Verwendung des weiter oben eingeführten verteilungsinternen Trennwertes wird ein danach im klinischen Sinne bedenklich hohes Maß an Selbstzuschreibungen depressiver Symptome nach der Itemliste von Skala 1 des DTK für eine Teilstichprobe von 14% gefunden.

Die für die Untersuchung 2034GR durchgeführte Analyse der von 2016 griechischen Schulkindern individuell erreichten Gesamtpunktwerte auf der Skala 1 des DTK führte zu einer Datenlage, die zu den nachstehenden Feststellungen berechtigt:

1. Wenn die von Schulkindern auf der Skala 1 einer griechischen Version des DTK erreichten Gesamtpunktwerte als Hinweise auf das Ausmaß der jeweils aktuell bestehenden individuellen Betroffenheit durch eine depressive Störung genutzt werden, ergibt sich die mit an Sicherheit grenzende Wahrscheinlichkeit des Auftretens depressiver Störungen in der untersuchten Stichprobe.
2. Der Anteil der von einer depressiven Störung mit klinischer Ausprägung betroffenen Kinder an der untersuchten Gesamtstichprobe kann empirisch begründet geschätzt werden.
3. Nach dieser, auf der Grundlage der Daten aus Skala 1 des DTK vorgenommenen Schätzung ist der mutmaßlich bestehende Anteil der Kindern mit einer depressiven Störung an der Gesamtstichprobe von insgesamt 2016 Kindern als zwischen 350 und 515 liegend zu veranschlagen. Danach läge die mutmaßlich bestehende Quote im Bereich von 17,36% bis 25,55%.
4. Die mit einem verteilungsinternen Trennwert bestimmte Quote von Kindern mit relativ hohen Selbstzuschreibungen an depressiven Symptomen beträgt 14%.

7.4 Analyse der über die Skala 2 des DTK erhobenen Daten

Der DTK ist als ein Messverfahren mit drei Skalen entwickelt worden. Auf jeder dieser Skalen kann für jeden Teilnehmer an einer Untersuchung mit dem DTK ein Gesamtpunkt看wert errechnet werden. Jede Skala nennt mit jedem Item einen Verhaltensaspekt, der über eine Ja- oder auch eine Nein-Antwort als für das Verhalten der jeweils antwortenden Persönlichkeit kennzeichnend oder auch nicht zutreffend beurteilt wird. Durch das jeweilige Antwortverhalten wird ein individueller Gesamtwert an gewählten Items bekannt, der einem prozentualen Anteil am erreichbaren Höchstwert entspricht. Da die Skalen des DTK mit der Testkonstruktion unterschiedlichen Konstrukten zugeordnet worden sind, weist jeder Gesamtpunkt看wert für jede Skala - und damit für jedes der jeweils gemessenen Konstrukte - mittelbar nach, wie relevant die antwortende Persönlichkeit das über die Items repräsentierte Konstrukt aktuell eingeschätzt hat. Zwar wären auch weitere und nicht wirklich gezählte Mengen an weiteren konstrukt-spezifischen Verhaltensäußerungen zu denken, die bei Skala 2 ebenfalls der Kategorie des „agitierten Verhaltens“ zuzuordnen wären, aber die auf der Skala 2 platzierten 16 Items haben sich in den

Pilotuntersuchungen während der Entwicklung des DTK als eine messtechnisch hinreichende Menge an Items erwiesen. Mit größeren Itemmengen war kein Alphawert zu erreichen gewesen, der über dem mit der jetzigen Skala erreichten gelegen hätte. Hohe Gesamtpunktwerte auf der Skala 2 wären, den empirischen Beleg dafür vorausgesetzt, demnach als Hinweise darauf zu werten, dass die betreffende Persönlichkeit mutmaßlich dazu tendiert, mehr agitiertes Verhalten zu zeigen als Persönlichkeiten mit niedrigeren Gesamtpunktwerten. Daraus ist dann auch zu folgern, dass die Gesamtpunktwerte von Skala 2 implizit einen prognostischen Wert auf weitere Proben agitierten Verhaltens haben sollten. In einem durch entsprechende Retest-Reliabilität abgesicherten Zeitraum wäre zu erwarten, dass Persönlichkeiten mit einem hohen Skala2-Gesamtpunkt看wert (S2-GPW) unter anderen Versuchsbedingungen sehr wahrscheinlich auch ein vergleichbares oder höheres Aufkommen an Hinweisen auf agitiertes Verhalten zeigen würden. Den empirisch verlässlichen Nachweis für solche Konstruktvalidität vorausgesetzt, wäre ein hoher S2-GPW im DTK prognostisch valide im Hinblick auf die Erwartung, solche Persönlichkeiten zeigten auch in Situationen außerhalb der Durchführung eines DTK mehr agitiertes Verhalten als Vergleichspersonen mit niedrigeren Gesamtpunktwerten auf Skala 2 de DTK.

S2-GPW.

In diese Darstellung ist das Problem der fraglichen Stabilität einmal bekundeter Agitiertheit nicht einbezogen worden.

Die Gesamtpunktwerte aus Skala 2 sind aber auch noch in anderer Hinsicht aussagekräftig. Nach Rossmann (2005) tragen die Gesamtpunktwerte aus Skala 2 allein oder in Kombination mit den Gesamtpunktwerten aus Skala 1 zu einer ohne weitere Vorinformationen relativ treffsicheren Trennung von Kindern mit klinischer Diagnose einer depressiven Störung und Kontrollgruppenkindern bei.

Die Feststellung der mit der Untersuchung 2034GR über die Durchführung des DTK gewonnenen Aufschlüsse über die jeweils erreichten Gesamtpunktwerte nach Skala 2 im DTK folgt also einerseits der Notwendigkeit, die Vollständigkeit der Datenanalyse für alle Skalen des DTK und deren Vergleichbarkeit zu erreichen. Zum anderen wird die skalenweise Berechnung von Gesamtpunktwerten wegen deren unterschiedlicher Verwendung in Formeln zur Bestimmung von Trennwerten (cut-offs) und auch wegen der angestrebten Vergleiche zwischen untersuchungsinternen und externen Daten unersetzbar. Zu prüfen sind insbesondere die Beziehungen der Skalendaten untereinander und deren Relationen zu Vergleichsdaten aus anderen Untersuchungen, nicht zuletzt zu denen aus der deutschsprachigen Normstichprobe von Rossmann (1993).

Die nun darzustellenden Arbeitsschritte zur Analyse der individuellen und gruppenspezifischen Gesamtpunktwerte aus der Skala 2 in der verwendeten griechischen Version des DTK folgen der Analysetätigkeit bei der Erfassung der Werte aus Skala 1.

7.4.1 Verteilung der über die Skala 2 des DTK bekannt werdenden skalenspezifischen Gesamtpunktwerte (S2-GPW)

Die Zahl der Kinder in der gezogenen Stichprobe von n=2034 Schulkindern, die, weil ihre Antwortbögen lückenlos ausgefüllt waren, auch einen Gesamtpunkt看wert nach Skala 2 erreichen konnten, ist identisch mit der in die Analyse der Skala 1 einbezogenen Kinder. Kriterium für die Einbeziehung in die Analyse der DTK Werte war die lückenlose Bearbeitung des DTK insgesamt. So ist wiederum für n = 2016 Kinder ein Gesamtwert zu berechnen, diesmal der nach Skala 2 (S2-

GPW). Welche Gesamtpunktwerte auf Skala 2 des DTK von welcher Zahl von Teilnehmern erreicht worden sind, zeigt Tabelle 7.10.

Die Skala 2 „Agitiertes Verhalten“ umfasst 16 Items. Somit konnten die individuellen Werte von 0 bis 16 variieren. Die Tabelle 7.10 führt die Anzahl an Teilnehmern (Frequenzen) auf, die den einzelnen Gesamtpunktwerten zugefallen sind und nennt den prozentualen Anteil der jeweiligen Frequenz an der Gesamtzahl der Mitglieder der Stichprobe, die den DTK vollständig ausgefüllt hatten, was auf n = 2016 Schulkinder zutrif.

Tabelle 7.10: Verteilung der Skala2-Gesamtpunktwerte (S2-GPW)

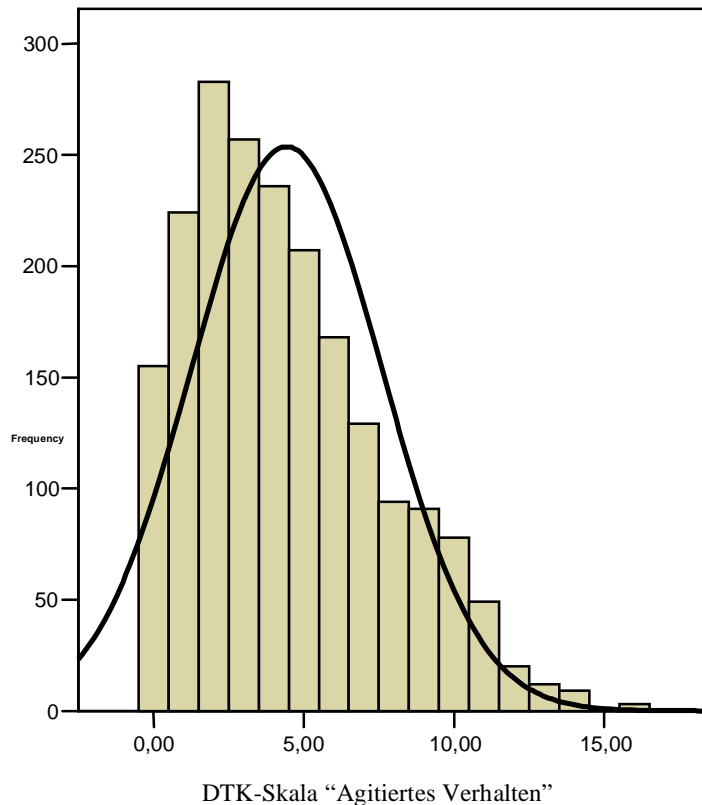
S2-GPW	Frequenz	Prozent	S2-GPW	Frequenz	Prozent
0	155	7,7	9	91	4,5
1	224	11,1	10	78	3,9
2	283	14,0	11	49	2,4
3	257	12,7	12	20	1,0
4	236	11,7	13	12	,6
5	207	10,3	14	9	,4
6	168	8,3	15	1	,0
7	129	6,4	16	3	,1
8	94	4,7	Summe	2016	100,0

Zur Häufigkeitsverteilung der S2-Gesamtpunktwerte gehören die nachstehend aufgeführten Kennwerte.

Mean	4,4276	Sum	8926,00	Skewness	,717
Std. Error of Mean	,07060	Median	4,0000	Std. Error of Skewness	,055
Range	16,00	Mode	2,00	Kurtosis	-,036
Minimum	,00	Std. Deviation	3,17007	Std. Error of Kurtosis	,109
Maximum	16,00	Variance	10,049		

Die von den Teilnehmern erreichten S2-GPW variieren über die gesamte Breite der erreichbaren Gesamtpunktwerte. Den Zielen der Testkonstruktion folgend, nimmt die Besetzung der einzelnen Punktwerte oberhalb der Skalenmitte deutlich ab. Die Verteilung der Teilnehmer auf die erreichbaren Punktwerte ist sehr deutlich asymmetrisch. Den Wertebereich von 0 bis zur Skalenmitte 7,5 besetzen 82,2% der Teilnehmer. Dementsprechend verbleibt für die deutliche Minderheit von nur 17,8% der Teilnehmer die andere Hälfte der Skala, die von 8,5 bis 16 Gesamtpunkten. Insoweit gleicht die linkssteile Verteilung der Werte auf Skala 2 der schon zuvor bei Skala 1 beobachteten (1.AS). Die siebenfache oder auch noch höherfrequente Wahl von Items der Skala 2 sind seltene Ereignisse. Der Schwierigkeitsgrad der n-fachen Itemwahl liegt ab und mit dem Gesamtpunktwert 7 (6,4%) unter dem des Gesamtpunktwertes 0 (7,7%)(2.AS).

Das nachstehende Histogramm (Abbildung 7.2) gibt die graphische Veranschaulichung der für Skala 2 des DTK in der griechischen Untersuchung festgestellten Häufigkeitsverteilung.

Abbildung 7.2: Häufigkeitsverteilung der Gesamtpunktwerte auf Skala 2 des DTK (n = 2016)

Den Intentionen der Testkonstruktion entsprechend sollte die für die Verteilung der Gesamtpunktwerte nach Skala 2 gefundene Verteilungsform, ebenso wie bei Skala 1, von einer Normalverteilung signifikant unterschieden sein. Die dem entsprechende Hypothese (H_1), nach der die Verteilung der Gesamtpunktwerte aus Skala 2 nicht als aus einer normalverteilten Grundgesamtheit kommend angesehen werden darf, sollte sich bestätigen lassen. Nach den Ergebnissen aus dem durchgeführten Kolmogorov-Smirnov Test ist die H_1 zu verifizieren und die Annahme einer Normalverteilung zurückzuweisen (Kolmogorov-Smirnov $Z = 5,820$; $p = 0,000$). Die ansteigenden Werte des Modus, Medianes und arithmetischen Mittels verweisen auf eine linkssteile Verteilung.

Somit erweist sich auch Skala 2 als theoriekonform messend. Eine über die 6-fache Wahl von Items der Skala 2 hinausgehenden Mengen an Selbstzuweisungen von Items durch die teilnehmenden Schulkinder wird zum relativ seltenen Ereignis.

Die aus der Verteilung der S2-GPW abzuleitenden Informationen über das Antwortverhalten der beteiligten Schulkinder in der Bearbeitung der Skala 2 „Agitiertes Verhalten“ des DTK werden die nachstehenden Aussagen gebildet.

7.14 AS: Die über die Antworten zu den Items der Skala 2 von den beteiligten Schulkindern abgegebenen Selbstzuweisungen an Symptomen verteilen mit ihren individuellen Gesamtsummen an Zuweisungen über die volle Breite der Skala mit den von 0 bis 16 Zuweisungen reichenden

Wahlmöglichkeiten . Der mit der Skala angebotene numerische Entscheidungsraum wird voll ausgenutzt. Die vorgefundene Verteilung spricht mit den geringen Besetzungen der Werte 15 und 16 für eine hinreichende Zahl der angebotenen Items.

7.15 AS: Die S2-GPW des DTK erweisen sich in ihrem Auftreten, wie schon zuvor die Gesamtpunktwerte der Skala 1 im DTK, als nicht normalverteilt. Ihre spezielle linkssteile Verteilungsform mit einem signifikant unter dem Verlauf der Normalverteilung abgeflachten Bereich hoher Punktwerte entspricht den gesetzten Vorgaben aus der Testkonstruktion. Indem die griechische Version der Verteilungsform mit der deutschsprachigen Normierung des DTK übereinstimmt, erweist sich, dass auch nach der Übersetzung der Items ins Griechische das dem Test unterliegende Messprinzip des DTK in der Skala 2, soweit erkennbar funktional ungemindert erhalten geblieben ist.

7.16 AS: Die mit dem DTK befragten Schüler und Schülerinnen erreichen in der griechischen Version individuelle S2-GPW, die von Angaben über die eigene Nichtbetroffenheit von agitiertem Verhaltensproblemen bis hin zum maximal erreichbaren Höchstwert an Betroffenheit reichen. Sie gleichen in diesem Antwortverhalten den deutschsprachigen Kindern aus der Normierungsstichprobe von Rossmann (in Rossmann, 1993). Diese Aussage fasst den Sachverhalt der 2.AS aus einer persönlichkeitspsychologischen Perspektive. Das Antwortverhalten der griechischen Kinder erlaubt die Beibehaltung einer Annahme, nach der die bei den Kindern der griechischen Stichprobe durch den Prozess der Bearbeitung des DTK angesprochenen psychischen Prozesse denen bei deutschsprachigen Kindern zumindest ähnlich sind.

7.17 AS: Auch in der griechischen Version des DTK streuen die S2-GPW so, dass eine Majorität von 82,2% aller Mitglieder der Stichprobe auf den ersten 50% der erreichbaren Gesamtpunktwerte platziert ist. So wird die zweite Hälfte der Skala der Gesamtpunktwerte, einer Minderheit von 17,8% der Mitglieder dieser Stichprobe überlassen.

7.18 AS: In der griechischen Übersetzung erweist sich über die von der Normalverteilung abweichende Verteilungsform der individuellen Gesamtpunktwerte auf Skala 2, dass die Konstruktionsprinzipien der deutschen Testform ungemindert erhalten sind und sich, nachgewiesen durch die Messdaten, auch in der griechischen Anwendung als sensibel gegenüber dem Auftreten der relativ seltenen Ereignisse einer individuellen Bestätigung von subjektiv gehäuft auftretenden Skala 2 –Symptomen erweisen. Das ist in der Begründung einer durch die allgemeinen Untersuchungsbedingungen für die griechische Untersuchung 2034GR vorerst notwendig werdenden Orientierung an den Ergebnissen der Normierung des deutschsprachigen DTK ein die Vorgehensweise stützendes Argument.

7.19 AS: Die Items der Skala 2 des DTK sind dem Konstrukt „Agitiertes Verhalten“ depressionsspezifisch zugeordnet. Die Verteilung der S2-GPW weist insoweit für die Kinder der untersuchten Stichprobe inhaltssvalide nach, dass 57,2% von ihnen die in ihrem Erleben auftretenden unterschiedlichen Symptome auf Zahlen von 0 bis 4 einschließlich beschränken. Damit beeinflussen sie insoweit eine für diese Untersuchung geltende griechische Norm für die Wahl agierter Verhaltensweisen aus der Itemliste der Skala 2 des DTK. Die absolute Mehrheit der Kinder dieser Stichprobe sieht den Aspekt des agitierten Verhaltens als mit maximal 4 Items als hinreichend repräsentiert an. Sie stützen damit eine Aussage wie die nachfolgende: „Für die absolute Mehrheit der untersuchten griechischen Schulkinder ist der in ihrem Verhalten für sie selbst erkennbare Aspekt des agitierten Verhaltens durch maximal 25% der durch Skala 2 (DTK) zur Wahl angebotenen Verhaltensaspekte hinreichend repräsentiert. Die Tendenz dieser Aussage wird durch das Antwortverhalten von ungefähr einem weiteren Drittel der an der Untersuchung beteiligten Schulkinder gestützt. Sie wählen zwar jeweils etwas höhere Mengen an Items, modifizieren die Normsetzung aber nur insoweit, als unter Kenntnisnahme ihres Beitrags eine präzisierende neue Aussage zur Norm für agitiertes Verhalten in dieser griechischen Untersuchung gegeben werden kann. Da, wie vorab bereits erklärt, weitere 34,2% der

untersuchten Kinder jeweils 5 bis 9 Items aus der Liste als zutreffend erklären, gilt die folgende Aussage: „In der untersuchten Stichprobe ist die Wahl von 10 oder mehr Items agitierten Verhaltens ein überaus seltenes Ereignis, das mit jedem individuellen Auftreten eine in mehrfacher Hinsicht zur Beachtung nötige Information gibt.“ Etwa 8,4% der untersuchten Kinder geben Werte von 10 bis 16 an. Daraus ergibt sich der Sachverhalt, dass 91,6% der teilnehmenden Schulkinder mit etwa 60% des Skalenbereiches auskommen und die restlichen 8,4% der Kinder sich mit ihren Gesamtpunktwerten der Skala 2 über 40% der Skalenbreite ausdehnen.

7.20 AS: Die 7.19 AS wertet die Besonderheiten der festgestellten Verteilung zur Skala 2 (DTK) in der Perspektive der Aussagekraft von Messwerten. In einer anderen Perspektive gibt diese Verteilung dann hinreichend spezifische Hinweise auf die Position des zu vermutenden Trennwertes zwischen im Normalbereich belasteten Kindern einerseits und in klinischer Ausprägung betroffenen andererseits. Nach den in der 7.19 AS mitgeteilten Selbstzuweisungen der Schulkinder dieser Stichprobe, erreicht diese Skala nicht das Niveau von Intervallskalen. Für die Abstände zwischen den jeweiligen Gesamtpunktwerten ist der Verdacht zu äußern, dass sie im unteren Skalenbereich anders zu quantifizieren sind als im oberen Teil. Dann aber wären wegen der festgestellten Verteilung die zunächst auch als überzogen wahrnehmbaren nachstehenden Aussagen zu rechtfertigen. Als überzogen könnten sie eingeschätzt werden, weil mit ihnen 9/10 der Teilnehmer einer Stichprobe qualitativ anders als die Teilnehmer in dem verbleibenden 1/10 der Stichprobengröße eingeschätzt werden. Das erfolgt aber mit der nachstehenden Aussage: „Etwa 92% der in 2034GR untersuchten griechischen Kinder geben Hinweise auf eine in Normalbereichen liegende Betroffenheit durch Symptome agitierten Verhaltens, während gut 8% Indizien für klinisch beachtenswerte Ausprägungen agitierten Verhaltens bieten.“ Der Anschein einer durch die Zuschreibung von Normalität an gut 90% einer Verteilung wahrzunehmende Widersprüchlichkeit löst sich bei genauerer Betrachtung der zu Grunde liegenden Daten weitgehend auf.

7.21 AS: Indem nach den Selbstangaben der untersuchten 2016 griechischen Schulkinder nur etwa 8,4% von ihnen durch höhere Zahlen depressionsspezifischer Symptome agitierten Verhaltens als davon belastet bekannt werden, verdeutlicht sich damit das störungsspezifische Gewicht größerer Mengen an selbstzugewiesenen Symptomen agitierten Verhaltens. Die Feststellung einer störungsspezifischen Bedeutung informiert aber noch nicht darüber, für welche Störung das Auftreten agitierten Verhaltens als spezifisch angesehen werden sollte. Agitiertes Verhalten wird nicht nur in Störungsbildern affektiver Störungen explizit oder implizit angeführt. Es gehört zumindest ebenso oder noch spezifischer zum Störungsbild von Aufmerksamkeitsstörungen. Die für die Skala 2 (DTK) anfallenden Gesamtpunktwerte verweisen nicht ausschließlich auf vermutete depressive Störungen, obschon der Testname zu diesem Fehlschluss verleiten könnte. So ist das Aufkommen an Gesamtpunktwerten nach Skala 2 (DTK), zumindest theoretisch, kein explizit depressionsbezogenes Ereignis. Den verschiedenen Störungsbildern in den Manualen DSM und ICD folgend, könnten Gesamtpunktwerte aus Skala 2 (DTK) ihren Hinweisscharakter für depressive Störungen erst durch eine Kombination mit anderen Skalenwerten gewinnen. Diese Erwartung entspricht auch den theoretischen Grundlagen der Konstruktion des DTK. Ob diese Erwartungen sich auch bestätigen lassen, wird zu prüfen sein.

7.22 AS: Die aufgeführten acht Aussagen zur Verteilung der Gesamtpunktwerte auf der Skala 2 des DTK entsprechen den Daten. Als von den teilnehmenden Schulkindern selbst ausgewählte Aussagen haben sie, darauf bezogen, uneingeschränkte Validität. Die angegebenen Zahlen von Schulkindern haben die über den S2-GPW bekannt werdenden Mengen an Ja-Entscheidungen vorgenommen. Somit ist dann zu wiederholen, was schon für Skala 1 in Anspruch genommen worden war. Unter Berufung auf Zimmermann (siehe Rossmann, 2005) wird empirisch begründet für die Gesamtpunktwerte aus Skala 2 (wie schon zuvor für die aus Skala 1) eine

skalenspezifische Konstruktvalidität angenommen. Danach sind die S2-GPW der Untersuchung 2034GR valide Indikatoren für das variierende Ausmaß von störungsspezifischen Ausprägungen agitierten Verhaltens.

7.23 AS: Nach der 7.22 Aussage wird mit den S2-GPW für die Teilnehmer an der Untersuchung 2034GR die jeweils individuelle Ausprägung der störungsspezifisch gerichteten Symptomatik „Agitierten Verhalten“ für jedes an der Untersuchung teilnehmende Schulkind bekannt. Ungeachtet der Beweislage für die anzunehmende oder abzulehnende Annahme von Validität sollte die hohe Selbstzuschreibung der Symptome als Anlass zu individuell zu richtenden Nachforschungen und psychologischen Beratungen gewertet werden.

7.4.2 Informationen aus den Gesamtpunktwerten der DTK-Skala 2 als Hinweise auf depressive Störungen

Wie zuvor bei der Analyse der Gesamtpunktwerte aus Skala 1 des DTK sind nun die Gesamtpunktwerte aus der Bearbeitung der Skala 2 des DTK durch n=2016 griechische Schulkinder nach Hinweisen zum Auftreten depressiver Störungen abzusuchen.

In diesem Zusammenhang war hier vorab bereits darauf hinzuweisen, dass von Daten zur Skala 2 „Agitiertes Verhalten“ keineswegs nur Hinweise auf depressive Störungen erwartet werden. Agitiertes Verhalten tritt in unterschiedlichen Störungsbildern auf, so etwa in den Aufmerksamkeitsstörungen. Die dem DTK unterliegende Theorie zielt darauf ab, durch Messwerte auf drei unterschiedlichen Skalen die Voraussetzungen zu schaffen, um durch einen kombinierten Gesamtwert aus drei unterschiedlichen Skalen die Erkennensleistungen des Gesamttests erhöhen zu können. Eine darauf zu beziehende simplifizierende aber eben auch veranschaulichende Verallgemeinerung wäre die Aussage, dass mit einem maximalen Punktwert auf der DTK-Skala „Dysphorie/Selbstwert“, zu dem ein maximaler Punktwert aus der DTK-Skala 2 „Agitiertes Verhalten“ gehörte und zu denen dann ein weiterer maximaler Punktwert, der von Skala 3 „Müdigkeit/autonome Reaktionen“ hinzu käme, ein kombinierter Gesamtwert aus drei Skalen gebildet werden könnte, für den erwartet werden dürfte, dass er mit seinem Auftreten jeweils eine Persönlichkeit bekannt werden ließe, die sich als durch eine erheblich ausgeprägte depressive Störung betroffen erweisen sollte.

Eine solche Erwartung ließe sich theoriekonform konstruieren, empirische Nachforschungen hätten zu untersuchen, ob sich dann die an der Durchführung des Verfahrens teilnehmenden Persönlichkeiten in ihrem Antwortverhalten auch hinreichend theoriekonform verhalten würden. Bekannt und in den vorangegangenen Ausführungen ausführlich erörtert worden ist der dem widersprechende und nun wiederum theoretisch zu fassende, empirisch geprüfte Sachverhalt. Dementsprechend ist auch für die mit Skala 2 des DTK in der Untersuchung 2034GR erarbeiteten Daten die Frage zu stellen und zu beantworten, welche Hinweise auf depressive Störungen in der untersuchten Stichprobe durch sie gegeben werden.

Wenn selbst durch drei depressionsspezifische Skalen in einem Depressionstest aus mehreren Gründen eine Diagnosen ermöglichende Treffsicherheit nicht erreicht werden kann, eine Steigerung der Validität eines solchen Verfahrens gegenüber eindimensional messenden Instrumenten sollte erreichbar sein.

Über diese oder ähnliche Annahmen ist bei dem erreichten Analysestand für die Untersuchung 2034GR aber nicht zu entscheiden, denn Analyseobjekt sind nun ausschließlich die Daten aus der Durchführung einer Skala, der Skala 2 des DTK.

7.4.2.1 Das Phänomen der hohen Selbstzuschreibung von Symptomen aus der Itemliste der Skala 2 des DTK

Die DTK-Skala 2 bietet 16 Items mit unterschiedlichen Beispielen für agitiertes Verhalten zur Beantwortung an. Über eine Ja-Nein-Entscheidung sollten auch die griechischen Schulkinder der Untersuchung 2034GR mitteilen, ob das jeweils genannte Verhaltensbeispiel in ihrem Selbsterleben auftritt. Aus der Gesamtstichprobe von $n=2034$ Schulkindern haben $n=2016$ Kinder den gesamten DTK bearbeitet, was als Voraussetzung für die Einbeziehung in die Analyse der DTK-Daten eingeführt worden war. Somit ist im Folgenden zu betrachten, welche Gesamtpunktwerte die in die Analyse einbezogenen $n=2016$ griechischen Schulkinder erreichten. Aus der Beschreibung und der Analyse der Verteilung der S2-GPW ist bekannt geworden, dass über 90% dieser Kinder für sich selbst jeweils nur Mengen aus den unteren 60% der Skala an Gesamtpunktwerten erreichten. Die Bildung von Untergruppen an untersuchten Schulkindern, die diesen Gruppen entsprechend den von ihnen erreichten S2-GPW zugewiesen werden, verschafft einen Überblick über die durch unterschiedliche Frequenzen an Gesamtpunktwerten entstandene Stufung.

Entsprechend der in 9.3.3.1 beschriebenen Verfahrenspraxis wird die Verteilung der Gesamtpunktwerte der DTK-Skala 2 an einem zuvor definierten Skalenwert getrennt. Definierter Trennwert ist der GPW, dessen Auftretensfrequenz in der Verteilung der Gesamtpunktwerte erstmals unter der Auftretensfrequenz für den $GPW=0$ liegt. Dieser Punkt ist nach Tabelle 7.10 zu identifizieren. Die Auftretensfrequenz von $GPW=0$ wird erstmals in der Verteilung durch den $GPW=7$ unterschritten. Nach der an dieser Stelle zu vollziehenden Trennung sind Kinder mit S2-GPW von 7 bis 16 bekannt geworden, die definitionsgemäß als Gruppe der durch eine hohe Zahl von Selbstzuweisungen an Symptomen agitierten Verhaltens als in kritischer Weise belastet gelten. Diese Teilgruppe wird nachfolgend als Gruppe AGV3 bezeichnet. Die vom Trennpunkt S2-GPW=7 in Richtung auf $GPW=0$ auf der Skala positionierten Schulkinder werden, wiederum nach der vorab gesetzten Verfahrensweise, in zwei weitere Teilgruppen getrennt. Trennpunkt dafür ist der in der Mitte des genannten Skalenteils befindliche Gesamtpunkt看wert. Da 6 Gesamtpunktwerte auf zwei Gruppen aufgeteilt werden sollen, ist die Frage zu beantworten, ob die Punktwertmitte von $GPW=3,5$ als Trennwert benutzt werden soll. Dies soll vermieden werden, weil dann eine Aufteilung von Kindergruppen vollzogen werden müsste, die jeweils durch ihre Selbstzuweisungen einer in dieser Perspektive übereinstimmenden Teilmenge von Kindern angehörten. Falls eine solche Gruppe bei bestehendem Ordinalskalenniveau mit starken Frequenzunterschieden auf den Skalenteilen jeweils zur Hälfte auf zwei unterschiedliche Gruppen AGV1 und AGV2 verteilt werden sollte, begünstigte das eine mögliche und nicht erwünschte Einebnung von Gruppenunterschieden. Aus diesem Grund wird eine zweite Verfahrensregel eingeführt. Die Teilung der einem Gesamtpunkt看wert zugewiesenen Zahl von Kindern soll vermieden werden. Der fragliche Gesamtpunkt看wert wird jeweils der neu zu bildenden Teilgruppe zugewiesen, die in Richtung auf $GPW=0$ liegt. Danach ist eine erste Gruppe AGV1 mit den Kindern der Gesamtpunktwerten 0, 1, 2 und 3 zu bilden. Die Gruppe AGV2 umschließt die Kinder mit den Gesamtpunktwerten 4, 5 und 6. Die nachstehende Tabelle zeigt die Frequenzen an Schulkindern für die in dieser Weise gebildeten Teilgruppen AGV1, AGV2 und AGV3.

Tabelle 7.11: Besetzung der nach den Gesamtpunktwerten aus DTK-Skala 2 „Agitiertes Verhalten“ gebildeten Teilgruppen AGV1, AGV2 und AGV3

Bezeichnung	Skala2-Gesamtpunktwerte	Selbstzuweisungen an Symptomen aus der Itemliste DTK-Skala 2	Frequenz	Prozentsatz von n
AGV 1	0-3	Niedrige Frequenz an Selbstzuweisungen von 0% bis 23,5 % der wählbaren Symptome	919	45,5 %
AGV 2	4-6	Mittlere Frequenz, bis zu 41,2% der wählbaren Symptome	611	30,3 %
AGV3	7-16	Hohe Frequenz an Selbstzuweisungen von 41,2% bis 100% wählbarer Symptome	486	24 %
Gesamte Stichprobe	0 – 16	Alle S2-GPW	2016	100 %

Aus Tabelle 7.11 ist zu entnehmen, dass 486 Schulkinder aus der untersuchten Stichprobe sich selbst eine jeweils relativ hohe Zahl von Symptomen agitierten Verhaltens als für ihr eigenes Verhalten kennzeichnend zugewiesen haben. Diese Kinder repräsentiert 24% der Angehörigen dieser Stichprobe griechischer Schulkinder.

Als Indiz für das Bestehen eines störungsspezifischen Risikos sind die hohen Zuschreibungen an Symptomen agitierten Verhaltens anzusehen, weil danach das Verhalten der Kinder aus der Gruppe AGV3 sehr viel dichter von agitiertem Verhalten durchsetzt erscheint als das der Kinder aus AGV1 und AGV2. Das ist durch die Berechnung der durchschnittlich pro Gruppenmitglied genannten Symptome agitierten Verhaltens zu verdeutlichen.

Tabelle 7.12: Durchschnittlich gewählte Items agitierten Verhaltens pro Kind in drei Gruppen

Gruppe	Summe Items	Kinder	Items pro Kind
AGV1	1561	919	1,6985
AGV2	2987	611	4,887
AGV3	4378	486	9,9982
Stichprobe	8926	2016	4,4275

Die Tabelle 7.12 nennt die durchschnittlich pro Kind in den drei Gruppen AGV1, AGV2 und AGV3 gewählten Items agitierten Verhaltens. Die Kinder der Gruppe AGV3 haben durchschnittlich über 5 mal größere Mengen an Symptomen agitierten Verhaltens als die Kinder in AGV1 gewählt. Eine Denkmöglichkeit wäre, das Zustandekommen solcher Unterschiede als zufällig zu verstehen. Unmöglich wäre das Aufkommen eines solchen Phänomens nicht, aber immerhin sehr unwahrscheinlich. Etwa ein Viertel der untersuchten Schulkinder fühlt sich veranlasst, ein Vielfaches an Symptomen dessen in Anspruch zu nehmen, was $\frac{3}{4}$ der anderen Kinder als hinreichend erscheint. Eine Annahme mit höherem Erklärungswert entsteht, wenn den mit der Itemliste der DTK-Skala 2 präsentierten Verhaltensbeispielen ein persönlicher Wert zugeordnet wird, der aus dem Wiedererkennen von selbst erlebten Verhaltenssituationen entsteht. Die Kinder der Gruppe AGV3 verbinden mit den vorgegebenen Beispielen bedeutsame eigene Erfahrungen. Warum die bedeutungsvoll sind, ist für die Analyse der Daten über das Antwortverhalten zur DTK-Skala 2 nicht nachweisbar. Unterstellt wird aber, dass die Kinder der Gruppe AGV3 ebenso wenig zufällig in diese Gruppe gekommen sind, wie die anderen Kinder in die ihrige. Das spezielle Antwortverhalten wird als bedeutsam aufgefasst. Auf welche Bedeutung es dann zurückgeht, bliebe aufzuklären. Die über die Analyse der S2-GPW ermittelte Information

wird nachfolgend in einer 11. Aussage innerhalb der Ausführungen zur Analysearbeit an Daten über das Antwortverhalten zur DTK-Skala 2 zusammengefasst.

7.24 AS: Die Bestimmung eines verteilungsinternen Trennwertes für die Verteilung der Gesamtpunktwerte aus Skala 2 des DTK führt zur Erfassung einer 24% der Gesamtstichprobe einschließenden Teilstichprobe von Kindern mit relativ hohen Mengen an Selbstzuschreibung von Symptomen aus der agitiertes Verhalten aufführenden Itemliste der Skala 2.

7.25 AS: Der Hinweis auf die risikofolle psychische Belastung der über hohe Selbstzuschreibungen von Symptomen agitierten Verhaltens nach Skala 2 des DTK identifizierten Teilgruppe wird durch den Nachweis eines im Durchschnitt gegenüber allen anderen Kindern der Stichprobe pro Kind mehr als verdoppeltes Aufkommen an Symptomen agitierten Verhaltens gegeben.

Diese Feststellungen beruhen auf dem durch die entsprechenden Daten dokumentierten Wahlverhalten der Schulkinder aus der Stichprobe 2034GR. Ungeprüft ist, inwieweit dieses Wahlverhalten der Kinder mit den alltäglichen Erfahrungen, auf die sie bezogen werden müssen, in Übereinstimmung steht. Der Erkenntnisgewinn, den die Feststellungen bieten, dokumentiert sich demnach vor allem damit, dass sie die Möglichkeit der Überprüfung eröffnen. So wie zuvor in der Analysearbeit mit den Daten aus der Skala 1 ist nun wiederum die Frage zu stellen und zu beantworten, welche Hinweise die über die Skala 2 bekannt gewordenen Daten aus der Durchführung des DTK in der Untersuchung 2034GR auf das Bestehen von depressiven Störungen in dieser Stichprobe geben.

7.4.2.2 Hinweise der GPW aus Skala 2 des DTK auf depressive Störungen in der untersuchten Stichprobe

Die Annahme einer eigenen Konstruktvalidität der DTK-Skala 2 ist explizit oder implizit eng mit der Konstruktion des DTK verbunden. Sie nimmt damit eine spezielle Realitätsbeschreibung als für die Messungen mit dem DTK nutzbar und nützlich in Anspruch. Die drei Skalen des DTK sollen über die Messung von unterschiedlichen Konstrukten eine Datenbasis schaffen, die den bestehenden psychischen Zusammenhängen im Konzept der emotionalen Störungen nach DSM und ICD so weit entspricht, dass über die Anwendung des DTK beachtbar treffsichere Prognosen auf das Bestehen solcher Störungen ermöglicht werden.

Wenn die Skala 2 dazu in spezifischer Weise beitragen soll, liegt nahe, von ihr zu erwarten, dass sie eine dafür zureichende Konstruktvalidität erreichen wird.

Diese Erwartung darf nach Zimmermann (siehe Rossmann, 1993, 2005) begründet als erfüllt angenommen werden. Der fand in einer Untersuchung, dass Kinder, die auf Skala 2 hohe Werte erreichen, zwei unterschiedlichen Diagnosegruppen entstammen, zum einen der Gruppe mit depressiven und zum anderen der Gruppe mit aufmerksamkeitsgestörten Kindern. Das Argument, beides sei widerspruchsfrei durch die Annahme einer Konstruktvalidität der DTK-Skala 2 „Agitiertes Verhalten“ zu erklären, ist auch das Basisargument der innerhalb dieser Untersuchung 2034GR auf die Daten aus Skala 2 gerichteten Analysearbeit. Erwartet wird, dass die Skala 2 eine spezifische Konstruktvalidität erreicht.

Insoweit, als der DTK konstruktvalide Skalen hätte, realisierte sich die dem Verfahren unterliegende Testtheorie.

Danach misst der DTK auf drei konstrukvaliden Skalen. Skala 1 erfasst das Konstrukt „Dysphorie/Selbstwert“. Es dürfte als eine spezifisch wirkende regulative Instanz interpretiert

werden. Das von dieser Instanz initiierte Verhalten wäre auch außerhalb der Testsituation in jedem relevanten Verhaltenskontext beobachtbar. Die DTK-Skala 2 misst konstruktvalide das „Agitiertes Verhalten“ genannte Konzept. Die über die Gesamtpunktwerte dieser Skala bekannten Kinder und Jugendlichen würden sich über die von ihnen gezeigte Ausprägung vermehrten agitierten Verhaltens von anderen unterscheiden. Die Skala 3 schließlich sollte konstruktvalide die persönlichen Ausprägungen von „Müdigkeit und Autonomen Reaktionen“ erfassen. Dementsprechend sollten Kinder, die auf dieser Skala hohe Gesamtpunktwerte erreichten, sich gegenüber anderen auch durch konstrukt-konformes Verhalten absetzen. Ein Mehr an Müdigkeit und autonomen Reaktionen gegenüber anderen Personen wäre bei ihnen zu erwarten. Wenn die einzelnen Skalen ihrer Bestimmung entsprechend reliabel und valide messen, erfassen sie alle Kinder mit hohen Werten für „Dysphorie/Selbstwert“ (1.), für „Agitiertes Verhalten“ (2.) und für das Konzept „Müdigkeit/Autonome Reaktionen“ (3.). Die dann mit den drei Skalen gewonnenen Informationen würden Eigenheiten von Kindern erkennen lassen, die mit diesen Ausprägungen in den Störungsbildern unterschiedlicher psychischer Störungen auftreten. Die betreffenden Kinder könnten, sobald nur die von einer der drei DTK-Skalen erfassten Ausprägung eines Konstruktes zu beachten wären, als von mehreren Störungen betroffen gelten. Ein hoher Wert auf der Skala „agitiertes Verhalten“ rechtfertigte sowohl den Verdacht auf das Bestehen einer Aufmerksamkeitsstörung als auch den auf das Bestehen einer depressiven Störung. Der Messerfolg des DTK im Erkennen depressiver Belastungen ist danach dann also davon abhängig, ob die zu verschiedenen anderen Störungsbildern gehörenden Kinder durch ein summatives Ausschlussverfahren von den depressiven Kindern getrennt werden können. Der DTK misst depressionsspezifisch valide, wenn sich die folgende Bedingung erfüllt: „Depressiv gestört ist, wer sowohl hohe Werte auf Skala 1 als auch auf Skala 2 und auf Skala 3 erreicht.“ Falls das nicht zum Ausschluss von depressiv unbelasteten, weil zu anderen Störungsbildern gehörenden Kindern führen sollte, wäre der DTK in Gefahr, auch Kinder als depressiv zu bezeichnen, die aus anderen als depressionsspezifischen Zusammenhängen hohe Werte auf einer oder gar auf zwei der konstruktvaliden Skalen erreichten. Da depressive Störungen bei Kindern beachtliche Komorbiditätsraten aufweisen, müssten bei der Identifizierung der einzelnen Störungsbilder auch messtechnische Probleme in Betracht gezogen werden.

Die für die DTK-Skala 2 nach Zimmermann angenommene spezielle Konstruktvalidität legt nahe, hohe Gesamtpunktwerte auf dieser Skala als validen Beleg für eine in der jeweiligen Persönlichkeit wirkende regulatorische Instanz für agitiertes Verhalten zu werten. Indem hier nun dieser Argumentation gefolgt wird, sind die individuell erreichten Gesamtpunktwerte als Indikatoren für jeweils eine entsprechende Ausprägung agitierten Verhaltens in dem jeweiligen Schulkind zu sehen.

Dann bemessen die unter Kenntnisnahme der Verteilung der Gesamtpunktwerte aus DTK-Skala 2 und unter Anwendung des definierten verteilungsinternen Trennpunktes festzustellenden Quoten für Gruppen mit unterschiedlichen Frequenzen an Selbstzuweisungen von Items der Skala 2 individuelle Ausprägungen agitierten Verhaltens. Annehmen ließe sich, hohe GPW auf Skala 2 seien zwar ein in der Vorhersage auf agitiertes Verhalten prognostisch zu verwertendes Indiz, würden sich aber noch nicht zur Vorhersage depressiver Störungen eignen. Erst ein über alle drei Skalen hinweg sich aufsummierender Gesamtpunktwert hätte die erwartete prognostische Effizienz. Unter Kenntnisnahme der hohen Komorbiditätsraten psychischer Störungen bei Kindern und Jugendlichen scheinen solche oder ähnliche Argumentationen gegenstandslos. Wegen der nachweisbaren hohen Komorbidität der Störungsbilder könnten alle drei Skalen des DTK jeweils eigene differentielle Validitäten erreichen. Mit den Daten der Untersuchung 2034GR werden sich weder die einen noch die anderen Hypothesen dazu prüfen lassen. Bestimmt werden kann aber, welche Informationen sich ergeben, wenn die von Pötz (s.o.) auch für die Skala 2

berechneten und auf ihre differentielle Validität geprüften Trennwerte für die Unterscheidung von Kindern ohne und mit diagnostizierter depressiver Störung auf die Daten aus der griechischen Untersuchung 2034GR angewendet werden.

Der S2-GPW ist nach Pötz (s. o.) ein Messwert, der Hinweise auf das Bestehen einer depressiven Störung geben kann. Diese Hinweise sind wiederum, wie schon bei der Analyse zur Skala 1, nicht absolut treffsicher, sie haben einen relativen Aussagewert. Die Treffsicherheit der Zuweisung eines Gesamtpunktwertes nach DTK-Skala 2 ist in seinen diskriminanten Leistungen über die Werte für Sensitivität und Spezifität zu kennzeichnen.

Um die Trennung der gesamten Stichprobe in zwei Teilstichproben vornehmen zu können, muss ein Trennwert (cut-off) bestimmt werden. Ein solcher empirisch bekannt gewordener Trennwert wird von Pötz (s.o.) auch für DTK-Skala 2 genannt. Dieser Trennwert wird für die nachfolgend zu vollziehende Trennung der gesamten Stichprobe in zwei Teilgruppen, die Teilgruppe der vermutlich mit klinischer Ausprägung depressiv belasteten Kinder und die Teilgruppe der nicht belasteten Kinder, übernommen. Pötz (s.o.) benennt den Trennpunkt für Skala2 mit >9 , bei einer Sensitivität von 72,73 und der Spezifität 63,64.

Tabelle 7.13: Verteilung der nach Trennwert >9 (Pötz, 2003) gebildeten Teilgruppen von Kindern mit und ohne Hinweise auf mutmaßlich bestehende depressive Störungen klinischer Ausprägung

Gruppe	N	%
Gruppe O	1844	91,4
Gruppe M	172	8,4
Gesamte Stichprobe	2016	99,8

Zeichenerklärung: Gruppe O = Gruppe von Kindern, mutmaßlich ohne klinischer Ausprägung einer depressiven Störung; Gruppe M = Gruppe von Kindern, mutmaßlich mit klinischer Ausprägung einer depressiven Störung; 2016 Kinder aus einer Stichprobe von $n = 2034$

Aus der Tabelle ist ersichtlich, dass unter Anwendung des durch Pötz (s.o.) für Skala 2 des DTK berechneten Trennwertes $GPW > 9$ der Hinweis gegeben wird, eine Teilgruppe von 8,4% der Kinder in der untersuchten Stichprobe sei mutmaßlich von depressiven Störungen betroffen, die klinisch bedeutsame Ausprägungen angenommen hätten. Die festgestellte Quote liegt deutlich unter der nach den Daten aus Skala 1 vollzogenen Abschätzung.

Die nachstehende 7.26 AS fasst das Ergebnis zusammen, während die 7.27 AS auf die Vertrauensgrenzen der mitgeteilten Angaben eingeht.

7.26 AS: Unter Anwendung eines von Pötz mitgeteilten Trennwertes für die DTK-Skala 2 ergibt sich der Hinweis auf den mit 8,4% der Gesamtstichprobe angegebenen Anteils einer Teilgruppe mutmaßlich mit klinischer Ausprägung einer depressiven Störung belasteter Kinder.

7.27 AS: Da die von Pötz vorgenommenen prognostischen Zuweisungen an die unterschiedenen zwei Gruppen mit Zuweisungsfehlern behaftet waren, ist auch die Quote der Hinweise auf mutmaßlich bestehender depressive Störungen fehlerhaft. Je nach dem Modus der Fehlerschätzung könnte die Hinweisquote zwischen 7% und 30% liegen.

7.4.2.3 Der nach Analyse der Gesamtpunktwerte aus Skala 2 des DTK vorliegenden Informationsstand

Auch mit der Analyse der Daten aus Skala 2 der griechischen Anwendung des DTK wurde nach Hinweisen auf das potentielle Bestehen depressiver Störungen nach DSM-IV und ICD-10 in der gezogenen Stichprobe gesucht. Die Bedingungen für das Auffinden solcher Hinweise sind erörtert worden und haben sich auch im Hinblick auf die Daten aus Skala 2 des DTK nicht verändert. Nach wie vor gilt:

1. Aus der Untersuchung 2034GR wird bekannt, welche und welche Mengen von Selbstzuschreibungen an Items aus dem DTK die Kinder individuell, in Teilgruppen und als gesamte Stichprobe abgegeben haben.
2. Die Analyse dieser Daten kann diese Selbstzuschreibungen benutzen, um eine differentielle Beschreibung des Antwortverhaltens und der dateninternen Abhängigkeiten sowie deren Relationen zu weiteren die Stichprobe kennzeichnenden Daten zu geben.
3. Für die über die Skala 2 erfassten Daten kann unter Annahme ihrer Konstruktvalidität für „agitiertes Verhalten/ autonome Reaktionen“ des Weiteren erfasst werden, welche Kinder nach verteilungsintern gesetzten Trennwerten gegenüber der absoluten Mehrheit in der Stichprobe signifikant höhere Mengen an Selbstzuschreibungen der konstruktspesifischen Symptome aus Skala 2 abgegeben haben.
4. Hinweise auf das mutmaßliche Auftreten depressiver DTK Störungen unter den Kindern der untersuchten Stichprobe werden erst erkennbar, wenn die erhobenen Daten der Untersuchung 2034GR über die Inanspruchnahme von Ergebnissen aus externen und deutschsprachigen Untersuchungen zur differentiellen Validität des DTK so bewertet werden, als ob diese Kennwerte auch auf sie zuträfen.
5. Die Inanspruchnahme dieser externen Daten aus der Studie von Pötz (2003) ist durch die messtheoretisch weitreichende Übereinstimmung der griechischen mit der deutschsprachigen Variante des Verfahrens zu rechtfertigen. Die Validität der griechischen Version ist dennoch empirisch ungeklärt. Aussagen dazu beruhen auf Annahmen.
6. Dennoch sind die Hinweise auf das mutmaßliche Bestehen von depressiven Störungen unter griechischen Schulkindern ein Informationszuwachs, weil sie eine über den bisherigen defizitären Wissensstand hinausgehende Annäherung an die mutmaßlich korrekte Quantifizierung des Auftretens depressiver Störungen unter griechischen Schulkindern sind.
7. Nach Analyse der über die Skala 2 des DTK bekannt gewordenen Daten ist eine Teilgruppe von 24% der Kinder aus der Gesamtstichprobe als eine Teilgruppe von Schulkindern mit vergleichsweise sehr hohen Selbstzuschreibungen von Items der Skala „agitiertes Verhalten“ (Skala 2, DTK) nachzuweisen gewesen.
8. Unter Anwendung der von Pötz (2003) mitgeteilten Trennwerte wird der Hinweis bekannt, dass nach der Verteilung der GPW aus Skala 2 ein Prozentsatz von 8,4% der erfassten Kinder mutmaßlich von depressiven Störungen betroffen sein könnte.
9. Die Diskrepanz zwischen den unter 7. und 8. mitgeteilten Hinweisen lässt die Annahme zu, dass, wie nach den Ergebnissen von Zimmermann (s.o.) zu erwarten, die Skala 2 zumindest auch auf das Bestehen von ADHS in der untersuchten Stichprobe verweist.

7.5 Analyse der über die Skala 3 des DTK in 2034GR erhobenen Daten

Die dem DTK von Rossmann unterlegte Theorie benutzt drei Konstrukte, deren diagnostisch relevantes Erfassen zu Informationen verhilft, die auf das Bestehen depressiver Störungen bei den untersuchten Persönlichkeiten verweisen. Jedem dieser Konstrukte ist im DTK eine eigene Skala zugeordnet. Nach Kenntnisnahme der über „Dysphorie/Selbstwert“ auf der Skala 1 erhobenen Daten und der über Skala 2 gewonnenen Informationen über „Agitiertes Verhalten“ wendet sich die Datenanalyse den Gesamtpunktwerten der DTK-Skala 3 (S3-GPW) zu. Auch diese Skala bietet den jeweiligen Probanden eine skalenspezifische Liste von Items an, die eine Antwort darüber anfordern, ob die Probanden meinen, von den genannten Phänomenen betroffen zu sein.

Diese von der Testanweisung verlangte Befassung mit den Items der Skala 3 wird wie bereits bei den zuvor analysierten Skalen 1 und 2 als Selbstzuschreibung bezeichnet. Von solchen Selbstzuschreibungen wird zunächst nicht mehr gewusst, als dass sie vollzogen werden. Welche Bedeutung die Selbstzuschreibung bestimmter Items hat, ergibt sich nicht unmittelbar aus der vollzogenen Handlung. Die Informationen über die Selbstzuschreibung sprechen zunächst eben nur für sich selbst. Ob sie darüber hinaus eine weitere Bedeutung haben, lässt sich nur mittelbar über die Heranziehung weiterer Informationen feststellen oder analysieren. In der klassischen Testtheorie repräsentiert der Begriff der Validität wesentliche Aspekte dieses Sachverhalts.

In der Itemliste der dritten DTK-Skala werden Items angeboten, die dem Konstrukt „Müdigkeit/ autonome Reaktionen“ zugerechnet werden sollen. Diese Benennung ist der Skala 3 beigegeben worden. Als Konstrukt wird „Müdigkeit/ autonome Reaktionen“ zunächst durch die Items erklärt, die als seine Operationalisierungen zu verstehen sind. Die im DTK verwendeten 13 Items sind allerdings schon Stellvertreter für eine weitaus größere Zahl von Items, die während der Entwicklung des Messverfahrens verwendet worden sind. Die 13 beibehaltenen Items führen zu Messwerten, deren innere Konsistenz trotz reduzierter Itemzahl den Messleistungen entspricht, die mit höheren Zahlen inhaltsähnlicher Items auch nur zu erreichen war. Das veranlasste die Aufgabe dieser größeren Itemmenge. Insofern sind die verwendeten Items messtechnisch gesehen Stellvertreter für die zuvor verwendete weitaus größere Itemmenge. Der Sachverhalt wäre demnach folgerichtig wie folgt darzustellen. Wer die angebotenen 13 Items oder eine Teilmenge von ihnen wählte, würde damit darauf hinweisen, dass er aus der bei der Konstruktion zuvor verwendeten größeren Itemmenge eine sehr wahrscheinlich in der Relation zwischen Gesamtmenge und gewählter Itemmenge nur zufällig abweichende Selbstzuschreibungen vornehmen würde. Das Konzept der Konstruktvalidität fasst den beschriebenen Sachverhalt in ähnlicher Weise.

Auf diesen Sachverhalt ist im Hinblick auf Skala 3 des DTK erneut eingegangen worden, um die Annahme der Konstruktvalidität auch für Skala 3 begründet beibehalten zu können.

Vor allen Überlegungen zur Bedeutung der über DTK-Skala 3 erhobenen Daten ist deren Aufkommen festzustellen und zur Kenntnis zu nehmen.

7.5.1 Verteilung der über die Skala 3 des DTK bekannt werdenden skalenspezifischen Gesamtpunktwerte (S3-GPW)

Aus der mit der Durchführung des DTK in einer griechischen Stichprobe gewonnenen Datenmenge sind die von den Mitgliedern der Stichprobe erreichten Gesamtpunktwerte als Analyseeinheiten für einen ersten Analysegang ausgewählt worden. Sie teilen als individuelle Werte mit, wie viele Symptome sich die jeweils betrachtete Persönlichkeit durch die Bearbeitung der entsprechenden Skala als sie selbst betreffend zugeschrieben hat. Sie lassen zugleich auch wissen, wie viele Items als nicht relevant zurückgewiesen wurden. Der individuelle

Gesamtpunktwert ist zwar über die Summe der gewählten Items festgestellt worden, lässt aber nicht die Zusammensetzung der Menge erkennen. Diese Zusammensetzung, die Konfiguration der Items in der Zahl der gewählten Items, ist durchaus informationsreich und bedeutsam. Die Konfiguration der gewählten Items wird aber vom Messverfahren zugunsten der Aussagekraft von Summen gewählter Items, den Gesamtpunktwerten, zurückgestellt. Die hier zu beschreibende Analysearbeit folgt dieser Setzung.

Die Darstellung der aus der Durchführung des DTK bekannt werdenden Gesamtpunktwerte wird nun mit dem Eingehen auf die Häufigkeitsverteilung zur dritten DTK-Skala „Müdigkeit/ autonome Reaktionen“ fortgesetzt. Diese Häufigkeitsverteilung wird für 2024 Kinder, die die Skala3 vollständig bearbeitet hatten, in Tabelle 7.14 wiedergegeben.

Tabelle 7.14: Verteilung der Skala3-Gesamtpunktwerte (S3-GPW)

S3-GPW	Frequenz	Prozent	S3-GPW	Frequenz	Prozent
0	76	3,8	7	202	10,0
1	136	6,7	8	159	7,9
2	202	10,0	9	115	5,7
3	242	12,0	10	68	3,4
4	277	13,7	11	47	2,3
5	257	12,7	12	16	,8
6	219	10,8	13	8	,4
			Summe	2024	100,0

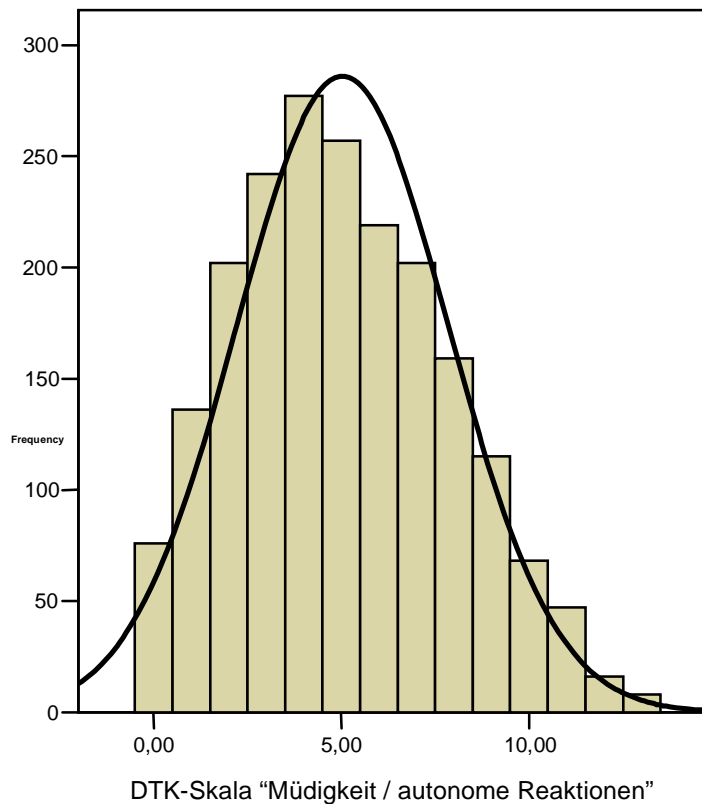
Die in Tabelle 7.14 dargestellte Häufigkeitsverteilung wird von den nachfolgend genannten statistischen Werten gekennzeichnet.

Tabelle 7.15: Statistische Kennwerte der voran stehenden Tabelle zur Häufigkeitsverteilung der S3-GPW im DTK

Mean	5,0331	Range	13,00	Skewness	,313
Std. Error of Mean	,06275	Minimum	,00	Std. Error of Skewness	,054
Median	5,0000	Maximum	13,00	Kurtosis	-,497
Mode	4,00	Sum	10187,00	Std. Error of Kurtosis	,109
Std. Deviation	2,82307	Variance	7,970		

Den Kurvenverlauf dieser Verteilung gibt die nachfolgende Abbildung wieder.

Abbildung 7.3: Die Häufigkeitsverteilung der DTK-Skala „Müdigkeit/autonome Reaktionen“



Wie schon zuvor bei den DTK-Skalen 1 und 2 ist zu prüfen, ob die Verteilungsform den Bedingungen für eine Normalverteilung entspricht. Die Durchführung eines Kolmogorov-Smirnov Tests für die Prüfung der Verteilungsform dieser Skala führt wiederum zur Ablehnung der Annahme einer Normalverteilung ($Z = 4,668$; $p = 0,000$).

Damit sind die nachstehenden Feststellungen zur Verteilung der Gesamtpunktwerte aus Skala 3 gerechtfertigt.

7.28 AS: Die Verteilung der Gesamtpunktwerte weicht wie schon bei den Skalen 1 und 2 des DTK auch bei Skala 3 von der Form der Normalverteilung ab. Die linksschiefe Verteilung ist gegenüber der Normalverteilung in ihrem Verlauf bei den höheren S3-GPW signifikant abgeflacht.

7.29 AS: Die Verteilungsform entspricht, wie auch die bei den Skalen 1 und 2, dem für den DTK geltenden Messkonzept. Die aussagekräftigeren höheren S3-GPW haben eine signifikant geringere Frequenz als die GPW aus den unteren zwei Dritteln des Skalenbereiches.

7.30 AS: Die von der Normalverteilung signifikant abweichende linksschiefe Verteilung der S3-GPW führt zu den sie selbst kennzeichnenden Werten. Mit den ersten sieben Gesamtpunktwerten von 0 einschließlich und bis und mit GPW 6 als Angebot kommen bereits etwa 70% der teilnehmenden Schulkinder für das Ausmaß ihrer Selbstzuschreibungen aus. Bei der Marke von 64,3% des Bestandes an Items, das ist der GPW 8, der neunte wählbare Wert einschließlich GPW 0, sind bereits 87,6 % der Kinder dieser Stichprobe repräsentiert. Knapp 2/3

der angebotenen Items sind hinreichend für die Selbstzuschreibung von fast 88% der beteiligten Schulkinder.

7.31 AS: Der S3-GPW 10 liegt als erster unter dem Frequenzanteil an der Gesamtstichprobe, der dem S3-GPW 0 zukommt. Damit ist die hier zuvor definierte verteilungsinterne Trennstelle zwischen Kindern mit relativ hohen Selbstzuschreibungen an Items der Skala 3 und den mit über 8/10 der Gesamtstichprobe die Mehrheit bildenden anderen Kindern, die mit signifikant weniger Selbstzuschreibungen auskommt, als $GPW < 10$ bestimmt.

7.32 AS: Der nach der 7.31 AS bestimmte Trennwert ist nach der Regel ausgewählt worden, die auch für Skala 1 und Skala 2 Anwendung fand. Diese Regel setzt ein verteilungsabhängiges Kriterium zum Auffinden des Trennwertes ein. Das Maß für das Auffinden des Trennwertes ist das Ausmaß der Nutzung des GPW 0. Trennwert für das Abtrennen einer als mit relativ hohen Selbstzuschreibungen versehenen Teilgruppe ist jener Frequenzwert, der niedriger ist als die Inanspruchnahme des GPW 0 in der zu analysierenden Verteilung von GPW aus Skalen des DTK. Die Verteilung beginnt bei GPW 0, gipfelt bei einem der Folgewerte und fällt jenseits des Gipfelwertes bei einem anderen Wert erstmals unter die Frequenz von GPW 0. Das ist der Trennwert, der den ersten Beitrag zur Größe der gesuchten Teilgruppe von Kindern mit hohen Selbstzuschreibungen an Skalenitems liefert.

Der Trennwert wird also jeweils durch den Frequenzwert bestimmt, der durch die Persönlichkeiten zustande kommt, die durch ihre Selbstzuweisungen bekunden, von keinem der Symptome betroffen zu sein, die mit der Itemliste der Skala zur Auswahl angeboten worden sind. Die mit hohen GPW abgetrennte Teilstichprobe hat dann jeweils ein Skalenstück allein für sich beansprucht. Dieses Skalenstück ist durch die Zahl der dort befindlichen Items quantifizierbar. Diese Teilmenge an GPW erreicht jeweils einen Anteil an der Gesamtzahl aller erreichbaren GPW, der durch den Prozentsatz zu diesem Grundwert quantifiziert ist. Dieser Prozentsatz kann dann als ein von Skala zu Skala benutzbarer Vergleichswert genommen werden. Er präsentiert den Teil der GPW-Skala, den die abgetrennte Teilstichprobe für ihre eigenen Selbstzuweisungen in Anspruch genommen hat. Je höher dieser Prozentsatz ist, desto ausgeprägter setzen sich die Teilnehmer in der abgetrennten Teilstichprobe von dem durch Selbstzuschreibungen bekundeten Verhalten der hochfrequent überwiegenden Mehrheit der Teilnehmer im Skalenbereich unterhalb des Trennwertes ab.

7.33 AS: Der von der signifikant abweichenden Minderheit okkupierte Anteil an den in der Verteilung zur Verfügung stehenden Gesamtpunktwerten ist für die drei DTK-Skalen in dieser griechischen Untersuchung unterschiedlich. Für Skala 1 „Dysphorie/Selbstwert“ sind 50 % der GPW von der Minderheit mit signifikant hohen durchschnittlichen Selbstzuweisungen in Anspruch genommen worden. Für die Skala 2, „Agitiertes Verhalten“, wurden 68,75% ermittelt. Auf Skala 3 werden 10 von 14 auftretenden GPW durch die Mehrheit und nur 4 von der Minderheit besetzt, die damit auf 28,75% der GPW beschränkt ist.

7.34 AS: Der für die Skala 3 geltende Prozentsatz an okkupierten GPW von 28,75 verweist darauf, dass ein Grossteil der wählbaren Mengen an skalenspezifischen Items von der Teilstichprobe der mit ihrem GPW unterhalb des Trennwertes befindlichen Mehrheit an Schulkindern beansprucht wird.

7.35 AS: Das Konstrukt „Müdigkeit/ autonome Reaktionen“ variiert als Gesamtpunktwert in der vom DTK benutzten Bestimmung mit relativ größeren Mengen auch in der Mehrheit der Schulkinder dieser Stichprobe als die GPW der Konstrukte zu den Skalen 1 und 2. Symptome von „Müdigkeit und autonomen Reaktionen“ treten in der vom DTK verwendeten Bestimmung auch bei der Mehrheit der Schulkinder dieser Stichprobe mit relativ hohen Frequenzen auf.

7.5.2 Aus den Gesamtpunktwerten der DTK-Skala 3 ableitbare Hinweise auf das mutmaßliche Auftreten depressiver Störungen in der untersuchten Stichprobe

Die Bearbeitung der DTK-Skala 3 durch 2024 griechische Schulkinder hat zu ebenso vielen Gesamtpunktwerten aus Skala 3 geführt, deren Verteilung hier vorangehend dargestellt, erläutert und analysiert worden ist. Diese Gesamtpunktwerte sind Informationsträger über die von den teilnehmenden Schulkindern vorgenommenen Selbstzuschreibungen von Items der Skala 3. Den Arbeitsschritten bei der Analyse der Daten aus Skala 2 folgend, wird nun mitgeteilt, welche Hinweise damit erfassbar wurden.

7.5.2.1 Informationen über die Selbstzuschreibung von Symptomen aus der Itemliste der Skala 3

Mit der Durchführung des DTK und der aufgabengerechten Bearbeitung der Items zu Skala 3 durch die an der Untersuchung beteiligten griechischen Schulkinder werden weisungsgemäß Selbstzuschreibungen an Symptomen aus der Itemliste zu Skala 3 vorgenommen. Über welche psychischen Prozesse diese Selbstzuschreibungen erfolgen, beachtet das Messverfahren nicht. Die Messung orientiert sich an den durch das Ankreuzen der entsprechenden Felder auf dem Antwortbogen mitgeteilten Entscheidungen.

Über einige Aspekte dieses Wahlverhaltens der Mitglieder der untersuchten griechischen Stichprobe informierte die Verteilung der Gesamtpunktwerte aus Skala 3. Die Informationen sind auf Mengenangaben begrenzt, die als absolute Mengen an individuell erreichten GPW wie auch als Verteilung von GPW auf Teilgruppen mitgeteilt werden kann. Wie schon zuvor bei den beiden anderen DTK-Skalen werden nun auch die individuell erreichten Gesamtpunktwerte aus Skala 3 benutzt, um drei Teilgruppen von Schulkindern zu bilden. Unterschieden werden die mit relativ hoher Selbstzuschreibung an Symptomen zum Konstrukt „Müdigkeit/ autonome Reaktionen“ (HMaR) zum einen von den Kindern, die als Gruppe der mittleren Selbstzuschreibung (MMaR) und zum anderen von denen mit relativ niedrigen Selbstzuschreibungen (NMaR). Die Trennwerte zwischen den Gruppen werden über die schon zuvor (7.3.3.1 und 7.4.2.1) angewendete Regel zur Bestimmung eines verteilungsinternen Trennwertes zwischen HMaR einerseits und MMaR und NMaR andererseits eingeleitet. Die unterhalb des Trennwertes verbleibende Skalenstrecke für die Teilgruppen MMaR und NMaR wird halbiert und entsprechend zugewiesen.

Tabelle 7.16: Wertebereiche dreier Teilgruppen von Teilnehmern mit unterschiedlichen Selbstzuweisungen von Mengen an Items der Skala 3 (HMaR, MMaR und NMaR)

Teilgruppe	S3-GPW	Zahl der GPW
NMaR	0 – 4	5
MMaR	5 – 9	5
HMaR	10 - 13	4

Die Bereiche an Gesamtpunktwerten aus Skala 3 für die drei zu bildenden Gruppen sind der oben angeführten Tabelle zu entnehmen. Welche Zahlen an Kindern den einzelnen Gruppen zugeteilt werden können, zeigt die nachfolgende Tabelle 7.17.

Tabelle 7.17: Teilgruppen von Teilnehmern mit unterschiedlichen Selbstzuweisungen von Mengen an Items der Skala 3 (HMaR, MMaR und NMaR)

Teilgruppe	Anzahl Kinder	Prozentsatz
NMaR	933	46,1
MMaR	952	47,04
HMaR	139	6,87
Summe	2024	100,01

Die gebildeten drei Teilgruppen mit unterschiedlichen Selbstzuweisungen an Symptomen aus der Itemliste der DTK-Skala 3 sind unterschiedlich frequentiert. Die Teilgruppen NMaR und MMaR sind prozentual fast übereinstimmend besetzt. Da die Trennung der Teilgruppen über einen aus der Verteilungsform bestimmten ersten (zwischen MMaR und HMaR) Trennwert GPW=10 erfolgte, kann die Gleichverteilung der beiden unteren Teilgruppen auf der Skala wie in der 7.36 AS beschrieben werden.

7.36 AS: Die annähernde Gleichverteilung der Frequenzen für die Teilgruppen NMaR und MMaR zeigt an, dass fast identische Mengen an Schulkindern in der Stichprobe die GPW von 0 bis 4 und von 5 bis 9, also fast spiegelgleich, besetzen.

7.37 AS: Von dieser Gleichverteilung auf die beiden Teilgruppen NMaR und MMaR weicht die Besetzung der Teilgruppe HMaR deutlich ab. Eine Teilmenge von 6,87% aller Schulkinder der Stichprobe, das sind 139 Kinder von 2034, wählt höhere Mengen an Symptomen aus der Itemliste Skala 3 als auf sie zutreffend aus, als über 90% der Kinder in dieser griechischen Stichprobe.

7.38 AS: Die Datenlage erlaubt den Hinweis, dass zumindest 6,87% der Schulkinder in dieser Stichprobe mit selbstzugewiesenen Symptomen nach Skala 3 „Müdigkeit/autonome Reaktionen“ belastet scheinen (Teilgruppe mit hohen Selbstzuschreibung von Symptomen aus Skala 3).

7.5.2.2 Hinweise auf das mutmaßliche Auftreten störungsrelevanter Ausprägungen des Konstruktes „Müdigkeit/autonome Reaktionen“

Wie schon zuvor bei der Beantwortung der Frage nach den möglichen Bedeutungen der Gesamtpunktwerte aus den DTK Skalen 1 und 2 sollten die Gesamtpunktwerte aus DTK-Skala 3 auch nach Hinweisen auf mutmaßlich bestehende depressive Störungen abgesehen werden. Die Übertragung der zuvor realisierten Verfahrenspraxis auf die Analyse der S3-GPW entbehrte aber der Rechtfertigung durch zumindest ein externes Untersuchungsergebnis zur Konstruktvalidität der DTK-Skala 3. Für die Skalen 1 und 2 waren Untersuchungsergebnisse von Zimmermann (s. dazu Rossmann, 2005) als Rechtfertigung für die Annahme einer Konstruktvalidität der Skalen 1 und 2 heran gezogen worden. Ganz im Gegensatz zu den bei Zimmermann (s.o.) zu diesen Skalen auffindbaren Hinweisen auf eine Konstruktvalidität fehlen für Skala 3 entsprechende Belege. Die von Zimmermann in seine Untersuchung einbezogenen klinischen Gruppen unterscheiden sich nicht in den ihnen jeweils zukommenden Mittelwerten an Gesamtpunkten auf Skala 3. Das erscheint nach Kenntnisnahme der Verteilung der S3-GPW in der hier betrachteten Stichprobe fast folgerichtig, obschon die allein auf die Untersuchung von Zimmermann repräsentierte und somit äußerst schmale Datenbasis deutlich überinterpretiert sein dürfte. Immerhin war die Breite des möglicherweise als klinisch relevant anzusehenden Anteils an Gesamtpunktwerten mit 28,75 nur etwas mehr als die Hälfte der für Skala 1 bekannt gewordenen Breite an Gesamtpunktwerten außerhalb der Inanspruchnahme durch die absolute Mehrheit der Stichprobe.

Eine Aussage zu den nach über die nach den Gesamtpunktwerten aus DTK-Skala 3 zu ermittelnden Informationen über etwaige klinische Ausprägungen eines verhaltenswirksamen Symptomkomplexes „Müdigkeit/autonome Reaktionen“ unterbleibt, weil Gründe bestehen, die Konstruktvalidität für Skala 3 anzuzweifeln. Das wird auch in der nachstehenden Aussage vermerkt.

7.39 AS: Die Teilgruppe der Kinder mit hohen Selbstzuschreibungen an Symptomen der Skala 3 zum Konstrukt „Müdigkeit/autonome Reaktionen“ umfasst in der Untersuchung 2034GR, bestimmt über den verteilungsinternen Trennwert, insgesamt 6,87% der Schulkinder in der Gesamtstichprobe.

7.5.2.3 Informationen aus den Gesamtpunktwerten auf Skala 3 im DTK über Hinweise auf klinisch ausgeprägte depressive Störungen in der untersuchten Stichprobe

Obschon mit dem DTK drei Subskalen zur Erfassung depressionsspezifischer Merkmale eingesetzt werden, erwies sich in Untersuchungen von Pötz (s. Rossmann, 2005) für jede der Skalen eine beachtenswerte depressionsspezifische diskriminante Validität. Unter Beibehaltung der in diesen Erörterungen schon an mehreren Stellen aufgeführten Einschränkungen werden die von Pötz (s.o.) mitgeteilten Trennwerte auf die hier zu analysierenden griechischen Daten übertragen. Die Vorbehalte sind nicht aufgehoben. Immerhin erweitern aber die so zu gewinnenden Informationen den bisherigen defizitären Informationsstand über Prävalenzen in repräsentativen griechischen Stichproben mit Schulkindern.

Nach Pötz (s.o.) wird unter den Gesamtpunktwerten aus Skala 3 jener zum Trennwert, der die im Vergleich zu anderen GPW höchste Sensitivität beim Erkennen von Kindern erreicht, für die Diagnosen über depressive Störungen bekannt waren. Bei gleicher Sensitivität ist der Wert mit der zusätzlich höchsten Spezifität auszuwählen.

Tabelle 7.18: S3-GPW als Trennwert zwischen vermutlich mit klinischen Ausprägungen depressiver Störungen belasteten Kindern und anderen

Skala	Sensitivität	Spezifität
Skala1 > 10 oder skala3 > 6	75,76	60,61
Skala1 > 10 oder skala3 > 7	75,76	66,67

Der S3-GPW > 7 ist der gesuchte Trennwert.

Wird dieser von Pötz (s.o.) vorgeschlagene Trennwert aus DTK-Skala 3 benutzt, ergibt sich die in der folgenden Tabelle aufgeführte Verteilung der Teilgruppen von Kindern mit unterschiedlichen GPW. Die Gruppen werden im Unterschied zu den weiter oben zu findenden Bezeichnungen solcher Teilgruppen als MAR1, MAR2 und MAR3 benannt. Dem unterliegt die Hypothese, der vermutete Grad der Belastung mit depressiven Symptomen werde aufsteigend, von GPW=0 in Richtung auf GPW=3, Hinweiser auf zunehmend höhere Ausprägungen depressiver Störungen als wahrscheinlich auftretend geben können. Die benutzte Formulierung soll die mit solchen Hinweisen zwangsläufig verbundene Unsicherheit erkennen lassen.

Tabelle 7.19: Teilgruppen mit vermutlich unterschiedlichen Belastungen durch depressive Symptome (Cutoff GPW >7, nach Pötz, s. Rossmann, 2005)

Teilgruppe	Anzahl Kinder	Prozentsatz	S3-GPW
MaR1	656	32,5	0 – 3
MaR2	955	47,2	4 -7
MaR3	413	20,1	8 – 14
Summe	2024	99,8*	
	Rundungsfehler	0,2%	

MaR1, MaR2, MaR3 = nach Gesamtpunktwerten DRT-Skala 3 unterschiedene Teilgruppen von Schulkindern mit vermutlich unterschiedlicher Belastung durch depressive Symptome

Nach dem unter Berufung auf Pötz (s. Rossmann, 2005) verwendeten Trennwert S3-GPW > 7 ergibt sich eine mutmaßlich mit depressiven Störungen belastete Teilgruppe von 413 Kindern, was einer prozentualen Quote von 20,1% der Gesamtstichprobe entspricht.

7.40 AS: Unter Anwendung eines nach Pötz (s. Rossmann, 2005) bestimmten Trennwertes S3-GPW > 7 ergeben sich Hinweise auf 413 Kinder, das sind 20,1% der Gesamtstichprobe, die mutmaßlich in klinisch beachtenswertem Ausmaß durch depressive Symptome betroffen sind.

7.6 Eine Zusammenlegung von zwei Skalen des DTK als zusätzliche Datenbasis: Skala 1 plus Skala 2

Die der Testkonstruktion des DTK implizit oder explizit unterliegende Theorie führte zur Konstruktion dreier Skalen. Die erste von ihnen, Skala 1, ist mit ihren Items dem Konstrukt „Dysphorie/Selbstwert“ zugeordnet. Die Skala 2 wird inhaltlich vom Konstrukt „Agitiertes Verhalten“ bestimmt, und die Skala 3 misst „Müdigkeit/ autonome Reaktionen“.

Schon in der Erörterung des Informationsstandes nach Analyse der Gesamtpunktwerte aus Skala 3 war in der hier geführten Erörterung ansatzweise auf die nicht nur messtheoretisch interpretierbare Problematik der Beziehungen zwischen den Messergebnissen auf den drei DTK-Skalen und dem über Symptome zu beschreibenden Störungsbild verwiesen worden. Die Symptomatik in den Handbüchern DSM und ICD ist oberbegrifflich dargestellt. Die Handbücher enthalten keine Listen aller nur denkbaren depressiven Symptome. Die psychologischen Vorstellungen darüber, wie denn in der jeweils betroffenen Persönlichkeit die Regulation der auftretenden depressiven Symptome erfolgt, sind insgesamt nicht hinreichend homogen, um eine „regulative Struktur“ von psychischen Instanzen verbindlich vertreten und nachfolgend nutzen zu können. Die Repräsentation übergreifender psychischer Funktionszusammenhänge als regulative Vorgaben im Zentralen Nervensystem von der initialen Orientierung bis hin zum Hervorbringen speziellen Verhaltens ist in anderen Zusammenhängen beispielsweise von H.J. Eysenck versucht worden. Indem er unter anderem Extraversion und Introversion, beziehungsweise die individuelle Position auf der von ihnen gebildeten Dimension, als eine der regulative Eigenheiten des Zentralen Nervensystems bestimmte, ließen sich regulative Hierarchien entwerfen, die systemimmanent bis zum individuellen Hervorbringen von Verhalten auszuformulieren waren. Wie weit sie dann über entsprechende empirische Befunde nachweisbaren Erklärungswert erreichten, wird hier nicht zu beantworten versucht. Das Beispiel bildet hier nur den Kontrast, vor dem sich die Situation in der Depressionsforschung abhebt. In den psychologischen Theorien zur Depression wären zwar ähnliche regulative Hierarchien konzipierbar, die aber hätten ihrerseits nicht nur den

Ursprungstheorien, so etwa den von A. Beck (1976) oder Seligman (1979) vertretenen zu folgen, solche Systeme müssten gleichzeitig voll kompatibel mit den derzeit geltenden Störungsbildern nach DSM und ICD sein. Das aber wäre allem Anschein nach gegenwärtig nicht zu sichern. Für die Testpraxis gilt deshalb, dass sie sich mühen muss, mit ihren Konstruktionen dem Ziel näher zu kommen, das für einen diagnostisch relevanten Test gesetzte Kriterium zu erreichen. Dazu findet sich in entsprechenden Handbüchern der Hinweis, ein solcher Test müsse mindestens eine auf die jeweilige Diagnose zu beziehende Trefferquote von 95% nachweisen. Da nun aber die Theorie der Testkonstruktion von depressionsspezifischen Messverfahren nicht eindeutig zu beschreiben vermag, wie dieses Kriterium sicher erreichbar wäre, bleibt immer wieder nur der Versuch, über diese oder jene Vorgehensweise eine möglichst beachtenswerte Annäherung an die als Ziel gesetzte Trefferquote von mindestens 95% zu vollziehen. Eine solche Strategievариante ist das Erproben der diskriminanten Leistungen eines Verfahrens durch fortgesetzt variierende und empirisch geprüfte Bewährungsaufgaben. Die Untersuchung von Pötz (s. Rossmann, 2005) zur diskriminanten Validität des DTK ist ein solcher Versuch. Die Untersuchung von Zimmermann (ebd.) ist ein weiterer. Da Pötz auch Verfahrensvarianten zu nennen weiß, die unterschiedlich diskriminante Leistungen des DTK in ihrer Untersuchung zur Folge hatten, liegt nahe, einem ihrer Vorschläge zu folgen und möglichst einem relativ effizienten. Eine solche Verfahrensvariante ist unter vergleichbaren anderen die Kombination der DTK-Skalen 1 und 2 zu einer neuen gemeinsamen Skala. Die Gesamtpunktwerte aus dieser neu gebildeten Skala erreichen nach Pötz ab einem von ihr gefundenen Trennwert eine beachtenswerte diskriminante Leistung bei der Aufgabe, möglichst fehlerfrei zwischen Kindern mit klinisch relevanten Ausprägungen depressiver Störungen und anderen Kindern zu unterscheiden. Diesem Vorschlag zu r Kombination von zwei DTK-Skalen zur Gewinnung einer neuen Integrationskala wird hier nun gefolgt. Das Argument für die Nutzung der Kombination ist die von Pötz berichtete diskriminante Leistung des Gesamtpunktwertes 18 als Trennwert. Mit ihm wird eine Sensitivität von 81,5 erreicht (Rossmann, 2005). Der numerische Wert relativiert sich durch die mit dem Trennwert 18 insgesamt erreichte Trefferquote von etwa 75% insgesamt.

Die neue Skala aus den DTK-Skalen 1 und 2 wird nachfolgend, um Verwechslungen mit den anderen DTK-Skalen zu vermeiden, als Skala „Depressive Belastung“ bezeichnet. Diese Benennung ist auch anders zu begründen. Keine andere von Pötz genannte Skala oder Skalenkombination benennt mehr Kinder zutreffend als nicht oder doch depressiv belastet. Dennoch ist statistisch betrachtet, jede vierte Zuweisung falsch. Von 2005 durch die Skala DTK Skala 1+2 repräsentierten Kindern werden danach etwa 501 falsch zugewiesen.

7.6.1 Die Verteilung der DTK Gesamtpunktwerte aus einer Skalenkombination unter der Bezeichnung „Depressive Belastung“ (DTK-Skala 1 plus 2)

Die Analyse der drei Häufigkeitsverteilungen von DTK-Gesamtpunktwerten auf den drei DTK Skalen ist als ein erster Arbeitsschritt zur Gewinnung von depressionsspezifischen Informationen aus den mit der Durchführung dieses Verfahrens gewonnenen Daten zu betrachten. Die Analyse der Verteilung der mit den drei Skalen jeweils bekannt werdenden Gesamtpunktwerte wurde unter vielen anderen möglichen Vorgehensweisen als ein erster Arbeitsschritt ausgewählt, weil so sowohl Deskriptionen der Verteilungen von Punktwerten als auch Kennwerte für das Antwortverhalten der untersuchten Kinder auf die nach Skalen geordneten Fragen innerhalb des DTK gleichzeitig zu erfassen sind, und beides ist der angestrebten Beantwortung der dieser Untersuchung unterliegenden Forschungsfragen dienlich.

Da Daten aus drei Subskalen vorliegen, wären permutierte Kombinationen von jeweils zwei Skalen und eine Dreierkombination denkbar. Nach Pötz ist aber die Kombination Skala 1 + Skala

2 bei der angestrebten Vorhersage von Kindern mit und ohne klinischen Ausprägungen depressiver Störungen erfolgreicher als andere.

Das Verfahren DTK erfasst das Antwortverhalten der untersuchten Kinder auf die Items der Skalen und erlaubt deren Interpretation als Indizien für existente Ausprägungen eines verfahrensexternen Konstruktes, dem der depressiven Störung. Gesamtpunktwerte können unter entsprechenden Vorbehalten einerseits als Indizien für graduelle Abstufungen der individuellen Betroffenheit der untersuchten Kinder durch depressive Symptome verstanden werden (Skala1+2 „depressive Belastung“). Andererseits sind sie Indikatoren für empirisch unterscheidbare Äußerungen einer selbstberichteten depressiven Symptomatik und dies dann auf getrennten Skalen als Aufzählung von Symptomen subjektiver dysphorischer Stimmung (Skala1 „Dysphorie“), als Symptomkomplex einer subjektiven Tendenz zur Agitiertheit (Skala2 „Agitiertes Verhalten“) und als Aufzählung psychosomatischer Belastungen (Skala 3 „Müdigkeit/autonome Reaktionen“).

Obschon die durchgeführte Untersuchungsarbeit die Suche nach Hinweisen auf depressive Belastungen bei griechischen Schulkindern betreibt, wird einer Theorie der Konstruktion eines deutschsprachigen DTK gefolgt, die mit der Überführung der Items ins Griechische verbunden bleibt. Auf diese Konstrukte bezogen sind weiterhin die Befunde aus seiner Anwendung in deutschsprachigen Stichproben. Alle die dort gefundenen Daten betreffen ein deutschsprachiges Instrument. Indem solche Ergebnisse auf die griechische Übersetzung angewendet werden, müssen sie als Vorgaben angesehen werden, deren Nützlichkeit sich an deutschen Untersuchungen erwiesen haben mag, deren Transponierbarkeit auf griechische Verhältnisse sich dann aber erst nachträglich prüfen ließe. Zunächst bliebe zu hoffen, die mit deutschsprachigen Untersuchungen gewonnenen Ergebnisse und Befunde würden sich durch griechische Anwendungen replizieren lassen. Die transkulturelle Problematik wird einer bestehenden Pragmatik unterworfen. Ob das so hilfreich wie erhofft ist, wird sich erweisen müssen. Die künftige Erfahrung muss solche Hoffnung keinesfalls bestätigen, aber sie könnte es.

Im Sinne dieser Ausführungen sind die nachfolgenden Tabellen zu verstehen, denn sie sind im Rückblick auf Befunde aus deutschsprachigen Anwendungen des DTK formatiert worden, ohne die Ergebnisse einer Prüfung der inhaltlichen Berechtigung zur Übernahme in eine griechische Untersuchung abgewartet zu haben.

Der Zusammenlegung zweier DTK-Skalen, Skala1 und Skala, ging die Analyse der sich aus den Antworten der untersuchten Kinder auf den Einzelskalen ergebenden Verteilungen voraus. Dementsprechend sind auch die Ergebnisse aus den Analysen zur DTK-Skalen 1, „Dysphorie/Selbstwert“, und 2, „Agitiertes Verhalten“ und zur Skala 3, „Müdigkeit/autonome Reaktionen“ bekannt. Mit der aus Skala 1 und Skala 2 kombinierten neuen DTK-Skala „Depressive Belastung“ steht eine weitere Verteilung zur Untersuchung an.

Nach Pötz (s.o.) dürften die Daten aus dieser kombinierten Skala gegenüber einigen anderen aus dem DTK als valider gegenüber dem Kriterium der depressiven Störung gelten. Die vorangestellten Ausführungen relativieren dies mit dem Hinweis auf die deutschsprachige Bindung der Bedeutung solcher von Pötz gegebenen Informationen.

Die Tabelle 7.20 zeigt die Verteilung der von den untersuchten Kindern erreichten Gesamtpunktwerte in der DTK – Skala 1+2, „Depressive Belastung“. Die in dieser Tabelle aufgeführten Zahlen gehen auf die Daten derjenigen Kinder zurück, die auf alle 41 Fragen der Skala1+2 geantwortet haben. Ihre Zahl (n = 2005) entspricht 98.6% aller Kinder, die an der Untersuchung 2034GR teilgenommen haben. Die zu vermissenden Kinder hatten ein Item oder

eben auch mehrere Items nicht bearbeitet und wurden mit ihren Daten bestimmungsgemäß aus der Datenbasis der nachfolgenden Berechnungen genommen.

Tabelle 7.20: Verteilung der Skala1+2-Gesamtpunktwerte (S1+2-GPW)

S1+2-GPW	Frequenz	Prozent	S1+2-GPW	Frequenz	Prozent
0	23	1,1	18	81	4,0
1	48	2,4	19	52	2,6
2	45	2,2	20	51	2,5
3	81	4,0	21	45	2,2
4	75	3,7	22	49	2,4
5	84	4,2	23	36	1,8
6	102	5,1	24	20	1,0
7	101	5,0	25	24	1,2
8	125	6,2	26	11	,5
9	101	5,0	27	20	1,0
10	124	6,2	28	17	,8
11	110	5,5	29	9	,4
12	99	4,9	30	17	,8
13	104	5,2	31	3	,1
14	84	4,2	32	4	,2
15	97	4,8	33	4	,2
16	82	4,1	34	2	,1
17	74	3,7	36	1	,0
			Summe	2005	100,0

Die Verteilung der Skala 1+2 – Gesamtpunktwerte ist von den nachstehenden statistischen Werten gekennzeichnet.

Mean	12,1930	Range	36,00	Skewness	,526
Std. Error of Mean	,15528	Minimum	,00	Std. Error of Skewness	,055
Median	11,0000	Maximum	36,00	Kurtosis	-,170
Mode	8,00	Sum	24447,00	Std. Error of Kurtosis	,109
Std. Deviation	6,95302	Variance	48,344		

Um zu prüfen, ob die in Tabelle 1 dargestellten Werte in ihrer Verteilung der Normalverteilung entsprechen, wird der Kolmogorov-Smirnov Test eingesetzt. Die Ergebnisse zeigen, dass keine Übereinstimmung mit einer Normalverteilung vorliegt. (Kolmogorov-Smirnov $Z = 3,454$; $p = 0,000$). Die Positionen des arithmetischen Mittels ($AM = 12,1930$), des Medianwertes ($Md = 11,0$) und des Modalswertes ($Mo = 8,00$) verweisen auf eine linksteile Verteilung, die durch den negativen Exzeß (Kurtosis = $-0,170$) als flacher und breitgipfliger gegenüber einer Normalverteilung zu charakterisieren ist (Bortz, 1999, 5. Aufl., S. 34, 39 und 46; Brosius und Brosius, 1998, S. 303). Die nachfolgende Abbildung 7.4 veranschaulicht die gefundene linksasymmetrische Verteilung der Gesamtwerte, die die Kinder der untersuchten Stichprobe in der DTK-Skala „Depressive Belastung“ erbracht haben. Zuvor bleibt zu bedenken, welche Bedeutung der mit dem Kolmogorov-Smirnov Tests bewiesenen signifikanten Abweichung der

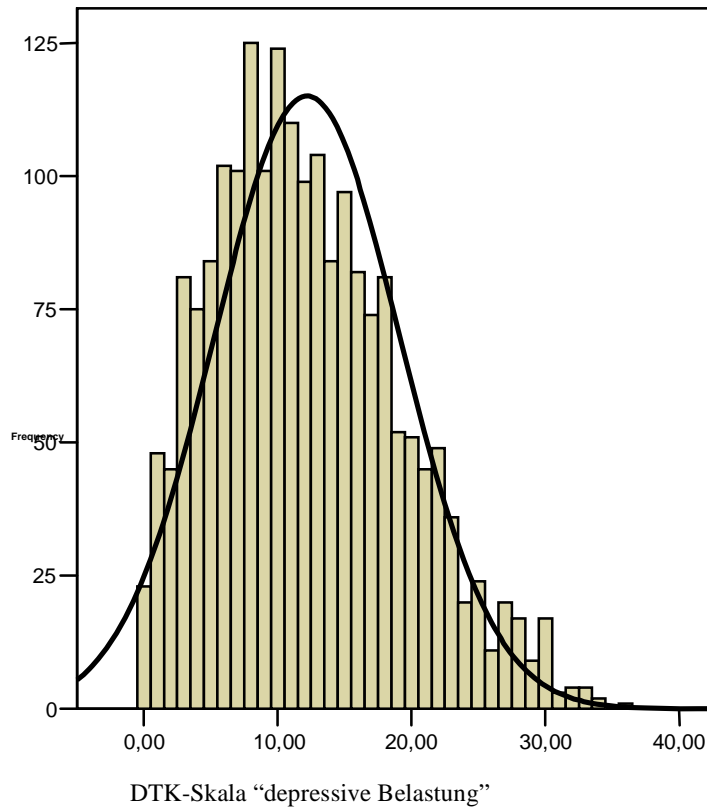
Verteilung der Gesamtpunktwerte aus Skala 1+2 von der Normalverteilung innerhalb dieser Untersuchung zukommt.

Nach Lienert und Raatz (1994, 5. Aufl., S. 151) ist die Anomalität einer Rohwertverteilung auf eine mögliche Stichprobenheterogenität, auf ungünstige messtechnischen Eigenschaften der verwendeten Messmethode oder auf die speziellen Eigenheiten des Messobjektes zurück zu führen. Die Autoren führen „...insbesondere Linksasymmetrie...“ (Lienert und Raatz, s.o., S. 158) auf die Natur des untersuchten Merkmales selbst zurück. Im Falle „pathologischer“ Konstrukte, so auch von Depressionen bei Kindern, ist davon auszugehen und auch zu hoffen, dass sie relativ ungewöhnlich sind, sonst hätten sie nicht ihren Abweichungscharakter, sie wären „normal“. In einem solchen Fall wäre aber, weil Normalität und Abweichung identisch wären, auch keine Existenzberechtigung für die Psychiatrie zu erkennen. Aus der Durchführung einer zur Erfassung psychopathologischer Symptome konstruierten Depressionsskala bei einer eher unauffälligen Population von Schulkindern dürfte sich keine normale Häufigkeitsverteilung depressiver Symptome ergeben. Die für die DTK-Skala „Depressive Belastung“ gefundene Verteilungsform entspricht demnach den theoretischen Erwartungen. Ob die untersuchte Stichprobe homogen genannt werden sollte, oder ob sie als heterogen bezeichnet werden müsste, weil im Hinblick auf die eingebrachten Gesamtpunktwerte divergierende Untergruppen von Kindern zu erkennen wären, was wiederum von der Depressionstheorie her nicht auszuschließen ist, wird an anderen Stellen dieser Erörterungen geprüft werden müssen.

Die Bedeutung der hier betrachteten Abweichung von der Normalverteilung liegt für die darzustellende Untersuchung insbesondere in dem mit ihr empirisch fassbaren Hinweis auf den Erhalt der mit der deutschsprachigen Konstruktion des DTK Realität gewordenen Messintentionen für die griechische Anwendung des DTK. Die aufgenommene Verteilungsform der Gesamtpunkte aus Skala 1+2 belegt die Unversehrtheit der Messintentionen nach der Transformation der Items aus einer deutschsprachigen in eine griechische Fassung. Diese Transponierbarkeit ist nicht selbstverständlich, bestenfalls ist sie erwartbar, aber immer muss sie sich dann in der Anwendung empirisch beweisen. Die Abweichung der analysierten Verteilung Skala 1+2 von der Normalverteilung ist ein Indikator für den Erhalt der Messintentionen des DTK in seiner griechischen Anwendung.

Bortz und Döring differenzieren, ähnlich wie Lienert und Raatz (s.o.), zwischen Merkmalen, die von der Theorie her normalverteilt oder eher nicht so verteilt sein sollten. „Ist die Nicht-Normalverteilung der Testwerte theoriekonform, kann der Test unverändert bleiben. Allerdings muss die statistische Auswertung (z.B. Gruppenvergleiche) auf die Verletzung der Normalverteilungsvoraussetzung abgestimmt werden. Zwei Strategien sind möglich: Entweder man operiert mit größeren Stichproben (ab ca. 30 Untersuchungsobjekten), wodurch sich die Forderung nach normalverteilten Messwerten in der Regel erübrigt (vgl. Bortz, 1999, S.93), oder man verwendet (v.a. bei kleinen Stichproben) statt der „normalen“ (verteilungsgebundenen) statistischen Verfahren die sog. Verteilungsfreien Analysetechniken (vgl. Bortz und Lienert, 1998; Bortz und Döring, 2002, S. 217).

Abbildung 7.4: Verteilung von Gesamtpunktwerten nach Skala „Depressive Belastung“ (DTK-Skala 1+ 2)



Aus der gefundenen Verteilungsform dürfen die folgenden Informationen als Arbeitsergebnisse entnommen werden:

7.41 AS: Innerhalb einer möglichen Gesamtscorebreite der DTK-Skala „Depressive Belastung“ von 0 bis 41 Punkten berichten die an der Untersuchung 2034GR beteiligten Kinder über das subjektiv festgestellte Auftreten von 0 bis 36 DTK-Symptomen. Sie erhalten 0 bis 36 Gesamtpunkte.

7.42 AS: Die für diese Skala gefundene linksasymmetrische Verteilung der Gesamtpunktwerte weist darauf hin, dass die Mehrheit der Kinder vom Auftreten einer jeweils geringeren Zahl solcher Symptome berichten, als sie für den Schwellenwert bestimmt werden, der als Trennwert zwischen depressiv unbelasteten oder weniger belasteten Kindern einerseits und den nach DTK-Messungen klinisch belasteten Kindern benutzt wird. Diese Trennwerte grenzen dann Teilgruppen von Kindern gegeneinander ab, für die, den bekannten Restriktionen für die Aussagemöglichkeiten der Testwerte folgend, Hinweise auf die begründete Möglichkeit des Vorliegens klinisch relevanter Störungen vorliegen.

7.43 AS: Ein sehr niedriger Prozentsatz der beteiligten Schüler und Schülerinnen in Höhe von 1.1% gibt an, dass sie zum Zeitpunkt der Untersuchung überhaupt kein depressives Symptom erlebten.

7.44 AS: Die maximale Menge an den durch die Kinder wählbaren Items ist 42. Ein Großteil der Kinder dieser Stichprobe nutzt diese Breite des Angebotes nicht. Bereits die ersten 50% des Angebots an wählbaren Items, das sind die von 0 bis 21 Items, sind für 91 % der Kinder dieser Stichprobe hinreichend.

7.45 AS: Die zweite Hälfte des Angebotes an wählbaren Items, das sind die Items 22 bis 41 werden nur noch von 9% der Kinder dieser Stichprobe in Anspruch genommen.

7.46 AS: Das Itemangebot in der DTK-Skala 1+2 ist für die untersuchte Stichprobe hinreichend, weil es durch die Wahlen nicht ausgeschöpft wird. Beansprucht werden vom Punktwert 0 bis 36 durchgehend nur 88,1% der Wahlmöglichkeiten.

7.47 AS: Durch die Zusammenlegung der Skalen 1 und 2 zur Skala „Depressive Belastung“ wird ersichtlich, dass der durch die Addition der wählbaren Items entstehende Entscheidungsraum als Angebot größer ist als die dann realisierte Nachfrage. Die Divergenz zwischen Angebot und Nachfrage ist aber größer als bei jeder Einzelskala im DTK. Das weist darauf hin, dass die extensive Nutzung des Entscheidungsraumes auf Skala 1 seltener mit einer extensiven Nutzung auf Skala 2 gekoppelt ist. Das muss so sein, weil sonst die höheren Gesamtpunktwerte auf Skala 1+2 auch mit höheren Frequenzen besetzt wären.

7.48 AS: Die 7.47 AS ist ein Argument für die konstruktsspezifische Messung der DTK-Skalen 1 und 2.

Aus der vorangegangenen Prüfung der Häufigkeitsverteilung für die Gesamtpunktwerte aus Skala 1+2 „Depressive Belastung“ ergab sich die noch sehr allgemeine Information, dass, insoweit diese Gesamtpunktwerte verlässliche Hinweise auf unterschiedliche Belastungen durch depressive Symptome zu geben vermögen, wovon hier ausgegangen wird, dass die Kinder der untersuchten Stichprobe in unterschiedlicher Weise von depressiven Symptomen belastet erscheinen. Wenn für die untersuchte Skala Ordinalskalenniveau angenommen wird, was hier als berechtigt gilt, ist sicher, dass 91% der griechischen Kinder als relativ gering belastet gelten dürfen als die bei den höchsten GPW aufzufindenden weiteren 9% von ihnen. Ungewiss bleibt, wie groß die Distanz der Ausprägung von Belastung zwischen den unteren 91% und der oberen Teilgruppe aus 9% der Kinder dieser Stichprobe ist, weil auf Ordinalskalen keine sicheren Informationen über die Abstände zwischen den Rangplätzen vorliegen. Immerhin sollte die nachfolgende Aussage als empirisch gerechtfertigt gelten dürfen:

„Die Messwerte aus der Anwendung des DTK in einer Stichprobe von 2005 (Missing 29) griechischen Kindern geben Hinweise darauf, dass etwa 9% von ihnen in deutlich stärkerem Maße mit depressiven Symptomen belastet sein dürften als alle anderen Kinder. Das sind zumindest etwa 180 von 2005 Kindern. Diese Aussage beruht nicht auf Trennwerten, sondern auf der prozentualen Nutzung des Entscheidungsraumes an DTK-GPW (Skala 1+2) links und rechts vom Median der benutzten GPW.“

Das Verfahren DTK wurde bei der Durchführung der Untersuchung 2034GR als ein zur Erfassung von Hinweisen auf das Bestehen depressiver Belastungen verlässliches Instrument eingesetzt. Demnach sind mit dem DTK auch keine Diagnosen zu erstellen. So ist denn auch der eben zuvor angegebene Prozentsatz keine Prävalenzrate. Er gibt Menge von Kindern an, für die Hinweise der genannten Art auf das Bestehen depressiver Belastungen vorliegen.

Die dargelegte Schätzung basiert zum einen auf der zumindest für deutschsprachige Untersuchungen empirisch belegten Annahme, dass DTK Gesamtpunktwerte und depressive Belastungen in gleichsinnig gepolten korrelativen Beziehungen stehen (s. Rossmann, 2005). Weiterhin nutzen sie die empirisch belegte Auffassung, dass Anwachsen des numerischen Wertes

von DTK Gesamtpunkten korrespondiere auch mit dem Anwachsen der Ausprägung depressiver Belastungen. Beides reicht nur, um eine oben dargestellte Form der Schätzung abzugeben.

Ungewiss bleibt zunächst, welchen Stärkegrad die jeweils gemeinte Ausprägung hat. Sie könnte beispielsweise subklinisch oder aber klinisch sein. In beiden Fällen unterrichteten die mitzuteilenden Informationen über das Bestehen von Handlungsbedarf, der für jeden, der um die gesunde und optimale Entwicklung von Kindern bemüht sein möchte oder dazu gesellschaftlich beauftragt ist, in seinem Problemgehalt erkennbar ist. Die Information verweist auf einen sehr wahrscheinlich bestehenden Notstand. Ein sehr wahrscheinlich bestehender Notstand sollte jeden, der nicht mutwillig andere oder sich selbst gefährden wollte, zumindest zu einer umgehenden und sorgsam Überprüfung der gegebenen Situation veranlassen. Insoweit ist die mit dem Befund aus dieser Analyse abzugebende Schätzung eine bereits hinreichend ernst zu nehmende Warnung. Nachfolgend wird versucht, die abgegebene Schätzung nach Möglichkeit zu verbessern.

Die Realisierung eines solchen Bemühens ist auf die Verfügbarkeit von weiteren und hier bisher noch ungenutzten Informationen über Zusammenhänge zwischen individuell variierenden Gesamtpunktwerten im DTK und individueller Belastung durch Symptome der Depression angewiesen. Allgemein angenommen wird, dass höhere Gesamtpunktwerte auf schwerere depressive Belastungen hinweisen. Unterstellt wird somit also auch, ein Kind, das nur eine Frage im DTK als für sich zutreffen bejaht, werde weniger depressiv belastet sein als Kinder, die das Auftreten von vielleicht 10 oder noch mehr Symptomen als für sich selbst zutreffend angeben. Die Angaben eines mit 12 Gesamtpunktwerten bekannt werdenden Kindes sind in dieser Aussagenlogik dann wiederum als Indizien für eine relativ leichte Belastung einzustufen, wenn ein Kind zum Vergleich heran gezogen wird, das einen DTK-Gesamtwert von 30 Punkten erreichte. Um eine realistische Vergleichbarkeit von DTK-Gesamtpunktwerten erreichen zu können, müsste auf der vorhandenen Skala von individuell erreichten Testwerten zumindest ein GPW bekannt werden, der insoweit als ein kritischer Punkt zu betrachten wäre, als er die Grenze zwischen alltäglich auftretenden depressiven Episoden und klinischen Ausprägungen von Depression markierte. Mit einer solchen Erwartung wird ein Punktwert beschrieben, der in der hier schon zuvor durchgeführten Analysen zu den Gesamtpunktwerten der DTK-Skalen 1, 2 und 3 als Trennwert oder Cut off bezeichnet und verwendet worden ist. Ein solcher Trennwert kommt dem Bedürfnis nach, irgendwie zwischen dem einen und dem anderen Sachverhalt unterscheiden zu können. Der Trennwert ist auch eine Antwort auf die Frage nach der Grenze zwischen qualitativ unterschiedlichen Bereichen. Im gegenwärtigen europäischen Alltag ist das Verfallsdatum von Lebensmitteln ein viel beachteter Trennwert, aber auch einer, der in der Praxis mit all den Problemen aufwartet, die Trennwerten anscheinend zukommen, so auch mit dem der fraglichen Vertrauenswürdigkeit des zu beachtenden Wertes. In der einschlägigen psychologischen Literatur zum Umgang mit Prävalenzen und Prävalenzraten beklagen Verhulst und Koot (1992, S. 33) das Fehlen eines Standards, den sie „gold standard“ nennen, der bei der Identifizierung von Persönlichkeiten mit und ohne klinischen Ausprägungen einer Störungen durch seine verlässliche Präzision hilfreich wäre. Nicht so sehr die Unterscheidung von unterschiedlichen Ausprägungen scheint das schwerwiegendere Problem zu sein, sondern mehr noch, welche Ausprägungen als klinisch relevant zu betrachten seien.

Trennpunkte zu setzen ist sehr wahrscheinlich leichter, als sie zwingen zu begründen. So bleibt problematisch, sie gegenüber dem Vorwurf von Beliebigkeit zu verteidigen. Übereinkünfte könnten erreichen, kritische Trennwerte da zu setzen, wo 10% einer untersuchten Stichprobe sich von den anderen 90% absetzen. Methodisch und wissenschaftstheoretisch sind solche und ähnliche Setzungen überaus unbefriedigend. Schließlich lassen sich Stichproben mit klinisch relevanten Konstrukten denken, bei denen die gesamte Verteilung der Ausprägungen klinisch unbedenklich wäre. Nach einer anderen Verfahrensweise, eine, die in der Literatur durchaus

verbreitet ist, werden Cut-offs über vorab definierte Standardabweichung gesetzt. Eine weitere Möglichkeit der empirischen Begründung von Trennwerten ergibt sich über die Durchführung von Untersuchungen, bei denen den Testverfahren gewissermaßen die Aufgabe gestellt wird, mit ihren Testwerten in diagnostisch untersuchten Stichproben, möglichst fehlerfrei zwischen Persönlichkeiten mit und ohne klinisch relevanten diagnostischen Befund zu unterscheiden. Die nachgewiesenen Trefferquoten bieten Informationen über eine zweckmäßige Positionierung von Trennwerten. Diese Vorgehensweise hat sich der Kritik zu stellen, dass empirisch erfassbare Grenzen zwischen klinischen und subklinischen Ausprägungen von Störungen nicht nur verfahrensabhängig sondern auch stichprobenabhängig seien. Der in der einen Stichprobe gültige Trennwert müsse nicht zwangsläufig für eine andere Stichprobe gelten.

Immerhin besteht aber ein als allgemein zu bezeichnendes Interesse daran, über Messinstrumente schnell und verlässlich das Ausmaß von Gefährdungen erfassen zu können, am besten noch rechtzeitig vor dem Eintreffen von Ereignissen, die durch Diagnosen nicht mehr zu verhindern sind.

Für die vorliegende Arbeit kann ein Vorteil darin gesehen werden, dass empirische Angaben über die Diskriminierungsfähigkeit der eingesetzten Verfahren aus Studien an klinischen Stichproben von Schulkindern vorliegen. Durch eine Übernahme der dort gefundenen Grenzwerte für die Identifizierung von wahrscheinlich mit klinischer Ausprägung depressiv gestörten Kinder könnten zum einen Trennwerte nach Algorithmen gesetzt werden, die sich schon anderswo empirisch bewährt hätten. Andererseits ist dann auch empirisch gestützt abzuschätzen, mit welcher Wahrscheinlichkeit danach ausgesprochene Zuweisungen fehlerfrei oder fehlerhaft sein werden. Gesucht werden Trennwerte, die helfen, mit klinischer Ausprägung depressiv belastete griechische Schul Kinder in der untersuchten Stichprobe möglichst fehlerfrei aufzufinden. Unberührt von all solchen Zielvorstellungen bleibt erneut festzustellen, dass solche Trennwerte zum Zeitpunkt der Durchführung dieser Analyse nicht bekannt sind.

Empirisch fundierte Empfehlungen zur Bestimmung depressiver Störungen mit klinischer Ausprägung unter Anwendung des DTK sind, wie hier zuvor mehrfach berichtet, bislang nur in einer deutschsprachigen Studie von Pötz zu finden (s. Rossmann, 2005). Pötz (2003) untersuchte eine Stichprobe von Kindern mit einer klinischen Diagnose von Depression und eine Stichprobe von Schulkindern mit dem Ziel, Cut-off Werte im DTK zu finden, mit denen die als depressiv diagnostizierte Teilgruppe effektiv von den anderen unterschieden werden könnte. Die nachstehende Tabelle stellt die aus der Studie von Pötz bekannt gewordenen DTK-Grenzwerte dar, die im Hinblick auf ihre jeweilige Sensitivität und Spezifität als effizient eingestuft werden konnten.

Tabelle 7.21: DTK Gesamtpunktwerte als Cutoff-Werte zur Identifizierung diagnostizierter depressiver Kinder (aus Pötz, 2003, S. 74)

	Sensitivität	Spezifität	Overall Accuracy	+ prädiktiver Wert	- prädiktiver Wert	Kappa Index
Skala 1 + 2 \geq 17 oder Skala 3 = 7	81,82	66,67	74,25	0,71	0,79	0,48
Skala 1 + 2 \geq 18 oder Skala 3 = 7	75,76	72,73	74,25	0,74	0,75	0,48

Die zwei Kombinationen sind im Hinblick auf ihre Kenngrößen vergleichbar. Die Wahl des einen oder des anderen Kennwertes führt indessen nicht zu durchgängig identischen Resultaten. Die

Sensitivität bei der Identifizierung der klinisch diagnostizierten Gruppe führt über den Trennwert GPW 17 auf Skala 1+2 zur Bildung einer Teilgruppe, in deren mit 100% dargestellten Gesamtmenge an Kindern 81,82% der Teilnehmer aus der Diagnosegruppe kommen, die mit einer klinischen Ausprägung der depressiver Symptomatik belastet sind. Das sollte nicht mit einem prozentualen Anteil an der Diagnosegruppe selbst verwechselt werden. In der Gegengruppe aus Kindern mit GPW unterhalb des Trennpunktes ist der Anteil der falsch zugewiesenen Kinder, weil für sie doch eine Diagnose über eine depressive Störung vorliegt, insgesamt 33,33%. Die werden nun, weil nicht von den anderen der Teilgruppe trennbar, irrtümlich als klinisch unbelastet bezeichnet. Wird der Trennwert 18 gewählt, hat die abgetrennte und als im klinischen Sinn belastet geltende Teilgruppe (100%) nur noch 75,76% korrekt identifizierter Mitglieder in sich. Dafür ist in der als unbelastet geltenden Gegengruppe der Anteil der, weil doch belastet und somit falsch zugewiesenen Teilnehmer mit diagnostizierten klinischen Ausprägungen depressiver Störungen auf 27,27% gesunken. Eine fehlerfreie Trennung der Teilgruppen ist demnach weder mit dem einen noch mit dem anderen Trennwert erreichbar.

In Anwendung des von Pötz identifizierten Trennwertes 18 umfasst die so bestimmte Teilmenge der Kinder alle jene, die in der Verteilung der Gesamtpunktwerte aus DTK-Skala 1+2 „Depressive Belastung“ entweder einen GPW von 18 oder höher erreicht haben. Das sind, wie die oben angeführte Tabelle der Gesamtpunktwerte aus DTK-Skala 1+2 ausweist, 446 Kinder, die als identifizierte Teilgruppe 22,2% der gesamten untersuchten Stichprobe einschließen.

Von dieser Teilgruppe mutmaßlich in klinischem Ausmaß belasteter Kinder wird gewusst, dass sie unter Anwendung der von Pötz ermittelten Kennwerte 27,27% fälschlich in die klinische Gruppe einbezogene Kinder umfasst.

In einer Erweiterung des von Pötz vorgeschlagenen Trennverfahrens für Skala1+2-GPW =18 lässt sich die weiter oben wiedergegebenen Verteilung der Gesamtpunktwerte aus Skala 1+2 in drei Teilgruppen mit mutmaßlich unterschiedlicher depressiver Belastung aufteilen. Diese Vorgehensweise wird durch die Beobachtung angeregt, dass der Anteil der irrtümlich als unbelastet bezeichneten und also doch klinisch belasteten Kinder mit Annäherung an GPW = 0 sinkt.

Die Setzung von kritischen Werten als Indikatoren für den Belastungsgrad depressiver Verstimmungen nach DTK ermöglicht die Einteilung der untersuchten Kinder, je nach ihren individuellen DTK-Gesamtwerten, auf drei Gruppen mit mutmaßlich unterschiedlichen Schweregraden in der Ausprägung depressiver Störungen. Der Vorteil dieser Einteilungen ist der ihnen jeweils zukommende Grad an Wahrscheinlichkeit für die Korrektheit der Zuweisungen. Mit Trennwert 18 nach DTK-Skala 1+2 sollten nach Pötz etwa 75% der Teilnehmer treffsicher zugeordnet worden sein. Das wäre weniger als für diagnostische Zwecke verlangt, vermittelte aber mehr Aufschlüsse als bislang zugänglich gewesen wären. Die müssten mit dem Wissen verbunden bleiben, dass jede vierte Zuweisung eines Kindes zu einer der beiden links und rechts des Trennwertes nach Pötz beschriebenen Gruppen falsch gewesen sein dürfte. Der Konjunktiv ist zu benutzen, weil selbst diese Angabe auf Schätzungen beruht.

Dieser Sachlage folgend wird dann die Frage nach der Häufigkeit depressiver Störungen in der untersuchten Stichprobe 2034GR beantwortet. Die Antwort beginnt mit dem ausdrücklich wiederholten Verweis auf den gegenwärtig relativ unsicheren Wissensstand. Über den Einsatz des DTK in der untersuchten Stichprobe werden empirisch begründete Abschätzungen des mutmaßlichen Anteils einer Teilgruppe depressiv gestörter Schulkinder an dem Umfang der untersuchten Stichprobe mitgeteilt. Die Fehlerquote dieser Einschätzungen könnte insgesamt bei 25% liegen. Insgesamt bedeutet in diesem Zusammenhang, dass dieser Prozentsatz an

fehlerhaften Zuweisungen in einem nicht genau bekannten Verhältnis auf beide voneinander unterschiedenen Teilgruppen der Stichprobe zu verteilen wäre. Das gilt für die Aufteilung in zwei Teilgruppen, einer von Kindern mit wahrscheinlich klinischen Ausprägungen depressiver Störungen (KL) und einer Gegengruppe von nicht in dieser Weise betroffenen Kindern (nKL). Die untersuchte Stichprobe (ST) wird dabei am Trennwert in zwei Teilgruppen geteilt ($ST = KL + nKL$). Zuvor war hier begründet worden, dass die nKL Teilgruppe nochmals in zwei Untergruppen geteilt werden könnte, in eine mit gehobenen (gnKL) und eine andere mit niedrigen GPW (nnKL). Die Teilgruppe nKL hätte dann zwei Untergruppen ($nKL = gnKL + nnKL$).

Die Tabelle 7.22 zeigt die kritischen Grenzwertwerte, die für die Zuordnung der Kinder in drei Gruppen mit mutmaßlich unterschiedlicher Ausprägung der persönlichen Betroffenheit durch das Auftreten depressiver Symptome. Der obere Trennwert GPW 18 ist der von Pötz empfohlene. Der Trennwert zwischen den beiden anderen Gruppen trennt den vom Trennwert auf 0 zugehenden Skalenbereich in zwei gleiche Teilstücke von Gesamtpunktwerten. Dieser Trennwert ist willkürlich gesetzt.

Tabelle 7.22: Kritische DTK-Trennpunktwerte für die Zuordnung der Kinder zu drei Belastungsgruppen

„depBel“- GPW	Belastungsgruppe	Bezeichnung
0 – 8	wahrscheinlich keine bis leichte depressive Belastung	depBel 1
9 – 17	wahrscheinlich moderate depressive Belastung	depBel 2
18 – 41	wahrscheinlich klinisch ausgeprägte depressive Störung	depBel 3

Zeichenerklärung: „depBel“ – GPW: Gesamtpunktwerte in der DTK-Skala „depressive Belastung“

Tabelle 7.23 stellt die nach den in der obigen Tabelle genannten Trennwerten vorgenommene Verteilung der Schulkinder dar, die alle Fragen der DTK-Skalen 1+2 („depressive Belastung“) beantwortet haben (N=2005). Das führt zur Unterscheidung der Besetzungen für die Belastungsgruppen depBel 1, depBel 2 und depBel 3.

Tabelle 7.23: Verteilung von n=2005 Schulkinder auf nach DTK-GPW in Skala 1+2 unterschiedenen Belastungsgruppen

Belastungsgruppe	„depBel“- GPW	Schülerzahl	Prozentsatz
depBel 1	0 – 8	684	34,1%
depBel 2	9 – 17	875	43,6%
depBel 3	18 – 41	446	22,2%

Zeichenerklärung: „depBel“ – GPW: Gesamtpunktwerte in der DTK-Skala „depressive Belastung“

Belastungen durch depressive Symptome für die Kinder unserer Stichprobe zu Arbeitsergebnissen formuliert werden, sollte überprüft worden sein, ob die Manipulation der Gesamtpunktwerte durch die eingesetzten Cutoff-Scores zu einer durch zufällig variierende Mengen an Kindern geführt hat. Die Verteilung der GPW aus den Antworten auf die DTK-Fragen muss nicht von Zufallserwartungen abweichen, Die Verteilung der Kinder auf die Belastungsgruppen depBel 1, depBel 2 und depBel 3 könnte zufällig variierende Mengen erbracht haben. Um die

aufgeworfenen Fragen beantworten zu können, wird die Mengenverteilung auf Signifikanz geprüft.

Die Prüfung mit einem nonparametrischen Verfahren, χ^2 Test, ergibt, dass die Wahrscheinlichkeit für eine Ähnlichkeit zwischen der beobachteten und im Gegensatz dazu theoretisch erwarteten Gleichverteilung der untersuchten Anzahlen von Kinder auf die drei Gruppen (depBel 1,2,3) $p = 0,000$ ist. Demnach dürfen die in diesem Abschnitt mitzuteilenden Werte signifikant eingestuft werden:

7.49 AS: Aus der Anwendung des DTK in der Untersuchung 2034GR hat sich ergeben, dass etwa 22% der untersuchten griechischen Schulkinder mutmaßlich mit klinischer Ausprägung vom Auftreten depressiver Symptome betroffen sind, etwa 44% mutmaßlich moderate Belastungen durch solche Symptome aufweisen und etwa 34% als nur von vereinzelt auftretenden Symptomen betroffen zu betrachten wären.

Die hier zuvor bei der Analyse der Gesamtpunktwerte auf DKT-Skala 1 eingeführte Bestimmung eines verteilungsinternen Trennwertes war dort wie folgt beschrieben worden: „Als Trennwert (cut off) für die Abgrenzung der Untersuchungsteilnehmer mit klinisch relevanter Ausprägung des skalenspezifischen Symptomkomplexes nach Skala 1 des DTK (Konstrukt 1) von anderen Teilnehmern mit niedrigerer Zahl an Selbstzuweisungen wird der Gesamtpunkt看wert der Skala 1 bestimmt, der erstmals nach dem GPW = 0 unter dessen prozentuaem Anteil an der gesamten Stichprobe von Kindern und Jugendlichen bleibt. (21. AS).“

Die Anwendung dieses Trennwertes führt zu der nachstehenden Feststellung.

7.50 AS: Die Messwerte auf DTK-Skala 1+2 aus der Durchführung des DTK in einer Stichprobe von 2034 griechischen Kindern geben unter Bestimmung eines von Kampisiou definierten verteilungsinternen Trennwertes nach Frequenzunterschreitung von GPW=0 Hinweise auf das mutmaßliche Bestehen von depressiven Störungen bei etwa 126 (126,38) Kinder oder 6,3% der untersuchten Stichprobe.

Das Kapitel 7 wird von Ausführungen zur Analyse der Gesamtpunktwerte aus einer ersten griechischen Anwendung des DTK in einer Stichprobe von 2034 Schulkindern gefüllt. Der DTK sollte helfen, Informationen über das Auftreten depressiver Symptome bei griechischen Schulkindern zu sammeln. Anzunehmen ist, dass depressive Symptome bei so gut wie allen Schulkindern, so auch bei den griechischen Kindern der Stichprobe 2034GR auftreten werden. Begründet wird dies mit der Annahme einer alltäglichen Präsenz einzelner depressiven Symptome. Sie gehören zur psychischen Normalität. Das vermehrte und fortwährende Auftreten solcher Symptome sind aber Indizien für das Bestehen einer depressiven Symptomatik. Die kann als depressive Störung und als solche, wie jedes einzelne Symptom auch, mit unterschiedlichen Ausprägungen bestehen. Vor diesem Hintergrundwissen sind Informationen über die Existenz und die Frequenz depressiver Symptome unter Schulkindern insbesondere im Zusammenhang mit Bemühungen um deren möglichst störungsfreies Aufwachsen von Bedeutung. Das Forschungsinteresse der hier erörterten Untersuchung 2034GR ist vor allem auf die Beantwortung der Frage gerichtet, ob depressive Störungen unter griechischen Schulkindern in einem zu Interventionen drängendem Ausmaß auftreten. Dazu könnten stark verallgemeinernd, aber vielleicht auch deswegen zum Gewinnen eines die Sachlage präsentierenden Überblicks nutzbar, drei unterschiedliche Kategorien von Informationen erwünscht sein. Das wären vielleicht 1. Informationen darüber, bei wie vielen Schulkindern depressive Symptome in einem die Normalität nicht verlassenden Ausmaß auftreten, 2. wie viele demgegenüber von solchen Symptomen in einem Ausmaß betroffen sind, das die Normalität bereits verlassen hat und als

Risiko für persönliche Entwicklung und Gesundheit betrachtet werden müsste. Schließlich sollten 3. Informationen darüber verfügbar sein, wie viele Schulkinder von klinisch ausgeprägten depressiven Störungen betroffen sind. Diese Informationen werden, um angemessen handeln zu können, auf eine Matrix verteilt benötigt, die sowohl die Frage nach der Anzahl der Personen als auch die nach den Personen selbst beantwortet. Hilfe kann nicht für Prävalenzen und Prävalenzraten, wohl aber für Menschen und im gegebenen Beispielfall für Kinder gegeben werden. Dazu müssen sie namentlich bekannt sein.

Die mit der Anwendung des DTK zu gewinnenden Gesamtpunktwerte tragen Informationen zu jeder der genannten Fragen bei, die über dem für Griechenland bislang verfügbaren Wissensstand liegen.

Das trifft zu, obschon die so erarbeiteten Daten eine relativ hohe Ungenauigkeit aufweisen.

Dazu trägt das Messverfahren selbst auch bei, weil es Selbstzuschreibungen verwendet.

Die Informationen sind den Gesamtpunktwerten durch das Antwortverhalten der Schulkinder gegeben worden, die das Beantworten der Testfragen jeweils als eine persönliche Aufgabe übernommen haben. Welche persönlichkeitspsychologische und welche depressionsspezifische Bedeutung dem unterschiedlichen Antwortverhalten zukommt, wird durch die Analyse der mit einer entsprechenden Untersuchung gewonnenen Daten bislang nur unzureichend bestimmt. Die individuellen Gesamtpunktwerte auf den Skalen des DTK und deren gruppenbezogene Verteilung sind ein Teil der jeweils gewonnenen Datenbasis. Die Gesamtpunktwerte sind ein kategorial reduzierter Aspekt des Antwortverhaltens der beteiligten Schulkinder. Gesamtpunktwerte tragen explizit die Informationen wie viele Ja-Antworten das jeweils beachtete Kind aus der untersuchten Stichprobe gegeben hat. Die Gesamtpunktwerte erlauben auch das Aufnehmen von Informationen darüber, welche Ja-Antworten auf welche Fragen der Itemliste einer Skala das Gesamtergebnis konstituiert hatten.

Die Gesamtpunktwerte sind absolut valide dafür, dass eine bestimmte an der Untersuchung beteiligte Persönlichkeit eine ihr Antwortverhalten kennzeichnende Menge an depressiven Symptomen aus der angebotenen Liste auf einer DTK-Skala ausgewählt hat.

Alle weitergehenden Aussagen über den Informationsgehalt der Gesamtpunktwerte sind bereits von Ungewissheiten belastet, die dazu nötigen, Wahrscheinlichkeiten für den Erklärungswert der einen oder anderen Folgerung zu ermitteln und zu beachten.

Unterstellt wird, dass mit der Ja-Antwort ebenso wie mit einer Nein-Antwort bei der individuellen Bearbeitung des DTK das jeweilige Kind mit der Antwort mehr oder weniger valide mitteilt, ob das jeweils gelesene Item nach der Selbsterfahrung des antwortenden Kindes verbal ein Phänomen repräsentiert, das von diesem Kind an sich selbst erfahren worden ist. Unterstellt wird, das antwortende Schulkind zeige nicht nur Antwortverhalten, sondern gebe damit auch ein valides Urteil ab, nämlich das, was hier zuvor beschrieben wurde. Ob dies wirklich so von jedem der an einer Untersuchung teilnehmenden Kinder realisiert worden ist, vermag im Anschluss an die Durchführung des DTK niemand mit Sicherheit zu bestimmen.

Die Analyse von Testdaten ist häufig, so auch beim DTK, an die Übernahme der Aufgabe gebunden, aus vielseitig interpretierbaren Fakten möglichst eindeutige Befunde herzustellen.

Da die Analyse von Testdaten weiterhin darauf abzielt, die Bedeutung der vorliegenden Daten zu benennen und zu erläutern, ist hinzunehmen, dass der Erklärungswert der Mehrheit aller vorgelegten Befunde jeweils nur als mehr oder weniger wahrscheinlich einzuordnen ist. Die wissenschaftliche Praxis ist darauf eingestellt und weiß darauf angemessen einzugehen.

Schwieriger ist die Einschätzung der Bedeutung der mitzuteilenden Befunde für alle diejenigen, die zur Optimierung ihrer Praxis daran interessiert sind, durch wissenschaftliche Erkundungen auf

entstehenden Handlungsbedarf in ihrem eigenen Praxisfeld, etwa im Schulwesen, in der Sozialarbeit oder im Gesundheitswesen aufmerksam gemacht zu werden. Um eine dann nützlich werdende Markierung bieten zu können, werden die nachfolgend mitzuteilenden Befunde mit verbalen Hinweisen auf den relativen Grad ihrer Verlässlichkeit für die Praxis versehen. Obschon solche Angaben, die als Wahrscheinlichkeiten für den Erklärungswert der Aussagen zu verstehen sind, nicht eindeutig zu bestimmen waren, haben sie nachweisbaren Orientierungswert.

Unterschieden werden die nachfolgenden Hinweise. „Mit an Sicherheit grenzender Wahrscheinlichkeit“ ist der hier nachfolgend benutzte Ausdruck für eine sehr hohe Wahrscheinlichkeit. Die extrem geringe Gegenwahrscheinlichkeit dürfte kaum bemerkenswerte Einbußen des Erklärungswertes bewirken. „Sehr hohe Wahrscheinlichkeit“ soll bedeuten, dass eine Irrtumswahrscheinlichkeit von vielleicht 10% eingeräumt werden muss. „Hohe Wahrscheinlichkeit“ ist die Verbalisierung für Irrtumswahrscheinlichkeiten von bis zu 25%. „Wahrscheinlich“ ist dann die kategoriale Einstufung für Ereignisse, deren Gegenwahrscheinlichkeit bis um 33% liegen könnte. Alle diese Angaben haben begrenzten Orientierungswert, dies vor allem, weil sie selbst auf Daten beruhen, die aus Stichproben und Teilstichproben außerhalb des Praxisfeldes kommen, auf das sie jeweils angewendet werden sollen. Wenn dann der Wert solcher Hinweise selbst als problematisch gelten darf, dann ist dies aus statistischen Gründen gerechtfertigt. Ziemlich eindeutig sind demgegenüber aber die mit den genannten Angaben als notwendig anzusehenden Handlungskonsequenzen für die klinische, die pädagogische und die gesellschaftliche Praxis mit Kindern und Jugendlichen.

Aussagen, denen „sehr hohe Wahrscheinlichkeit“ oder „hohe Wahrscheinlichkeit“ zuerkannt wird, sollten, wenn Konsequenzen für das eigene Praxisfeld anstehen, vorrangig zu sofortigen schnellen Interventionen veranlassen, die 1. die Fakten im eigenen Praxisfeld empirisch überprüfen helfen und 2. vorsorglich bereits spezifische präventive und kurative Maßnahmen bereit stellen und für die Anwendung vorbereiten.

Als „wahrscheinlich“ klassifizierte Aussagen sollten, wenn Konsequenzen für das eigene Praxisfeld anstehen, 1. veranlassen, die Fakten im eigenen Praxisfeld und 2. die aktuell gegebene Handlungsfähigkeit im Hinblick auf etwa notwendig werdende präventive oder kurative Interventionen zu prüfen.

Alle Aussagen sollten als Hinweise auf die Notwendigkeit verstanden werden, die fortlaufende Sensitivität der Praxisarbeit gegenüber dem sehr wahrscheinlichen Auftreten von individuellen Fällen der Belastung durch depressive Störungen und die auf solche Fälle eingestellte Handlungsfähigkeit zu sichern. Weiterhin ist fortlaufend die Prävention gegenüber psychischen Störungen und auch gegenüber Depressivität speziell zu organisieren, zu realisieren und zu optimieren. Das potentielle Auftreten von Depressivität und anderen psychischen Störungen ist ein unabweisbarer Anlass für spezifische soziale und gesellschaftliche Maßnahmen zum Schutz von Kindern und Jugendlichen. Wo dieser Inhalt übergangen werden sollte, wird er sich durch die Faktizität der Ereignisse melden, die dann als individuell zu ertragende Belastung und als persönliches Leid von betroffenen Kindern, Jugendlichen und ihren Angehörigen erkennbar werden.

Die durchgeführte Analyse der Items und der Gesamtpunktwerte führte, wie in den Kapiteln 6 und 7 der Erörterungen im Einzelnen belegt, zu den nachstehenden Befunden.

1. Der DTK hat in seiner griechischen Anwendung innerhalb der Untersuchung 2034GR seine aus deutschsprachigen Anwendungen bekannten Messqualitäten ungemindert erhalten.

2. Mit der Untersuchung 2034GR wird eindeutig nachgewiesen, dass Selbstzuschreibungen depressiver Symptome durch griechische Schulkinder individuell zwischen keinem Symptom und angebotenen Höchstwerten variieren.
 - 2.1 Eindeutig nachgewiesen wurde, dass eine Teilgruppe von Schulkindern eine gegenüber wenig belasteten Kindern fast fünffach höhere Rate an Selbstzuweisungen depressionsspezifischer Symptome an Dysphorie und Selbstwertproblemen trägt. Diese Teilgruppe schließt 14% der Schulkinder der untersuchten Stichprobe von 2034 Kindern ein.
 - 2.2 Eindeutig nachgewiesen worden ist, dass eine Teilgruppe griechischer Kinder eine gegenüber den anderen Kindern in der untersuchten Stichprobe von 2034 Schulkindern deutlich höhere persönliche Betroffenheit von Symptomen agitierten Verhaltens berichtet. Diese Teilgruppe umfasst 24% der Kinder in der gesamten Stichprobe.
 - 2.3 Eindeutig ist, dass ein Großteil der untersuchten griechischen Schulkinder sich von Müdigkeit und weiteren körperlichen Reaktionen betroffen fühlt. Eine Teilgruppe von 6,9% ist in dieser Weise stärker als andere beeinträchtigt.
 - 2.4 Mit an Sicherheit grenzender Wahrscheinlichkeit ist ein wesentlicher Teil der unter 2.1 bis 2.3 genannten Beschwerden in eine die weitere Entwicklung dieser Kinder belastende depressive Störung eingebunden.
3. Mit an Sicherheit grenzender Wahrscheinlichkeit verweisen die Ergebnisse der Untersuchung auf die Verbreitung depressiver Störungen in der untersuchten Stichprobe von 2034 griechischen Schulkindern.
4. Nicht eindeutig zu bestimmen ist die nach diesen Hinweisen zu vermutende Häufigkeit klinisch ausgeprägter depressiver Störungen. Nach den aufgenommen Hinweisen sind mutmaßlich zwischen 6,3% und 22,4% der Kinder von depressiven Störungen mit klinischer Ausprägung betroffen.
5. Diese Schätzungen sind mit Fehlern behaftet, die Abweichungen bis zu etwa 15% (plus/minus) als hoch wahrscheinlich gelten lassen.
6. Die unter 1 bis 5 aufgeführten Befunde sind für alle Praxisfelder im Bereich des Gesundheitswesens, der psychosozialen Versorgung, der Schulen, des Erziehungswesens und der gesellschaftlichen Sorge für Kinder und Jugendliche von Bedeutung. Sie sollten als gewichtige Fakten verstanden werden, die darauf drängen, die in der eigenen Verantwortlichkeit stehenden Praxisfelder 1. möglichst umgehend verlässlich zu überprüfen, ob auch dort Hinweise zum Auftreten depressiver Belastungen bei Kindern und Jugendlichen erkennbar sind. Weiterhin sollten 2. bereits vorsorglich bis zum Vorliegen empirischer Erkenntnisse spezifische präventive und kurative Maßnahmen bereit zu stellen und für eine Anwendung vorzubereiten. Schließlich wären 3. in allen in der eigenen Verantwortlichkeit stehenden Arbeitsfeldern beim Auftreten von Hinweisen auf depressive Störungen unverzüglich spezifische Interventionen zu organisieren und zu realisieren. Depressive Störungen sind erhebliche Risikofaktoren für die Entwicklung von Kindern und Jugendlichen und ziehen bei ihrer Entfaltung ungewöhnlich hohe gesellschaftlich zu tragende Folgekosten nach sich (Laux, 2002).
7. Die mitgeteilten Befunde verweisen auf die Notwendigkeit, da, wo dies noch nicht geschehen, weiträumig wirksame Präventionen gegenüber psychischen Störungen und speziell gegenüber Depressivität zu organisieren, zu realisieren und zu optimieren.
8. Die mitgeteilten Befunde verweisen für die wissenschaftliche Arbeit zum Auftreten von depressiven Störungen bei griechischen Schulkindern auf die Unersetzlichkeit und die entsprechende Dringlichkeit der Durchführung von klinischen Studien unter Einbeziehung von Screening-Verfahren.

Kapitel 8

Verteilung der DTK-Gesamtpunktwerte in verschiedenen Teilgruppen der untersuchten Stichprobe 2034GR

Mit Stichproben arbeitende psychologische Untersuchungen betrachten im Allgemeinen Gruppierungen von Persönlichkeiten, die zumindest im Hinblick auf ein Merkmal die für eine Stichprobenbildung notwendige Gemeinsamkeit aufweisen. In der Studie 2034GR wird eine solche für die Bildung einer Stichprobe notwendige Gemeinsamkeit durch das vorab definierte Merkmal der zu einem gegebenen Zeitpunkt bestehenden Zugehörigkeit zu einer der für die Untersuchung ausgewählten griechischen Schulen hergestellt. Alle von einer solchen Stichprobe aufgenommenen empirischen Daten gelten für diese aus n Personen mit Ausprägungen des gemeinsamen Merkmals bestehenden Stichprobe.

Da Persönlichkeiten über dieses Merkmal hinaus durch eine Vielzahl weiterer persönlichkeitspezifischer Daten gekennzeichnet werden können, besteht bei jeder untersuchten Stichprobe die Möglichkeit, sie in Teilstichproben mit zumindest einem in zwei Ausprägungen variierenden persönlichen Merkmal aufzuteilen. In der Regel besteht ein methodisch begründetes Interesse daran. Der in seiner Verteilung homogen erscheinende Datensatz der Stichprobe könnte aus heterogenen und signifikant voneinander abweichenden Verteilungen der Daten in den Teilgruppen der Gesamtstichprobe hervorgegangen sein.

Die mit der Stichprobe 2034GR zusammengefassten Schulkinder sind auf nach Alter, Geschlecht, Wohnort und weiteren Merkmalen unterschiedenen Teilstichproben zu verteilen. Je größer die Zahl der die Persönlichkeiten einer Stichprobe kennzeichnenden und auch bekannt gewordenen Merkmale ist, desto mehr Teilstichproben lassen sich ausdifferenzieren und in ihren Messwerten untereinander vergleichen.

In diesem Sinne könnte sich für die auf depressive Störungen verweisenden DTK Gesamtpunktwerte ergeben, dass sie in ihrem Auftreten zwar die Gesamtstichprobe eindeutig kennzeichnen, aber zugleich dem entgegen stehende Kennwerte für Untergruppen von Jungen und Mädchen innerhalb der untersuchten Stichprobe einschließen.

In der Forschung zum Auftreten von depressiven Störungen bei Kindern und Jugendlichen sind insbesondere Alter und Geschlecht nahezu durchgehend als mögliche Quellen von Datenvariationen betrachtet und entsprechend systematisch untersucht worden. Anzunehmen ist, dass andere Persönlichkeitsmerkmale für die untersuchten Fragen zum Auftreten depressiver Störungen zumindest ebenso beachtenswert sind wie Alter und Geschlecht. Da aber einige die untersuchten Kinder oder Jugendlichen kennzeichnenden Daten leichter als andere zu erheben sind, ergibt sich eine Frequenzliste beachteter Merkmale, die eher der Verfügbarkeit und der auch sozial kontrollierten Erreichbarkeit von Daten folgt als den inzwischen durch Fortschritte in der empirischen Forschung aufzuwerfenden speziellen Forschungsfragen. Mit denen drängte sich beispielsweise auf, eher nach den Ursachen für die mit Alter und Geschlecht variierenden Prävalenzraten als nach deren Variation mit den unterschiedlichen Stichproben zu forschen. Schon der Hinweis auf die dann möglicherweise notwendig werdende Erfassung spezieller psychophysiologischer Daten verweist auf das damit einhergehende Ansteigen des Untersuchungsaufwandes.

8.1 Geschlechts- und Alterseffekte in Punktwerten aus der Anwendung des DTK

In fast allen bisherigen Untersuchungen zum Auftreten depressiver Störungen bei Kindern und Jugendlichen ist geprüft worden, ob die jeweilige Datenlage systematische Geschlechts- und Alterseffekte erkennen lässt. Der über die unterschiedlichen Verfahren hinausreichende übereinstimmende Befund stützt zwei Feststellungen. So kann

1. eine mit dem zunehmenden Alter steigende Prävalenzrate depressiver Störungen bei Kindern und Jugendlichen als gegeben und Stichproben übergreifend angenommen werden. Weiterhin ist im Allgemeinen
2. eine mit dem Alter bei Mädchen stärker als bei Jungen ansteigende Prävalenzrate beobachtet worden.

Diese Annahmen werden auch mit entsprechenden Untersuchungen zum DTK gestützt. Dabei sind zwei in ihrer Aussagekraft unterschiedliche Quellen für solche Belege zu unterscheiden. Sie können zum einen auf Untersuchungen zurückgehen, in denen sowohl DTK Punktwerte als auch Diagnosen depressiver Störungen für die Datenanalysen nutzbar sind. Zum anderen können die behaupteten Unterschiede und Eigenheiten als wahrscheinlich gegeben angesehen werden, wenn für die DTK Punktwerte eine entsprechende Aussagekraft nachzuweisen war.

Zur Kenntnis zu nehmen ist, dass der Griechenland betreffende Forschungsstand zum Auftreten depressiver Störungen bei Kindern und Jugendlichen zumindest bis 2004 weitaus weniger Daten und Ergebnisse umfasst als sie für die in der Forschung führenden Staaten verfügbar sind. In diesem Zusammenhang ist Fragen nach alters- und geschlechtsspezifischen Variationen der Prävalenzen für depressive Störungen unter griechischen Schulkindern eine andere Bedeutung zuzuschreiben, als sie für Staaten gelten würde, für die eine längere psychologische Forschungstradition als die in Griechenland vorfindbare bestünde.

Die Mehrheit der DTK-Studien mit deutschsprachigen Stichproben von Kindern kommt im Hinblick auf Geschlechts- und Alterseffekte bei den drei DTK-Subskalen übereinstimmend zu Ergebnissen, die den aus der Normstichprobe des Verfahrens (Rossmann, 2005) entsprechen.

So ergab sich für die Skala „Dysphorie/Selbstwert“ des DTK einerseits, dass die Mädchen signifikant höhere Mittelwerte an Gesamtpunktwerten (GPW) als ihre Mitschüler erreichen. Andererseits sind für Kinder aus höheren Schulstufen gegenüber jüngeren Kindern auch signifikant höhere Durchschnittswerte (GPW, DTK Skala1) festgestellt worden (Rossmann, 1993; Engel, 1994; Fuchs, 1995; Gratz, 1994; Hafner, 2001; Janisch, 1994; Malovic, 2000; Ofner, 1995; Reitbauer, 1999; Sitzenfrey, 1994; Slezak, 1993).

Abweichend davon fanden allerdings Acherman (1995), Buchmann (2000), Hafner (2001), Kelz (2000), Linsbogen (1996) und Malovic (2000) für die Mittelwerte ihrer Stichproben keine geschlechtsspezifischen Effekte auf den DTK-Skalen.

Mit dem immer wieder neu darzustellenden Widerspruch im aktuellen Forschungsstand über das Auftreten oder Ausbleiben geschlechtsspezifischer Unterschiede in den über Screening-Verfahren für depressive Störungen festgestellten wahrscheinlichen Häufigkeiten sollte nicht nur von Fall zu Fall, sondern eben auch wissenschaftstheoretisch orientiert und insoweit theoriegeleitet verhandelt werden. Das einfache Auszählen von Mehrheiten als Basis für empirisch nachgewiesene oder noch nachzuweisende geschlechtsspezifische Effekte ist weitgehend unbefriedigend. Hier wird davon ausgegangen, dass die erbrachten Nachweise für geschlechtsspezifische Unterschiede das Bestehen dieser Variationsquelle nachweisen. Dementsprechend besteht für das Ausbleiben des Effektes ein spezieller Erklärungsbedarf.

Relativ stabil erweisen sich auch Unterschiede in der Produktion der Gesamtpunktwerte zur Skala „Agitiertes Verhalten“ im Hinblick auf Alter und Geschlecht der Schulkinder in der untersuchten

Stichproben (Achermann, 1995; Engel, 1994; Fuchs, 1995; Gratz, 1994; Hafner, 2001; Linsbogen, 1996; Malovic, 2000; Ofner, 1995; Rossmann, 1993; Sitzenfrey, 1994; Slezak, 1993). Auch auf dieser Skala nehmen die festgestellten Mittelwerte mit dem Alter zu, aber diesmal erreichen die Jungen die signifikant höheren Punktwerte. Als Ausnahmen dazu sind die Ergebnisse der Studien von Buchmann (2000), Janisch (1994), Kelz (2000) und Reitbauer (1999) und zu betrachten, die für die DTK-Skala 2 keinen geschlechtsspezifischen Unterschied fanden.

In Bezug auf die Produktion der Durchschnittswerte zur Skala „Müdigkeit/autonome Reaktionen“ ergab sich aus den Studien von Engel (1994) und von Reitbauer (1999), dass die Mädchen höhere Mittelwerte als die Jungen erreichten. Für die Normstichprobe des Verfahrens (s.o.) waren in allen Altersstufen ebenfalls für die Mädchen geringfügig höhere Durchschnittswerte als für die Jungen mitgeteilt worden. Wegen der geringen Eta-Quadrat Werte war dieser Befund von Rossmann (1993) eher als ein Effekt des großen Umfangs der Stichprobe angesehen worden.

Ein Ausbleiben geschlechtsspezifischer Effekte für die Skala „Müdigkeit/autonome Reaktionen“ erlaubt, die Übereinstimmung der entsprechenden Ergebnisse aus den meisten Replikationsstudien mit dem DTK zu vermerken (Achermann, 1995; Buchmann, 2000; Fuchs, 1995; Gratz, 1994; Hafner, 2001; Janisch, 1994; Kelz, 2000; Linsbogen, 1996; Ofner, 1995; Sitzenfrey, 1994; Slezak, 1993).

8.2 Alters- und Geschlechtseffekte in den DTK-Daten aus der griechischen Untersuchung 2034GR

Nachfolgend wird nun erörtert, welche Effekte dem Alter und dem Geschlecht in der Untersuchung 2034GR zuzuschreiben sind.

Vorab soll auf die Bedeutung möglicherweise auftretender Übereinstimmungen zwischen den Ergebnissen aus der griechischen Untersuchung und denen aus internationalen Studien verwiesen werden. Solche Übereinstimmungen würden hier als Argumente für Gemeinsamkeiten der verwendeten Messinstrumente und damit auch als Indiz für die Qualität der griechischen Version des DTK angesehen werden.

Für die Prüfung der Zusammenhänge wird hier das Verfahren angewendet, mit dem auch Rossmann bei den Analysen der Daten aus der Normstichprobe gearbeitet hat. So wird für jede DTK Skala eine zweifaktorielle Varianzanalyse unter Hinzuziehung des Scheffé-Posttests auf einem Signifikanzniveau von $p < .01$ gerechnet.

8.2.1 Geschlechts- und Alterseffekte in den Gesamtpunktwerten zur Skala „Dysphorie/Selbstwertprobleme“ (DTK-Skala 1) in der Untersuchung 2034GR

Aus der Varianzanalyse mit Daten über die Gesamtpunktwerte zur DTK-Skala 1 ergab sich für die Faktoren Geschlecht und Alter ein signifikanter Effekt des Faktors Geschlecht [$F(1,2000) = 59,114$; $p = 0,000$; Eta-Quadrat = 2,9%]. Aus der Tabelle 1 ist zu erfahren, dass die Mädchen höhere Mittelwerte auf der Skala „Dysphorie/Selbstwert“ als die Jungen erreichen (8,77 und 6,82 entsprechend). Ein signifikanter Haupteffekt ergab sich auch für den Faktor Alter [$F(7,2000) = 9,701$; $p = 0,000$; Eta-Quadrat = 3,3%]. Die erste Spalte der Tabelle 8.1 zeigt, dass mit zunehmendem Alter die für alle Kinder errechneten Mittelwerte der GPW aus der DTK-Skala 1 zunehmen.

Neben den zuvor erwähnten Haupteffekten erbrachte die durchgeführte Varianzanalyse auch eine zweifache Interaktion zwischen Geschlecht und Alter [$F(7,2000) = 5,551$; $p = 0,000$; Eta-Quadrat = 1,9%]. Die Durchführung von t-Tests zur Überprüfung von geschlechtsspezifischen Unterschieden in jeder Altersstufe ergab, dass die 8jährigen und 12-14jährigen Mädchen signifikant höhere Durchschnittswerte als die gleichaltrigen Jungen erreichten (siehe letzte Spalte in Tabelle 8.1).

Die eben mitgeteilten Ergebnisse werden auch über entsprechenden nonparametrischen Prüfungen mittels Mann-Whitney Test nachgewiesen.

Tabelle 8.1: Mittelwerte und Standardabweichung (SA) der DTK-Skala „Dysphorie/ Selbstwertprobleme“ in der Gesamtstichprobe und in geschlechtsspezifischen Altersgruppen (nebst Angaben zum T-Test)

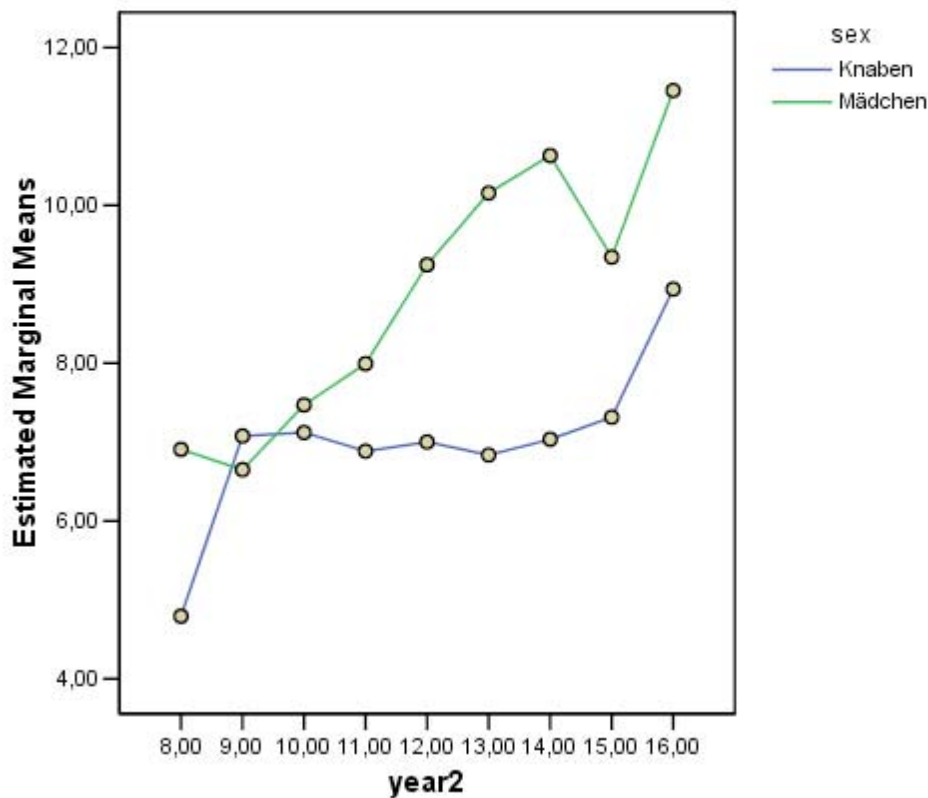
Alter	Gesamtstichprobe			Jungen			Mädchen			T-test	
	N	M	SA	n	M	SA	n	M	SA	t	p
8	171	5,7193	4,25	96	4,7917	3,64	75	6,9067	4,68	-3,324	,001
9	216	6,8565	4,90	105	7,0762	4,98	111	6,6486	4,83	,641	,522
10	257	7,2763	4,89	142	7,1197	4,88	115	7,4696	4,92	-,570	,569
11	284	7,4225	4,81	146	6,8836	4,78	138	7,9928	4,79	-1,953	,052
12	332	8,1566	5,12	161	7,0000	4,53	171	9,2456	5,40	-4,091	,000
13	337	8,4807	4,95	170	6,8353	4,45	167	10,1557	4,88	-6,528	,000
14	322	8,6646	5,25	176	7,0341	4,89	146	10,6301	5,01	-6,498	,000
15+	97	8,7835	5,12	51	7,8235	4,84	46	9,8478	5,25	-1,975	,051
Gesamt	2016	7,7604	5,02	1047	6,8243	4,69	969	8,7719	5,17	-8,877	,000

Zeichenerklärung: n= Zahl der Angehörigen einer Teilstichprobe, M= Mittelwert, SA= Standardabweichung

Abbildung 8.1 veranschaulicht graphisch die vorher dargestellten alters- und geschlechtsspezifischen Unterschiede im Aufkommen von GPW der DTK-Skala 1.

Abbildung 8.1: Score in der DTK-Skala „Dysphorie / Selbstwertprobleme“ in Abhängigkeit von Geschlecht und Alter

Estimated Marginal Means of Dysphorie / Selbstwertprobleme



Die bis hier dargestellten Analyseschritte zu den Zusammenhängen zwischen Gesamtpunktwerten auf den drei DTK-Skalen einerseits und Alter und Geschlecht andererseits führten zu folgenden Arbeitsergebnissen:

8.1 AS: In der Produktion der Mittelwerte zu den GPW der DTK-Skala „Dysphorie/Selbstwert“ wurden Effekte des Alters, des Geschlechts und ihrer Interaktionen gefunden.

8.2 AS: Die aus der griechischen Stichprobe bekannt werdenden GPW zur Skala 1- nehmen in ihren Mittelwerten mit dem Alter der Kinder zu.

8.3 AS: Die Mädchen aus der Gesamtstichprobe erreichen signifikant höhere Mittelwerte an GPW aus der Skala 1 als die Jungen.

8.4 AS: Statistisch signifikante geschlechtsspezifische Unterschiede bestehen wegen höherer Mittelwerte der Mädchen sowohl in der Alterstufe der 8-jährigen als auch ab dem 12. Lebensjahr.

8.5 AS: Die gefundene alters- und geschlechtsspezifische Unterschiede nach GPW zur Skala „Dysphorie/Selbstwert“ stimmen mit den entsprechenden Ergebnissen aus der Normstichprobe des Verfahren wie auch mit den meisten deutschsprachigen DTK-Studien überein.

Spezielle Aufmerksamkeit könnte das Ergebnis auf sich ziehen, nach dem die für die älteren Kinder gefundenen geschlechtsspezifischen Unterschiede auch schon bei den 8jährigen Kindern zu beobachten sind. Aufgeklärt werden sollte, ob dieser Befund replizierbar ist.

8.2.2 Geschlechts- und Alterseffekte in den Gesamtpunktwerten aus der Skala „Agitiertes Verhalten“ (DTK-Skala 2) in der Untersuchung 2034GR

Im Hinblick auf geschlechts- und altersspezifische Effekte zum Aufkommen an Gesamtpunktwerten auf der DTK-Skala „Agitiertes Verhalten“ hat sich über eine entsprechende Varianzanalyse ergeben, dass die Jungen der Stichprobe signifikant höhere Durchschnittswerte an Punkten auf Skala 2 erreichen als die Mädchen [$F(1,2015) = 67,570$; $p = 0,000$; Eta-Quadrat = 3,3%]. Wie aus der nachstehenden Tabelle 2 zu erfahren ist, erreichen die Jungen der Stichprobe in dieser Skala einen Durchschnittswert von 4,99 Punkten, während der entsprechende Wert der Mädchen 3,83 ist. Wie sich auch für die in der DTK-Skala 1 erbrachten Mittelwerte ergeben hat, nehmen die von Schülern in der Skala 2 „Agitiertes Verhalten“ erbrachten Mittelwerte ebenfalls mit ihrem Alter signifikant zu [$F(7,2015) = 32,388$; $p = 0,000$; Eta-Quadrat = 10,2%]. Diese altersabhängige Steigerung der arithmetischen Mittel der Gesamtpunktwerte aus Skala2 ist in der entsprechenden Spalte der Tabelle 2 erkennbar.

Die dann separat vorgenommene Durchführung von t-Tests zu den geschlechtsspezifischen Unterschieden zwischen den Alterstufen hat ergeben, dass die Jungen auf allen Alterstufen signifikant höhere Mittelwerte erreichen als die gleichaltrigen Mädchen (siehe Tabelle 2). Eine Ausnahme davon stellen nur die Mittelwerte der 8jährigen Mädchen und Jungen dar, die sich voneinander nur gering und nicht signifikant unterscheiden.

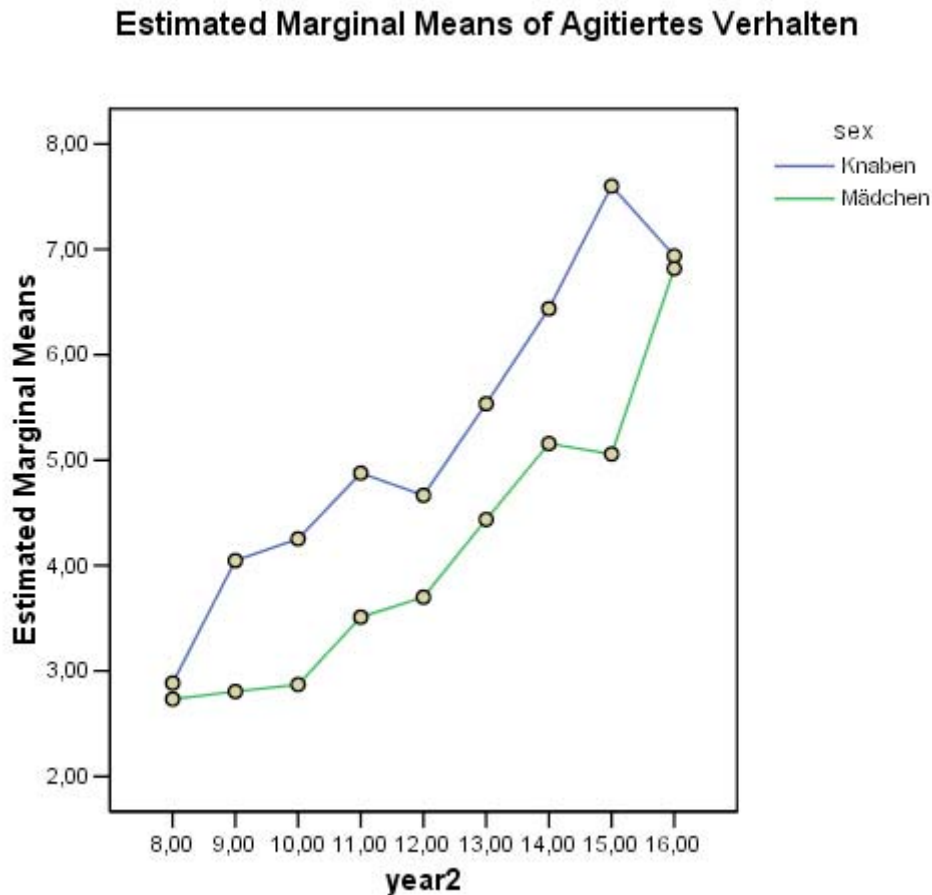
Tabelle 8.2: Mittelwerte und Standardabweichung (SA) der GPW auf der DTK-Skala „Agitiertes Verhalten“ in der Gesamtstichprobe und in geschlechtsspezifischen Altersstufen (nebst t-Test)

Alter	Gesamtstichprobe			Jungen			Mädchen			t-Test	
	n	M	SA	n	M	SA	n	M	SA	T	p
8	171	2,8187	2,24	96	2,8854	2,32	75	2,7333	2,16	,439	,661
9	219	3,4064	2,62	106	4,0472	2,66	113	2,8053	2,43	3,607	,000
10	258	3,6318	2,73	142	4,2535	2,87	116	2,8707	2,34	4,179	,000
11	283	4,2155	2,93	146	4,8767	3,10	137	3,5109	2,56	4,029	,000
12	329	4,1672	3,04	159	4,6667	3,17	170	3,7000	2,84	2,917	,004
13	337	4,9911	3,11	170	5,5353	3,21	167	4,4371	2,90	3,292	,001
14	322	5,8571	3,54	176	6,4375	3,72	146	5,1575	3,17	3,283	,001
15+	97	6,4845	3,61	51	7,3922	3,59	46	5,4783	3,39	2,689	,008
Gesamt	2016	4,4276	3,17	1046	4,9857	3,32	970	3,8258	2,88	8,347	,000

Zeichenerklärung: n= Zahl der entsprechenden Teilstichprobe, M= Mittelwert, SA= Standardabweichung

In der Abbildung 8.2 werden die zuvor dargestellten alters- und geschlechtsspezifischen Unterschiede im Hinblick auf das Aufkommen der Gesamtpunktwerte auf der DTK-Skala 2 veranschaulicht.

Abbildung 8.2: Score in der DTK-Skala „Agitiertes Verhalten“ in Abhängigkeit von Geschlecht und Alter



Auf der Grundlage der hier dargestellten Ergebnisse aus den statistischen Unterschiedsprüfungen dürfen die folgenden Aussagen formuliert werden:

8.6 AS: Im Aufkommen der Gesamtpunktwerte auf der DTK-Skala „Agitiertes Verhalten“ wurden sowohl signifikante Alters- als auch Geschlechtseffekte gefunden.

8.7 AS: Die in der griechischen Stichprobe beobachteten Mittelwerte für die Gesamtpunkte auf Skala 2 nehmen insgesamt mit dem ansteigenden Alter der Kinder zu.

8.8 AS: Die Jungen aus der Gesamtstichprobe erreichen signifikant höhere Mittelwerte in der Skala 2 als die Mädchen.

8.9 AS: Zwischen dem Alter und dem Geschlecht der Kinder besteht in dieser Skala keine signifikante Interaktion. Die Jungen erreichen in jeder Altersstufe – mit Ausnahme die Gruppe der 8jährigen Kinder - signifikant höhere Durchschnittswerte als ihre Mitschülerinnen.

8.10 AS: Die festgestellten alters- und geschlechtsspezifischen Unterschiede auf der Skala „Agitiertes Verhalten“ stehen sowohl mit den entsprechenden Ergebnissen aus der Normstichprobe des Verfahrens als auch mit den meisten deutschsprachigen DTK-Studien in Übereinstimmung.

8.2.3 Geschlechts- und Alterseffekte in der Produktion von Gesamtpunktwerten auf der DTK-Skala „Müdigkeit/autonome Reaktionen“ in der Untersuchung 2034GR.

Aus der mit den DTK-Daten der Skala „Müdigkeit/autonome Reaktionen“ durchgeführten Varianzanalyse ergab sich sowohl ein signifikanter Effekt für den Faktor Geschlecht [$F(7,2023) = 13,095$; $p = 0,000$; Eta-Quadrat = 0,6%] als auch einer für die Interaktion zwischen Geschlecht und Alter [$F(7,2023) = 9,701$; $p = 0,000$; Eta-Quadrat = 0,9%]. Wie aus Tabelle 8.3 zu ersehen ist, erreichen die Mädchen mit ihren Gesamtpunktwerten insgesamt geringfügig höhere Mittelwerte als die Jungen (5,30 versus 4,78). Dieser für die Teilgruppe der Mädchen insgesamt geltende Mittelwert geht auf unterschiedliche Relationen zwischen den für Jungen und Mädchen auf den verschiedenen Altersstufen gefundenen Mittelwerten zurück. Über separat geführte t-Tests ergab sich, dass auf den Alterstufen von 8, 12 und 13 Jahren die Mädchen höhere Durchschnittswerte in der Skala 3 des DTK erreichen als ihre gleichaltrigen Mitschüler.

Die zur Prüfung eines möglicherweise bestehenden allgemeinen Alterseffektes im Sinne steigender Mittelwerte mit steigendem Alter angesetzte Varianzanalyse führte wegen des Ausbleibens entsprechender Signifikanzwerte zur Ablehnung der Unterschiedshypothese [$F(7,2023) = 9,701$; $p = 0,066$].

Tabelle 8.3: Mittelwerte und Standardabweichung (SA) der DTK-Skala „Müdigkeit/ autonome Reaktionen“ in der Gesamtstichprobe und in geschlechtsspezifischen Altersstufen nebst separaten t- Tests

Alter	Gesamtstichprobe			Jungen		Mädchen			T-test		
	n	M	SA	N	M	SA	n	M	SA	t	p
8	174	4,6264	2,75	98	4,1122	2,63	76	5,2895	2,78	-2,860	,005
9	219	4,9315	2,83	106	5,2170	2,79	113	4,6637	2,85	1,451	,148
10	257	5,0233	2,95	141	4,7376	2,88	116	5,3707	3,02	-1,719	,087
11	284	4,7500	2,92	146	4,7945	3,02	138	4,7029	2,81	,264	,792
12	332	4,9699	2,73	161	4,6087	2,75	171	5,3099	2,68	-2,352	,019
13	337	5,2671	2,77	170	4,6824	2,69	167	5,8623	2,73	-3,993	,000
14	323	5,2786	2,75	177	5,0169	2,72	146	5,5959	2,76	-1,891	,060
15+	98	5,4286	2,95	51	5,2941	2,92	47	5,5745	3,00	-,469	,640
Gesamt	2024	5,0331	2,82	1050	4,7810	2,80	974	5,3049	2,82	-4,189	,000

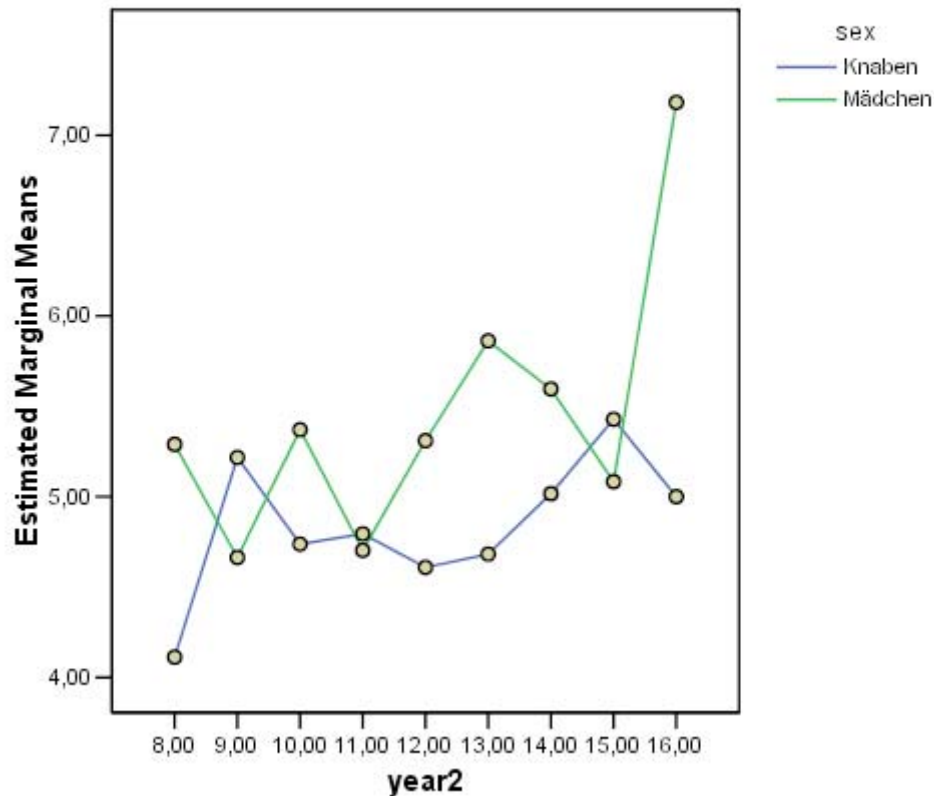
Zeichenerklärung: n= Zahl der entsprechenden Teilstichprobe, M= Mittelwert, SA= Standardabweichung

Sowohl für den nachgewiesenen Haupteffekt des Faktors Geschlecht als auch für den sich darüber hinaus ergebenden Nebeneffekt der Interaktion zwischen Geschlecht und Alter der Kinder sollte an dieser Stelle darauf aufmerksam gemacht werden, dass diese Effekte nur einen geringen Anteil der Variabilität der Messwerte erklären (Eta-Quadrat 0,6% und 0,9%). Interessanterweise wurden solche, in ihrer praktischen Relevanz fraglichen (siehe Rossmann, 1993), weil auf numerisch kleineren Unterschieden in der Skala „Müdigkeit/autonome Reaktionen“ beruhenden Effekte, auch in der Normstichprobe des DTK gefunden (siehe Abschnitt 8.1 dieser Arbeit oder im DTK-Manual, S. 20 und 21). Immerhin ließe sich die methodisch zu analysierende Frage aufwerfen, ob häufig in identischen Zusammenhängen wiederkehrende leichte und im Einzelfall als unbedeutend erscheinende Unterschiede angesichts ihrer relativen Konstanz in ihrer Aussagekraft gleich bleiben oder durch die Wiederholung etwa doch an Beweiskraft gewinnen.

In Abbildung 8.3 sind die zuvor mitgeteilten Ergebnisse graphisch dargestellt.

Abbildung 8.3: Score in der DTK-Skala „Müdigkeit / autonome Reaktionen“ in Abhängigkeit von Geschlecht und Alter

Estimated Marginal Means of Müdigkeit / autonome Reaktionen



Die Ergebnisse der zur DTK-Skala 3 durchgeführten und dargestellten Analyseschritte sind die Grundlage für die nachfolgend angeführten Befunde:

8.11 AS: In der Produktion der Gesamtpunktwerte zur DTK-Skala „Müdigkeit/autonome Reaktionen“ wurden Effekte des Geschlechts und der Interaktion zwischen Alter und Geschlecht gefunden.

8.12 AS: Die Mädchen aus der Gesamtstichprobe erreichen mit ihren Gesamtpunktwerten auf der Skala 3 signifikant geringfügig höhere Mittelwerte als die Jungen.

8.13 AS: Statistisch signifikante geschlechtsspezifische Unterschiede entstehen mit höheren Mittelwerten für die Mädchen in den Alterstufen 8, 12 und 13 Jahre.

8.14 AS: Da die oben erwähnten Effektstärke nur einen geringen Teil der Messvariationen erklären lassen, könnten die beobachteten Effekte durch den relativ großen Umfang der Stichprobe entstanden sein.

8.15 AS: Die hier diskutierten Ergebnisse stimmen mit den entsprechenden Befunden aus der Normstichprobe des DTK überein. Die Übereinstimmung betrifft sowohl die Richtung als auch die möglicherweise bestehende Abhängigkeit der Ergebnisse vom Stichprobenumfang.

8.3 Geschlechtsspezifische Variationen des nach den DTK-Ergebnissen abzuschätzenden mutmaßlichen Auftretens depressiver Störungen in der untersuchten griechischen Stichprobe

In der Fachliteratur über das Auftreten von depressiven Störungen bei Kindern und Jugendlichen finden sich Belege für eine mit dem Alter und dem Geschlecht einhergehende Variation der Auftretensraten depressiver Störungen. Insoweit dürfen Alter und Geschlecht auch als potentielle Risikofaktoren für das Auftreten depressiver Störungen bezeichnet werden. Diese Risikofaktoren scheinen nach den Ergebnissen aus entsprechenden Studien interaktiv auf das Auftreten depressiver Belastungen einzuwirken. So weisen viele Studien in Bezug auf das Geschlecht für Kinder unter 10 Jahren einen Gleichstand der Häufigkeit des Auftretens depressiver Störungen nach. Mit zunehmendem Alter steigt aber der Prozentsatz des Auftretens von depressiven Störungen bei Mädchen stärker als bei Jungen an. Ab der Pubertät sind die Auftretensraten bei den Mädchen allmählich etwa doppelt so groß als die der Jungen.

Demzufolge sind auch für die Untersuchung 2034GR mit Alter und Geschlecht variierende überzufällige Unterschiede im Auftreten von Hinweisen auf das Bestehen depressiver Störungen anzunehmen. Als Kriterium für Hinweise auf das Bestehen depressiver Störungen wird, wie schon in der Einleitung dieses Kapitels erwähnt worden ist, der DTK-Gesamtpunkt看wert aus der Skala 1+2 verwendet.

Durch die nachstehende Kreuztabelle wird in einem ersten diesbezüglichen Analyseschritt der isolierte Effekt des Geschlechts auf die Häufigkeit des Auftretens höherer DTK Gesamtpunktwerte als Indikatoren für das Bestehen von depressiven Störungen geprüft. Da höhere DTK-Werte als Indikatoren angesehen werden, ist auch zu bestimmen, ab welchem Punktwert eine hinreichende Wahrscheinlichkeit für das Bestehen depressiver Störungen angenommen werden soll. Auf die Problematik der Definition von Trennwerten für Punktwerte aus Screening-Verfahren ist in den hier geführten Erörterungen an anderer Stelle bereits ausführlich eingegangen worden (s. Kapitel 7). Nachfolgend wird hier nun der von Pötz (2003) errechnete Trennwert verwendet. Als Cutoff-Score wird der aus den Teilpunktwerten auf Skala 1 und Skala 2 addierte Gesamtpunkt看wert $\text{Skala 1} + \text{Skala 2} = 17$ verwendet (Pötz, s.o.) Der definierte Cutoff-Score trennt zwei in der Stichprobe aufgefundene Risikogruppen voneinander. Die erste Gruppe (depBel 1) besteht aus Kindern, deren Gesamtpunkt看wert aus Skala 1+2 weniger als 17 Punkte erreicht. Nach den von Pötz vorgelegten Untersuchungsergebnissen, erweisen sich Kinder mit den Gesamtpunktwerten Skala 1+2 von 0 bis 16, jeweils einschließlich, als empirisch belegt weniger ausgeprägt depressiv belastet als die Kinder der zweiten Gruppe. Für Kinder, die einen GPW Skala 1+2 unter 17 im DTK erreicht haben, dürfen somit im Vergleich zu der zweiten Gruppe als wahrscheinlich weniger depressiv belastet gelten. Auch diese Annahme ist sehr wahrscheinlich mit einer Quote um 25% fehlerhafter Zuweisungen belastet. Der zweiten Risikogruppe (depBel 2) wurden Kinder zugeordnet, deren individueller Skala 1+2-GPW gleich oder größer als 17 Punkte ist. Für diese Kinder darf empirisch begründet eine gegenüber Gruppe 1 höhere Wahrscheinlichkeit für das Bestehen einer klinisch ausgeprägten depressiven Störung angenommen werden.

Die folgende Kreuztabelle dient der statistischen Überprüfung der Annahme geschlechtsspezifischer Unterschiede im Auftreten von Punktwerten, die zur Zuweisung an die Gruppen depBel 1 und depBel 2 führen. Mit H_1 wird behauptet, die Mädchen der Stichprobe würden signifikant häufiger als die Jungen der Gruppe depBel 2 angehören.

Nach der Gegenhypothese H_0 wäre anzunehmen, die geschlechtsspezifische Verteilung der untersuchten Kinder auf die zwei gebildeten Risikogruppen bliebe als zufällig variierend zu erklären.

Tabelle 8.4: Geschlechtsspezifische Verteilung der Kinder aus der griechischen Stichprobe 2034GR auf zwei Gruppen mit unterschiedlicher Wahrscheinlichkeit für das Vorliegen depressiver Belastungen (Cutoff GPW = 17, nach DTK Skala 1+2)

Geschlecht		depBel 1	depBel 2	Gesamt
Jungen	Anzahl	790 **	250 **	1040
	Erwartete Anzahl	770,3	269,7	1040
	Prozentsatz	76,0%	24,0%	100%
Mädchen	Anzahl	695 **	270 **	965
	Erwartete Anzahl	714,7	250,3	965
	Prozentsatz	72,0%	28,0%	100%
Gesamt	Anzahl	1485	520	2005
	Prozentsatz	74,1%	25,9%	100%

Zeichenerklärung: depBel 1= Gruppe niedriger oder mittlerer depressiver Belastung nach DTK (bis zu 16 GPW), depBel 2= Gruppe hoher depressiver Belastung nach DTK (ab 17 GPW), **= Felder, die von den Erwartungen signifikant abweichen

Die in Tabelle 8.4 dargestellte geschlechtsspezifische Verteilung der Kinder auf die DTK-Risikogruppen depBel 1 und depBel 2 weicht signifikant von den Erwartungen ab [$\chi^2(1) = 4,047$; $p = 0,044$]. Da die Gesamttabelle signifikant von den Erwartungswerten abweicht, können die korrigierten und standardisierten Residualwerte zu Signifikanzaussagen für einzelne Felder benutzt werden. Danach sind dann die nachstehenden Aussagen zu vermerken.

8.16 AS: Die geschlechtsspezifische Verteilung der beteiligten Kinder auf zwei nach DTK Gesamtpunktwerten (DTK-Skala 1+2, Trennwert 17) gebildeten Risikogruppen (depBel 1: niedrige oder mittlere depressive Belastung, depBel 2: vermutlich klinische depressive Belastung) zeigt Variationen, die als nicht zufällig entstanden einzuordnen sind und signifikant von Zufallswerten abweichen. Die niedrige Risikogruppe (depBel 1) ist von den Mädchen unter- und von den Jungen überbesetzt. Die entgegengesetzte Belegung zeigt sich für die Gruppe mit der stärkeren Belastung (depBel 2).

8.17 AS: Die Überrepräsentation der Jungen in der Gruppe depBel 1 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten +2,56%.

8.18 AS: Die Unterrepräsentation der Jungen in der Gruppe depBel 2 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten -7,30%.

8.19 AS: Die Unterrepräsentation der Mädchen in der Gruppe depBel 1 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten -2,76%.

8.20 AS: Die Überrepräsentation der Mädchen in der Gruppe depBel 2 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten +7,87%.

Immerhin ergibt sich, dass die Stärke des positiven Zusammenhanges zwischen Geschlecht und DTK-Risikogruppen extrem gering ist (Phi und Cramer's V = 0,045), was darauf hinweise könnte, dass der beobachtete Effekt auf die Größe der Stichprobe zurückzuführen sei.

Auch wenn es so sein sollte, widersprüche dieses Ergebnis noch nicht den theoriegeleiteten Erwartungen. Die Verteilung basiert auf Daten der Gesamtstichprobe, die eine Altersspanne von 8 bis 18 Jahren deckt. Wie das Alter mit dem Geschlecht der Kinder auf deren Verteilung auf die zwei DTK-Risikogruppen eingewirkt hat, wird an einer nachfolgenden Stelle dieses Kapitels dargestellt werden. Hier sollte erkannt werden, dass die Gesamtskala „Depressive Belastung“ aus den Fragen der Subskalen 1, „Dysphorie/Selbstwert“, und 2, „Agitiertes Verhalten“ entsteht. Diese Skalen sind, wie sich aus den Analysen im Abschnitt 8.1 gezeigt hat, geschlechtsspezifisch

determiniert. Die höheren GPW der ersten Skala werden von den Mädchen und die der zweiten von den Jungen dominiert. So könnte eine zweite Erklärung für den eher geringen Unterschied in der geschlechtsspezifischen Verteilung der Kinder auf die Risikogruppen der Skala 1+2 in der genannten und somit besonderen Konstruktion der Skala DTK1+2 gesehen werden. Einander entgegenlaufende Unterschiede am Aufkommen der GPW in den Skalen heben sich in der Addition beider Skalen annähernd auf.

8.4 Mit dem Alter auftretende Variationen des nach Ergebnissen aus dem DTK abzuschätzenden Auftretens depressiver Störungen in der untersuchten Stichprobe

In der aktuellen Literatur zur Depression bei Kindern und Jugendlichen wird die Meinung vertreten, dass mit zunehmendem Alter auch die depressiven Störungen zunehmen. Als kritisches Lebensalter erscheint in vielen empirischen Studien die Lebenszeit zwischen 12 und 14 Jahren, die mit spezifischen biologischen und psychosozialen Änderungen wegen des Übergangs von der Kindheit zur Adoleszenz verbunden ist.

An der hier darzustellenden Untersuchung 2034GR haben Schüler und Schülerinnen zwischen 8 und 18 Jahren teilgenommen. Für die in dieser Stichprobe auftretenden unterschiedlichen Altersgruppen sollte demnach theoriegestützt erwartet werden, dass sie unterschiedliche Ausprägungen depressiver Belastungen aufweisen. Anzunehmen wäre, dass die älteren Kinder der Stichprobe von depressiven Symptomen in höherem Ausmaß als jüngere betroffen sein werden. Dieser hypothetischen Erwartung stünde eine ebenso hypothetische Vorhersage gegenüber, nach der die Häufigkeit auftretender depressiver Belastungen nicht mit dem Alter variierte. Zu erinnern wäre in diesem Zusammenhang zum einen daran, dass wegen der ausschließlichen Verwendung von Screening-Verfahren jedes Mal neu festzustellen bliebe, dass in der Datenbasis dieser Untersuchung nur Indizien oder bestenfalls kategorial treffender Indikatoren für das Auftreten depressiver Störungen bekannt werden können und somit auch nur Schätzungen von Prävalenzen abgegeben werden können. Zum anderen ist an die zuvor vollzogene Einführung zweier Risikogruppen zu erinnern. Die zur altersspezifischen Häufung von depressiven Störungen gehörende Hypothese wäre entsprechend zu modifizieren. Mit zunehmendem Alter sollte sich eine zunehmende Zahl von Schülerinnen und Schülern in der Risikogruppe 2 einfinden, der Gruppe von Kindern mit höherer DTK-Gesamtpunktwerten und entsprechend höherem Risiko für das Auftreten klinischer Ausprägungen depressiver Störungen.

Tabelle 8.5 stellt die entsprechende Verteilung der untersuchten Kinder auf die unterschiedlichen zwei Belastungsgruppen dar. Für die letzte Altersgruppe (15+) wird an dieser Stelle daran erinnert, dass sie, wie das Symbol + signalisieren soll, auch einige Kinder in sich aufgenommen hat, die schon älter als 15 Jahre sind (n=27). Die Zahl der Kinder auf den Altersstufen über 15 Jahre war nicht hinreichend hoch, um die Kriterien für eine eigenständige statistische Prüfung dieser Altersgruppen vornehmen zu können. Da ein zunehmender Anstieg der Zugehörigkeit mit zunehmendem Alter erwartet wird, müsste für die Altersgruppe 15+ ein stärkerer Anstieg als er dem Alter 15 zustünde, eintreten. Die Gruppierung 15+ folgt also ausschließlich statistischen Vorgaben. Die damit etwa eintretende Verzerrung der Ergebnisse ist allerdings theoretisch beschreibbar und somit abschätzbar. Insoweit ist sie zu akzeptieren.

An einer weiteren Stelle dieses Abschnittes wird auf die spezielle Teilgruppe der überalterten Kinder aus der Gruppe 15+ einzugehen sein.

Tabelle 8.5: Altersspezifische Verteilung der untersuchten Kinder auf die Risikogruppen depBel 1 und depBel 2

Alter		depBel 1	depBel 2	Gesamt
8	Anzahl	148 **	19 **	167
	Erwartete Anzahl	123,7	43,3	16
	Prozent	88,6%	11,4%	100%
9	Anzahl	183 **	33 **	216
	Erwartete Anzahl	160	56	216
	Prozent	84,7%	15,3%	100%
10	Anzahl	197	59	256
	Erwartete Anzahl	189,6	66,4	256
	Prozent	77,0%	23,0%	100%
11	Anzahl	224 **	59 **	283
	Erwartete Anzahl	209,6	73,4	283,0
	Prozent	79,2%	20,8%	100,0%
12	Anzahl	244	85	329
	Erwartete Anzahl	243,7	85,3	329
	Prozent	74,2%	25,8%	100%
13	Anzahl	228 **	109 **	337
	Erwartete Anzahl	249,6	87,4	337
	Prozent	67,7%	32,3%	100%
14	Anzahl	206 **	114 **	320
	Erwartete Anzahl	237,0	83,0	320
	Prozent	64,4%	35,6%	100%
15	Anzahl	55 **	42 **	97
	Erwartete Anzahl	71,8	25,2	97
	Prozent	56,7%	43,3%	100%
Gesamt	Anzahl	1485	520	2005
	Prozent	74,1%	25,9%	100%

Die in Tabelle 8.5 dargestellte altersspezifische Verteilung der Stichprobe auf die DTK-Risikogruppen depBel 1 und depBel 2 ist hochsignifikant [$\chi^2(7) = 74,196$; $p = 0,000$]. Die H_0 ist demnach nicht zu bestätigen. Wegen der nachgewiesenen Signifikanz gestatten die jeweils vorliegenden korrigierten und standardisierten Residualwerte, wenn sie die kritischen Werte erreichen und übersteigen, auch Aussagen über einzelne signifikante Felder der Tabelle. Die in Tabelle 8.5 dargestellten Ergebnisse sind die Datenbasis der nachfolgenden zusammenfassenden Aussagen über die Analyseergebnisse:

8.21 AS: Aus der Durchführung des DTK in einer Stichprobe von 2034 griechischen Kindern zwischen 8 und 18 Jahren ergeben sich statistisch signifikante Belege für einen altersspezifischen Effekt auf die Varianz des Auftretens von DTK Gesamtpunktwerten als kategorial verwendbare Indikatoren für eher geringere oder im Gegenteil höhere depressive Belastung der untersuchten Kinder.

8.22 AS: Die 8-, 9- und 11jährigen Kinder der Stichprobe sind in der DTK Risikogruppe 1 signifikant überrepräsentiert und in der Risikogruppe 2 unterrepräsentiert.

8.23 AS: Die Überrepräsentation der jüngeren Kinder (8, 9 und 11 Jahre) in der Gruppe depBel 1 beträgt gegenüber den über die jeweiligen Randsummen bestimmten Erwartungswerten +19,64% bei den 8jährigen, +14,38% bei den 9jährigen und +6,87 bei den 11jährigen Kindern. Diese Werte unterstützen die Vermutung, eine etwa in jüngeren Jahren vorhandene Resistenz gegenüber depressiven Störungen nehme in Stichproben während der Kindheit mit zunehmendem Alter prozentual ab. Danach wäre aber der Verlust der Resistenz ein individuell zu fassendes Ereignis, für das auch individuell vorliegende oder abwesende Ursachen zu vermuten wären.

8.24 AS: Die Unterrepräsentation der jüngeren Kinder (8, 9 und 11 Jahre) in der Gruppe depBel 2 beträgt gegenüber den über jeweiligen Randsummen bestimmten Erwartungswerten -56,12% bei den 8jährigen, -41,07% bei den 9jährigen und -19,62% bei den 11jährigen Kindern.

8.25 AS: Für die jüngeren Kinder erweist sich, dass der Prozentsatz der Über- und der Unterrepräsentation der beobachteten Werte gegenüber den über Randsummen bestimmten Erwartungswerten mit zunehmendem Alter abnimmt.

8.26 AS: Die in dieser Stichprobe befindlichen 10- und 12jährigen Kinder zeigen mit ihren Frequenzen in den DTK-Belastungsgruppen depBel 1 und depBel 2 nur zufällige Abweichungen.

8.27 AS: Die für die 13-, 14- und 15jährigen Kinder der untersuchten Stichprobe beobachteten Frequenzen zeigen für die Gruppe vermuteter niedriger depressiver Belastung gegenüber den Erwartungswerten signifikante Unterbesetzung und für die Gruppe erwarteter höherer depressiver Belastung deutliche Überbesetzungen.

8.28 AS: Die Unterrepräsentation der älteren Kinder in der Gruppe depBel 1 erreicht gegenüber den über die jeweiligen Randsummen bestimmten Erwartungswerten -8,65% bei den 13jährigen, dann -13,08% bei den 14jährigen und -76,80% für die Gruppe der 15jährigen und teilweise noch älteren Kinder.

8.29 AS: Die Überrepräsentation der älteren Kinder in der Gruppe depBel 2 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten +24,71% bei den 13jährigen, +37,35% bei den 14jährigen und +66,67% bei den 15jährigen oder auch älteren Kindern.

8.30 AS: Über den Einsatz des DTK in einer Stichprobe von 2034 griechischen Kindern ergibt sich deutlich, dass mit wachsendem Alter die Hinweise auf das Bestehen depressiver Störungen mit klinischen Ausprägungen zunehmen.

8.31 AS: Die mitgeteilten Ergebnisse stimmen tendenziell mit den in der einschlägigen Literatur vertretenen Positionen überein.

8.32 AS: Die Ergebnisse der griechischen Studie 2034GR5 stützen auch die These von einer erhöhten Vulnerabilität ab der Altersstufe von 13 Jahren.

8.5 Variationen des geschätzten Auftretens depressiver Störungen im Zusammenhang mit Interaktionen von Alter und Geschlecht

Nachdem vorangegangene und mit ihren Ergebnissen dargestellte Analyseschritte gezeigt haben, dass, jeweils allein für sich genommen, sowohl das Alter als auch das Geschlecht der Kinder mit signifikanten Unterschieden in der Verteilung der untersuchten Kinder auf die Risikogruppen depBel 1 und depBel 2 variieren, wird nun mit der nachstehenden Tabelle mutmaßlichen Zusammenhängen zwischen Kombinationen von Alter und Geschlecht einerseits und der Verteilung der Kinder auf die Risikogruppen andererseits nachgegangen. Zu erwarten sind solche Beziehungen, weil zahlreiche internationale Untersuchungsbefunde und deren Repräsentationen im vorherrschenden Störungsbild der Depression dies nahe legen. Dementsprechend wäre zu erwarten, dass auch die in der Untersuchung 2034GR über die Anwendung des DTK gesammelten Indizien für das Bestehen depressiver Störungen unter den Kindern der untersuchten Stichprobe in ihrem Aufkommen auf verschiedenen Altersstufen unterschiedliche geschlechtsspezifische Variationen zeigen. Die depressionsspezifischen Relationen zwischen den Geschlechtern sollten danach mit Veränderungen im Alter der Kinder nicht konstant sein, sondern auf unterschiedlichen Altersstufen verschiedenartig ausfallen. Wenn dies auf die Untersuchung 2034GR auch zuträfe, müsste sich das durch mit den Altersstufen variierende überzufällige Unterschiede in der geschlechtsspezifischen Verteilung der Kinder auf die eingeführten zwei Risikogruppen manifestieren.

Tabelle 8.6: Geschlechtsspezifische Verteilung der Kinder der Stichprobe 2034GR auf die eingeführten Risikogruppen in unterschiedlichen Altersstufen

Alter	Geschlecht		depBel 1	depBel 2	Gesamt
8jährige	Jungen	Anzahl	87 **	7 **	94
		Erwartete Anzahl	69,6	24,4	94
		Prozent	92,6%	7,4%	100%
	Mädchen	Anzahl	61	12	73
		Erwartete Anzahl	54,1	18,9	73
		Prozent	83,6%	16,4%	100%
9jährige	Jungen	Anzahl	86	19	105
		Erwartete Anzahl	77,8	27,2	105
		Prozent	81,9%	18,1%	100%
	Mädchen	Anzahl	97 **	14 **	111
		Erwartete Anzahl	82,2	28,8	111
		Prozent	87,4%	12,6%	100%
10jährige	Jungen	Anzahl	104	37	141
		Erwartete Anzahl	104,4	36,6	141
		Prozent	73,8%	26,2%	100%
	Mädchen	Anzahl	93	22	115
		Erwartete Anzahl	85,2	29,8	115
		Prozent	80,9%	19,1%	100%
11jährige	Jungen	Anzahl	112	34	146
		Erwartete Anzahl	108,1	37,9	146
		Prozent	76,7%	23,3%	100%
	Mädchen	Anzahl	112 **	25 **	137
		Erwartete Anzahl	101,5	35,5	137
		Prozent	81,8%	18,2%	100%
12jährige	Jungen	Anzahl	122	37	159
		Erwartete Anzahl	117,8	41,2	159
		Prozent	76,7%	23,3%	100%
	Mädchen	Anzahl	122	48	170
		Erwartete Anzahl	125,9	44,1	170
		Prozent	71,8%	28,2%	100%
13jährige	Jungen	Anzahl	126	44	170
		Erwartete Anzahl	125,9	44,1	170
		Prozent	74,1%	25,9%	100%
	Mädchen	Anzahl	102 **	65 **	167
		Erwartete Anzahl	123,7	43,3	167
		Prozent	61,1%	38,9%	100%
14jährige	Jungen	Anzahl	125	49	174
		Erwartete Anzahl	128,9	45,1	174
		Prozent	71,8%	28,2%	100%
	Mädchen	Anzahl	81 **	65 **	146
		Erwartete Anzahl	108,1	37,9	146
		Prozent	55,5%	44,5%	100%
15jährige	Jungen	Anzahl	28 **	23 **	51
		Erwartete Anzahl	37,8	13,2	51
		Prozent	54,9%	45,1%	100%
	Mädchen	Anzahl	27 **	19 **	46
		Erwartete Anzahl	34,1	11,9	46,0
		Prozent	58,7%	41,3%	100%
Gesamt		Anzahl	1485	520	2005
		Prozent	74,1%	25,9%	100%

Zunächst und vor jeder weiteren Aussage über die bestehenden Unterschiede ist festzustellen, dass die in Tabelle 8.6 dargestellte Verteilung der untersuchten Stichprobe auf die DTK-Risikogruppen depBel 1 und depBel 2 hochsignifikant [$\chi^2(15) = 99,126$; $p = 0,000$] von zufällig auftretenden Unterschieden zwischen erwarteten und beobachteten Besetzungen der Tabellenfelder abweicht. Insofern ist als erwiesen anzusehen, dass in dieser Stichprobe die Verteilung von Kindern auf depressionsspezifische Risikogruppen nach DTK Gesamtpunkten hochsignifikant und geschlechtsspezifisch mit wechselnden Altersstufen variiert. Da dann

weiterhin in dieser Tabelle die korrigierten standardisierten Residualwerte die signifikant abweichende Besetzung einzelner Tabellenfelder nachweist, sind außer der die Gesamttabelle betreffenden Feststellung auch auf einzelne Tabellenfelder bezogene Aussagen möglich.

Die in Tabelle 8.6 dargestellten Werte führen dann insgesamt zu den nachstehenden Aussagen:

8.33 AS: Die Verteilung der Schulkinder auf zwei nach DTK-Gesamtpunktwerten (DTK Skala 1+2, Grenzpunktwert 17) gebildeten Teilgruppen (Risikogruppe depBel 1, weniger stark ausgeprägte Hinweise auf depressive Störungen und depBel 2, stark ausgeprägte Hinweise auf das Bestehen depressiver Störungen) variiert signifikant mit geschlechtsspezifischen Altersgruppen.

8.34 AS: In der Altersgruppe der 8jährigen Kinder treten bei den Jungen signifikante und bei den Mädchen nur zufällige Abweichungen im Hinblick auf ihre Verteilung auf die zwei DTK-Risikogruppen auf. Die Jungen sind gegenüber den über Randsummen bestimmten Erwartungswerten in der Gruppe depBel 1 mit +25% überrepräsentiert und in der Gruppe depBel 2 mit -71,31% unterrepräsentiert. Die Jungen der Stichprobe bleiben auf dieser Altersstufe hoch signifikant unter den über Randsummen bestimmten Erwartungswerten für das Auftreten depressiver Belastungen, während die Mädchen denen weitgehend entsprechen. Die Gruppe der 8jährigen Jungen schöpft das über die Erwartungswerte bestimmte Potential der Entwicklung depressiver Belastungen nicht aus.

8.35 AS: In der Altersgruppe der 9jährigen Kinder treten bei den Jungen zufällige und bei den Mädchen signifikante Abweichungen im Hinblick auf ihre Verteilung auf die zwei DTK-Risikogruppen auf. Die Mädchen dieses Alters sind gegenüber den über Randsummen bestimmten Erwartungswerten in der Gruppe depBel 1 mit +18% überrepräsentiert und in der Gruppe depBel 2 mit -51,39% unterrepräsentiert. Die Jungen dieser Altersstufe entsprechen den Erwartungswerten, während die Mädchen unter den Erwartungswerten für das Auftreten depressiver Symptome bleiben. Die Gruppe der 9jährigen Mädchen zeigt geringere Frequenzen an depressiver Belastung als statistisch erwartet.

8.36 AS: Die 10jährigen Jungen und Mädchen der Stichprobe zeigen mit ihren Frequenzen in den Gruppen depBel 1 und depBel 2 nur zufällige Abweichungen. Sowohl die Jungen als auch die Mädchen von 10 Jahren folgen in ihren Verteilungen auf die Risikogruppen den Erwartungswerten.

8.37 AS: In der Altersgruppe der 11jährigen Kinder treten bei den Jungen zufällige und bei den Mädchen signifikante Abweichungen im Hinblick auf ihre Verteilung auf die zwei DTK-Risikogruppen auf. Die Mädchen dieses Alters sind gegenüber den über Randsummen bestimmten Erwartungswerten in der Gruppe depBel 1 mit +10,34% überrepräsentiert und in der Gruppe depBel 2 mit -29,58% unterrepräsentiert. Die 11jährigen Jungen entsprechen den Erwartungswerten, während die Gruppe der Mädchen geringere Frequenzen an depressiv belasteten Kindern als statistisch erwartet umfasst.

8.38 AS: Die 12jährigen Jungen und Mädchen der Stichprobe zeigen mit ihren Frequenzen in den Gruppen depBel 1 und depBel 2 nur zufällige Abweichungen.

8.39 AS: In der Altersgruppe der 13jährigen Kinder treten bei den Jungen zufällige und bei den Mädchen signifikante Abweichungen im Hinblick auf ihre Verteilung auf die zwei DTK-Risikogruppen auf. Die Mädchen dieses Alters sind gegenüber den über Randsummen bestimmten Erwartungswerten in der Gruppe depBel 1 mit -17,54% unterrepräsentiert und in der Gruppe depBel 2 durch +50,12% überrepräsentiert. Die Gruppe der 13jährigen Mädchen umfasst mehr Kinder mit stärkeren depressiven Belastungen als statistisch erwartet.

8.40 AS: In der Altersgruppe der 14jährigen Kinder treten bei den Jungen zufällige und bei den Mädchen signifikante Abweichungen im Hinblick auf ihre Verteilung auf die zwei DTK-Risikogruppen auf. Die Mädchen dieses Alters sind gegenüber den über Randsummen bestimmten Erwartungswerten in der Gruppe depBel 1 bei -25,07% unterrepräsentiert und in der

Gruppe depBel 2 mit +71,50% überrepräsentiert. Die Gruppe der 14jährigen Mädchen umfasst mehr Kinder mit stärkeren depressiven Belastungen als statistisch erwartet.

8.41 AS: In der Altersgruppe der 15jährigen oder auch älteren Kinder der Stichprobe treten sowohl bei den Mädchen als auch bei den Jungen signifikante Abweichungen im Hinblick auf ihre Verteilung auf die zwei DTK-Risikogruppen auf. Beide Geschlechter dieser Altersgruppe zeigen Unterbesetzung in der Gruppe depBel 1 und Überbesetzung in Gruppe depBel 2. Die Unterrepräsentation der Kinder in der Risikogruppe depBel 1 beträgt gegenüber den über jeweiligen Randsummen bestimmten Erwartungswerten -25,93% bei den Jungen und -20,82 bei den Mädchen. Die Überrepräsentation der Kinder in der Risikogruppe depBel 2 umfasst gegenüber den über jeweiligen Randsummen bestimmten Erwartungswerten +74,24% bei den Jungen und +59,66 bei den Mädchen. Die Gruppen der Jungen und Mädchen schließen jeweils mehr Kinder mit depressiven Belastungen ein als statistisch erwartet.

Die mit den aufgeführten 9 Aussagen mitgeteilten Ergebnisse weisen für die verschiedenen Altersstufen das Auftreten von signifikanten Variationen in der Verteilung von Jungen und Mädchen auf die unterschiedenen zwei depressionspezifischen Risikogruppen nach. Das ist gleichbedeutend mit überzufälligen Unterschieden in dem nach DTK Gesamtpunktwerten zu vermutenden Auftreten von Hinweisen auf das Bestehen depressiver Störungen. Die vorangegangene Tabelle kann auch mit Blick auf die für Jungen und Mädchen auf den einzelnen Alterstufen über die Randsummen errechneten Erwartungswerte gesehen werden. Dann bleiben die Jungen oder die Mädchen manchmal unter den erwarteten Frequenzen von Kindern mit DTK Werten die als Hinweise auf klinische relevante Ausprägungen depressiver Störungen verstanden werden. Das Aufkommen an solchen Kindern ist dann niedriger als erwartet. In diesem Sinne besteht auf solcher Altersstufe also ein Minus an klinisch belasteten Jungen oder Mädchen. Auf anderen Altersstufen werden wiederum die Erwartungswerte eingestellt (=). Die dritte Variante ist ein auf der Altersstufe entstehendes Plus im Aufkommen von Jungen oder Mädchen mit nach DTK mutmaßlich in klinischem Ausmaß depressiv belasteten Kindern (Plus). Die nachstehende Tabelle gibt eine solche stark vereinfachende Übersicht über die Tendenz der analysierten Befunde. Sie zeigt die Tendenz der Abweichungen von den statistischen Erwartungswerten.

Tabelle 8.7: Richtung der bei Jungen und Mädchen auf den Altersstufen vom statistischen Erwartungen abweichenden jeweiligen Frequenz mutmaßlich depressiv belasteter Kinder

Alter	Richtung der Abweichung	
	Jungen	Mädchen
8	Minus	=
9	=	Minus
10	=	=
11	=	Minus
12	=	=
13	=	Minus
14	=	<u>Plus</u>
15+	<u>Plus</u>	<u>Plus</u>

Zeichenerklärung: Minus = Gegenüber Erwartungswerten geringeres Aufkommen an mutmaßlich in klinischer Ausprägung depressiv belasteten Jungen oder Mädchen; mit = wird ein den Erwartungswerten folgendes Aufkommen angezeigt; Plus = ein gegenüber Erwartungswerten höheres Aufkommen.

Die stark vereinfachende Übersicht wird hier zu einer weiteren Aussage genutzt.

8.41 AS: Die altersspezifischen Variationen der geschlechtsspezifischen Verteilung auf die unterschiedenen zwei Risikogruppen zeigen bei den Mädchen tendenziell häufiger Abweichungen von Erwartungswerten auf den Altersstufe als bei den Jungen.

Die mit Alter und Geschlecht auftretenden Variationen in der Verteilung auf depressionsspezifische Risikogruppen stimmen tendenziell mit den nach dem aktuellen Forschungsstand zu vertretenden Erwartungen zu alters- und geschlechtsspezifischen Unterschieden in den Prävalenzen für depressive Störungen überein. Auch in der Untersuchung 2034GR tritt die mit dem Alter anwachsende Zunahme von Hinweisen auf das Bestehen depressiver Störungen mit klinischen Ausprägungen verstärkt bei den älteren Mädchen auf (ab 13 Jahren). Etwas unerwartet sind die Relationen zwischen Mädchen und Jungen in der Altersgruppe 15+. Während die Verteilungen der 13- und 14jährigen Jungen auf die zwei DTK-Risikogruppen, im Gegensatz zu den der Mädchen, nur zufällig von den Erwartungswerten abweichen, ergeben sich aus der Tabelle 8.6 für die Altersgruppe der 15jährigen oder älteren Kinder andere Relationen. Bei dieser Altersgruppe wird die Gruppe hoher depressiver Belastung (depBel 2) sowohl von den Mädchen als auch von den Jungen gegenüber Erwartungswerten überbesetzt. Bemerkenswert sollte werden, dass der Prozentsatz der hochbelasteten Jungen dieser Altersgruppe 15+ höher als jener der Mädchen ist (45,1% zu 41,3%). Auch wenn dieser prozentuale Unterschied relativ niedrig ist, widerspricht er der Aussage von einer mit wachsendem Alter allgemein auftretenden höheren Belastung der Mädchen. Als möglichen Ursachen für das bemerkte Phänomen können unterschiedliche Hypothesen heran gezogen werden.

Eine der möglichen Erklärungen für diesen Widerspruch wäre mit dem Hinweis auf den relativ geringen Umfang der Teilstichproben 15+ zu geben. Kleine Stichproben begünstigen das Auftreten extremer Datenvariationen. Ungewiss bleibt, ob die Zusammensetzung der 15+

genannten Altersstufe, die Jugendliche bis zu 18 Jahren umfasst, einen speziellen Einfluss auf die Ausprägung der Ergebnisse hatte. Ein solcher Effekt wäre erklärbar, weil die Zusammensetzung der Teilgruppe 15+ keine zufällige Altersstreuung hat. Die deutlich älteren Kinder in den betreffenden Schulklassen sind Jungen und Mädchen mit einem speziellen Schulschicksal. Diese Besonderheit könnte die psychische Befindlichkeit dieser Jugendlichen betreffen und beispielsweise das Entstehen depressiver Störungen begünstigt haben. Wie schon zuvor erklärt, erfolgte die Zusammenlegung der fraglichen Altersstufen nur, um Analyseschritte durch Frequenzanhebungen zu erleichtern.

Diese Vorgehensweise begünstigte die Entstehung einer Teilgruppe, die eben nicht nur inhomogen im Hinblick auf das Alter, sondern sehr wahrscheinlich auch in weiteren die Jugendlichen kennzeichnenden Merkmalen sind. Solche Eigenheiten dürften mit der Überalterung der Kinder einhergehen, die schon länger zur Schule gehen als die Mehrheit der Kinder in der jeweiligen Schulklasse. Die 16-, 17- und 18jährige Jugendlichen sind in ihren Schulklassen ausnahmslos überaltert.

Im Hinblick darauf wären vermehrt Informationen sammeln und zu analysieren, die entscheiden ließen, ob diese Jugendlichen im Zusammenhang mit ihrem besonderen Schulschicksal nicht auch besondere Persönlichkeitsmerkmale erkennen ließen.

An dieser Stelle wird nun nachfolgend noch einmal auf die Verteilung der individuellen DTK-Gesamtpunktwerte der untersuchten Kinder auf die zuvor definierten zwei DTK-Risikogruppen für das Bestehen depressiver Störungen eingegangen. Berücksichtigt werden dabei sowohl das Alter als auch das Geschlecht der Kinder. Diesmal wird aber die Verteilung nicht unmittelbar auf die jahrgangsweise unterschiedenen Altersklassen bezogen, sondern auf einen vermittelten, indirekten und zugleich relativen Ausdruck für Alter. Verwendet wird dem griechischen Sprachgebrauch folgend die Unterscheidung von Grundschulern und Gymnasiasten. Die erstere Teilgruppe umfasst die jüngeren und die zweite Teilgruppe die älteren Kinder oder Jugendlichen dieser Stichprobe. Die Zugehörigkeit zu einem Schultyp ist ein Äquivalent für eine von zwei Altersstufen, mittelbar unterschieden und klassifiziert nach Jahren des Schulbesuchs. Eine Unschärfe und Doppelsinnigkeit der Zuweisung entsteht damit, weil nicht nur das Alter der Schülerinnen und Schülern sondern zugleich auch der Schultyp variiert. Wer zu den jüngeren Kindern gehört, verweilt noch in einem Schultyp, der in seiner pädagogischen Anlage deutlich vom Gymnasium unterschieden ist. Aber diese Kinder sind zugleich eben auch jünger als die anderen. Beides ist nicht trennbar. Die Teilstichprobe der Grundschul Kinder besteht aus den jüngeren Kindern. Die sind zwischen 8 und 12 Jahren, einige 13jährige eingeschlossen.. Die Teilstichprobe der Gymnasiasten besteht aus Kindern zwischen 13 und 15 Jahren, sie umschließt einige 12- und auch 16- bis 18jährige. Die nachfolgende Kreuztabelle zeigt die Daten. Sie ist als Abbildung der durch das Schulwesen gebildeten und damit von den Kindern alltäglich erlebten Gruppierung in die Analyse einbezogen worden.

Tabelle 8.8: Nach Schultyp unterschiedene geschlechtsspezifische Verteilung auf DTK Risikogruppen

Schultyp	Geschlecht		depBel 1	depBel 2	Gesamt
Grundschule	Jungen	Anzahl	413 **	104 **	517
		Erwartete Anzahl	382,9	134,1	517
		Prozent	79,9%	20,1%	100%
	Mädchen	Anzahl	383 **	80 **	463
		Erwartete Anzahl	342,9	120,1	463
		Prozent	82,7%	17,3%	100%
Gymnasium	Jungen	Anzahl	377	146	523
		Erwartete Anzahl	387,4	135,6	523
		Prozent	72,1%	27,9%	100%
	Mädchen	Anzahl	312 **	190 **	502
		Erwartete Anzahl	371,8	130,2	502
		Prozent	62,2%	37,8%	100%
Gesamt		Anzahl	1485,0	520,0	2005
		Prozent	74,1%	25,9%	100%

Die als signifikant oder sehr signifikant (*; **) gekennzeichneten Einzelzellen sind durch entsprechende korrigierte und standardisierte Residualwerte als solche ausgewiesen. Die Tabelle ist statistisch hoch signifikant [$\chi^2(3) = 65,336$; $p = 0,000$].

Nach den in dieser Tabelle zu findenden Werten sind die nachstehenden Aussagen anzuführen.

8.42 AS: Die Verteilung der Schulkinder auf die zwei DTK-Belastungsgruppen variiert signifikant mit geschlechtsspezifischen Altersgruppen

8.43 AS: Die in dieser Stichprobe befindlichen jüngeren Jungen (aus der Grundschule) sind in der DTK Risikogruppe 1 signifikant überrepräsentiert und demnach in der Risikogruppe 2 unterrepräsentiert.

8.44 AS: Die Überrepräsentation der Jungen aus der Grundschule in der Gruppe depBel 1 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten +7,86%.

8.45 AS: Die Unterrepräsentation der Jungen aus der Grundschule in der Gruppe depBel 2 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten -22,45%.

8.46 AS: Die in dieser Stichprobe befindlichen jüngeren Mädchen (aus der Grundschule) sind in der DTK Risikogruppe 1 signifikant überrepräsentiert und dem entsprechend in der Risikogruppe 2 unterrepräsentiert.

8.47 AS: Die Überrepräsentation der Mädchen aus der Grundschule in der Gruppe depBel 1 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten +11,69%.

8.48 AS: Die Unterrepräsentation der Mädchen aus der Grundschule in der Gruppe depBel 2 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten -33,39%.

8.49 AS: Die in dieser Stichprobe befindlichen älteren Jungen (aus den Gymnasien) zeigen mit ihren Frequenzen in den DTK Risikogruppe 1 und 2 nur zufällige Abweichungen.

8.50 AS: Die in dieser Stichprobe befindlichen älteren Mädchen (aus den Gymnasien) sind in der DTK Risikogruppe 1 signifikant unterrepräsentiert und in der Risikogruppe 2 überrepräsentiert.

8.51 AS: Die Unterrepräsentation der Mädchen aus Gymnasien beläuft sich in der Gruppe depBel 1 gegenüber den über Randsummen bestimmten Erwartungswerten auf -16,08%.

8.52 AS: Die Überrepräsentation der Mädchen aus Gymnasien in der Gruppe depBel 2 erreicht gegenüber den über Randsummen bestimmten Erwartungswerten +45,93%.

Aus den Feststellungen ist wie aus der Tabelle zu ersehen, dass die Überrepräsentation der älteren, in Gymnasien beschulten Mädchen weit über den Erwartungswerten liegt.

8.6 Das Wiederholen von Schulklassen als potentieller Risikofaktor für das Auftreten von depressiven Störungen

Der Begriff Risikofaktor steht an dieser Stelle für den Sachverhalt, dass ein von diesem Faktor betroffenes Schulkind eine gegenüber anderen erhöhte Wahrscheinlichkeit trifft, durch depressive Störungen belastet zu sein. Die jeweils zu vermutenden kausalen Verbindungen für dieses Ereignis werden durch den Begriff nicht benannt.

Eine Teilgruppe, die bei der Durchführung der Untersuchung, obschon nicht eingeplant, entstand und der eine hypothetisch erhöhte depressogene Vulnerabilität zugesprochen werden könnte, umfasst alle diejenigen Kinder einer jeden Schulklasse, die das jeweilige klassentypische Alter, definiert durch das griechische Schulsystem, zum Zeitpunkt der Untersuchung überschritten hatten.

Die vierte Klasse der griechischen Grundschule beispielsweise sollten Kinder besuchen, deren Alter zwischen 9 und 10 Jahren liegt. Ältere Kinder in einer solchen Klasse, in diesem Beispiel etwa ein 11jähriges oder auch älteres Kinde wären als in dieser Klasse überaltert zu bezeichnen. Möglicherweise treten die Besonderen in der Schullaufbahn auch zusammen mit besonderen Risiken für das Bestehen depressiver Störungen auf. Solche Verbindungen wären, folgte man den verschiedenen Theorien über depressive Störungen, mehrfach zu begründen.

Bereits am Anfang der eigenen Individualentwicklung oder aber auch erst seit dem Schuleintritt können, wie eben zu jedem anderen Zeitpunkt auch, spezifische depressogene Entwicklungen begonnen haben, die dann einer als normentsprechend zu bewertende Schullaufbahn entgegen standen. Oder aber die aus anderen Gründen gescheiterte oder beeinträchtigte Schulkarriere hinterließ depressive Störungen. Trotz der im Einzelnen unterschiedlichen und unaufgeklärten Genese sind die in ihrer Schulklasse überalterten Kinder oder Jugendliche potentielle Risikoträger für depressive Entwicklungen. Gesundheitliche oder auch psychische Problem, Schwierigkeiten mit der Sprache eines Gastlandes nach Migration, schlechte Schulleistungen oder kombinierte Wirkungen solcher oder ähnlicher Ereignisse könnten mit depressiven Störungen verbunden sein und zugleich Gründe für das Wiederholen einer Schulklasse sein. Die nachfolgende Tabelle stellt die Verteilung der an der Untersuchung 2034GR beteiligten Kinder auf die 7 besuchten Klassenstufen dar, je nach Relation zu ihrem Alter.

Tabelle 8.9: Überalterte Kinder und Jugendliche der untersuchten Stichprobe in den verschiedenen Schulklassen

Klassenstufe	dem Alter entsprechend	Klassenbesuch				Gesamt
		überalterte Kinder				
		1 Jahr	2 Jahre	3 Jahre	4 Jahre	
Grundschule						
3e (8-9 J.)	222					222
4e (9-10 J.)	247	2	1			250
5e (10-11 J.)	280	3	1			284
6e (11-12 J.)	241	3				244
Gymnasium						
1e (12-13 J.)	327	18	5	1		351
2e (13-14 J.)	323	22	5	1	1	352
3e (14-15 J.)	312	14	4	1		331
Gesamt	1952	62	16	3	1	2034

Aus Tabelle 8.9 ist zu erfahren, dass beispielsweise in der dritten Klasse der Grundschulen, die wir besucht haben, kein Kind zu finden war, das überaltert gewesen wäre. Bei der Teilstichprobe der Kinder aus den vierten Klassen der Grundschule (250 Kinder) wurden dagegen 2 Kinder gefunden, die 1 Jahr älter als erwartet waren, und ein weiteres Kind erwies sich als um 2 Jahre älter. Deutlich wird in dieser Tabelle auch, dass die Mehrheit der überalterten Kinder in den Gymnasien aufzufinden war. Dem Zahlenverhältnis von 10 zu 72 für überalterte Schülerinnen und Schüler in Grundschulen und Gymnasien steht das Verhältnis von altersgerecht in den Schulklassen aufzufindenden Schülerinnen und Schüler von 990 zu 962 gegenüber. Der Unterschied ist signifikant. Dementsprechend ist die nachstehende Aussage zu vermerken.

8.53 AS: Die Zahl der überalterten Schülerinnen und Schüler in den Schulklassen der mit der Stichprobe 2034GR erfassten Schulen liegt bei etwa 4% der Gesamtzahl aller Schülerinnen und Schüler.

8.54 AS: Zwischen überalterten Schülerinnen und Schülern an Grundschulen und Gymnasien ergab sich ein Verhältnis von etwa 1% zu annähernd 7,5% der jeweiligen Gesamtzahl.

8.55 AS: An den Gymnasien der untersuchten Stichprobe fanden sich signifikant mehr überalterte Schülerinnen und Schüler als an Grundschule.

Eine mögliche Ursache für den im Vergleich von Grundschulen und Gymnasien beobachteten Unterschied im Auftreten von depressionsspezifischen Risikogruppen, einschließlich der überalterten Schülerinnen und Schüler, ist die an Gymnasien verstärkt praktizierte pädagogische Reglementierung über Leistungsforderungen und -kontrolle. Erst im griechischen Gymnasium werden ausschließlich Zensuren als Kriterien für das Erreichen oder Nichterreichen des jeweils gesetzten Klassenzieles verwendet, Damit vergrößern sich die psychischen Belastungen für ein Klientel, das diese Praxis als Überlastung erlebt. Aus dem damit verbundenen Auftreten von Risikofaktoren für die Entwicklung depressiver Störungen und auch aus deren Kombination mit anderen psychischen oder somatischen Beschwerden ergibt sich ein an Gymnasien zu vermutendes Potential zur Begünstigung von Häufungen depressiver Störungen. Die überalterten Kinder sind als durch Verzögerungen im regelgemäßen Ablauf des Schulbesuchs Betroffene ausgewiesen. Damit gehören sie zu denen, die ein stärkeres Risiko für die Entwicklung psychischer Störungen tragen. Die damit bestimmte Möglichkeit ist ein hinreichender Grund für die Einsetzung entsprechender Hypothesen. Wie dargestellt, kommt die Mehrheit der bekannt gewordenen überalterten Kinder (87,80%) aus Gymnasialklassen. Da weiterhin in vorausgegangen Analysen gefunden wurde, dass zwischen Grundschulkindern und Gymnasiasten signifikante Unterschiede im Hinblick auf deren Verteilung auf die zwei eingeführten DTK-Risikogruppen bestehen, darf die nachfolgende Analyse auf die Teilstichprobe der Gymnasiasten beschränkt bleiben. Überprüft wird die nachstehende Hypothese.

H₁: Die Verteilung der in gymnasialen Schulklassen altersgerecht befindlichen Schüler und Schülerinnen auf depressionsspezifische Risikogruppen (nach DTK) einerseits und die der überalterten Jugendlichen andererseits werden signifikante Unterschiede aufweisen.

Tabelle 8.10: Verteilung der in gymnasialen Schulklassen überalterten und der demgegenüber altersgerecht befindlichen Schülerinnen und Schüler auf definierte depressionsspezifische Risikogruppen (nach DTK)

Klassenbesuch		depBel 1	depBel 2	Gesamt
dem Alter entsprechend	Anzahl	649 **	305 **	954
	Erwartete Anzahl	641,3	312,7	954
	Prozent	68%	32%	100%
überalterte Kinder	Anzahl	40 **	31 **	71
	Erwartete Anzahl	47,7	23,3	71
	Prozent	56,3%	43,7%	100%
Gesamt	Anzahl	689	336	1025
	Prozent	67,2%	32,8%	100,0%

Zeichenerklärung: depBel 1= Teilgruppe mit nach DTK geringerem Risiko für das akute Bestehen depressiver Störungen; depBel 2= mit gegenüber depBel1 höherem Risiko belastet; **= Felder, die nach korrigierten standardisierten Residuen von den erwarteten Werten signifikant abweichen

Die in Tabelle 8.10 dargestellte Verteilung ist signifikant [$\chi^2(1) = 4,670$; $p = 0,031$]. Die H_1 -Hypothese gilt somit als verifiziert. Daraufhin dürfen auch die folgenden Aussagen formuliert werden:

8.56 AS: Die Verteilung der Schulkinder auf die zwei gebildeten Risikogruppen depBel 1 und depBel 2 variieren signifikant mit dem altersgerechten oder davon abweichenden Schulbesuch.

8.57 AS: Die Schüler, die eine ihrem Alter entsprechende Schulklasse besuchen, sind in der DTK-Gruppe niedriger Belastung (depBel 1) signifikant überrepräsentierter und in der Gruppe hoher Belastung (depBel2) unterrepräsentiert.

8.58 AS: Die Überrepräsentation dieser Schüler in der Risikogruppe 1 beträgt gegenüber den nach Randsummen der Tabelle bestimmten Erwartungswerten +1,2%.

8.59 AS: Die Unterrepräsentation dieser Schüler in der Risikogruppe 2 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten -2,46%.

8.60 AS: Die Schüler, die im Vergleich zu ihren Mitschülern in der Klasse überaltert sind, bleiben in der DTK-Gruppe niedriger Belastung (depBel 1) signifikant unterrepräsentiert, treten aber in der Gruppe hoher Belastung (depBel 2) überrepräsentiert auf.

8.61 AS: Die Unterrepräsentation der überalterten Kinder in der Risikogruppe 1 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten -16,14%.

8.62 AS: Die Überrepräsentation der überalterten Kinder in der Risikogruppe 2 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten +33,05%.

Die Einbeziehung überalterter Kinder in die hier beschriebene Untersuchung war, wie bereits erwähnt, nicht geplant. Die Teilnahme dieser speziellen Teilstichprobe, auch wenn sie nur eine Minderheit ist, an der Untersuchung 2034GR verweist aber auf eine Realität, die wegen der dargestellten sieben Aussagen, insbesondere im Hinblick auf die zuletzt angeführte, spezielle Beachtung finden sollte. Aus eben diesem Grund wird im Folgenden noch einmal auf die Untergruppe der überalterten Mädchen und Jungen eingegangen. Das kann allerdings nur in dem Umfang geschehen, den die verfügbaren Daten zulassen. Eine beantwortbare Zusatzfrage ist, ob auch in dieser Untergruppe geschlechtsspezifische Unterschiede in ihrer Verteilung auf die zwei DTK-Risikogruppen zu finden sind. Wie aus Tabelle 8.9 zu ersehen war, besucht die Mehrheit der überalterten Kinder, nämlich 72 von insgesamt 82, eine Gymnasialklasse. Für die Gesamtstichprobe der Gymnasiasten hatte sich aus vorherigen Analysen (Kapitel 8.4) ergeben, dass ihre Verteilung auf die Risikogruppen depBel1 und depBel2 geschlechtsspezifisch unterschiedlich ist. Die Mädchen aus Gymnasien treten stärker als erwartet in Risikogruppe

depBel2 auf. Dementsprechend könnte auch für die Teilstichprobe der überalterten Kinder vermutet werden, dass dort gleichsinnige geschlechtsspezifische Unterschiede beobachtbar wären. Wiederum sollten die stärker depressiv belasteten Mädchen überrepräsentiert sein. Untersucht wird also die folgende Hypothese:

H₁: Die Verteilung der Teilgruppe der überalterten Kinder auf die zwei DTK-Belastungsgruppen ist geschlechtsspezifisch signifikant unterschiedlich.

Die nun folgende Tabelle zeigt die Verteilung der überalterten Mädchen und Jungen auf die nach DTK Gesamtpunktwerten gebildeten Risikogruppen unterschiedlicher Belastung im DTK.

Tabelle 8.11: Verteilung der überalterten Mädchen und Jungen auf die nach DTK-Gesamtpunktwerten gebildeten depressionsspezifischen Risikogruppen

		depBel 1	depBel 2	Gesamt
Überalterte Jungen	Anzahl	20	17	37
	Prozent	54,1%	45,9%	100%
Überalterte Mädchen	Anzahl	20	14	34
	Prozent	58,8%	41,2%	100%
Gesamt	Anzahl	40	31	71
	Prozent	56,3%	43,7%	100%

Nach einer Prüfung über χ^2 ergibt sich, dass die in Tabelle 8.11 dargestellte Verteilung nicht signifikant ist [$\chi^2(1) = 0,164$; $p = 0,686$]. Die H₀ sollte demnach beibehalten werden. Dieses Ergebnis führt dann zu folgender Aussage:

8.63 AS: Die Verteilung der überalterten Mädchen und Jungen auf die DTK-Gruppen niedriger (depBel 1) und hoher (depBel 2) depressiver Belastung ist im Gegensatz zu dem Befund für die Gesamtgruppe dieser Altersspanne- nicht geschlechtsspezifisch signifikant unterschiedlich.

Dieses Ergebnis könnte ein Hinweis darauf sein, dass bei speziellen Gruppen von depressiv hochbelasteten Kindern die üblichen geschlechtsspezifischen Unterschiede schwächer werden. Eben deshalb sollte diesem Ergebnis ganz besondere Bedeutung beigemessen werden. Dies eben nicht in der Annahme, mit den Daten seien grundlegende Einsichten gewonnen worden, sondern wegen des begründbaren Verdachts, diese besonderen Daten könnten die Spur einer bislang nicht beachteten Beziehung sein. Sie sollte verfolgt werden.

Die besondere Konstruktion des DTK, depressive Belastungen durch drei Dimensionen zu erfassen, bietet darüber hinaus die Möglichkeit zu untersuchen, ob die Teilstichprobe der überalterten Kinder in einer oder in mehreren DTK-Dimensionen stärker als nicht überalterte Jungen und Mädchen belastet seien.

Tabelle 8.12: Verteilungsparameter für die Teilgruppen alterskonformer und überalterter Jungen und Mädchen in Gymnasialklassen

DTK-Skala	Teilgruppe	Statistische Kennwerte			T-test	
		n	M	SA	T	p
1. Dysphorie / Selbstwertprobleme	Altersgerecht	959	8,4004	5,08	-2,322	,020
	Überaltert	71	9,8592	5,52		
2. Agitiertes Verhalten	Altersgerecht	955	5,1288	3,35	-3,115	,002
	Überaltert	72	6,4028	3,23		
3. Müdigkeit / autonome Reaktionen	altersgerecht	960	5,1625	2,75	-1,820	,069
	Überaltert	72	5,7778	2,99		

Zeichenerklärung: M= Mittelwert, SA= Standardabweichung

Die Durchführung des nonparametrischen Verfahrens Mann-Whitney Test bestätigt die mittels t-Test gefundenen Signifikanzen.

8.64 AS: Die überalterten Kinder aus Gymnasialklassen berichten im DTK durchschnittlich signifikant mehr Symptome von Dysphorie und Selbstwertproblemen, aber auch über mehr Agitiertheit als alterskonforme Mitschüler.

Untersucht wird nun, ob und in einem solchen Fall dann welche geschlechtsspezifischen Unterschiede im Hinblick auf die von überalterten Kindern und Jugendliche erbrachten Mittelwerte in den drei DTK-Subskalen bestehen.

Tabelle 8.13: Mittelwerte und Standardabweichungen auf den drei DTK-Subskalen für überalterte Jungen und Mädchen (nebst t-Test für Geschlechtsunterschiede)

DTK-Skala	Teilgruppe	Statistische Kennwerte			T-test	
		n	M	SA	t	p
1. Dysphorie / Selbstwertprobleme	Jungen	37	8,8919	5,52662	-1,555	,124
	Mädchen	34	10,9118	5,40128		
2. Agitiertes Verhalten	Jungen	38	7,1316	3,01502	-2,070	,042
	Mädchen	34	5,5882	3,31313		
3. Müdigkeit / autonome Reaktionen	Jungen	38	5,3421	2,91572	-1,312	,194
	Mädchen	34	6,2647	3,04819		

Zeichenerklärung: M= Mittelwert, SA= Standardabweichung

Die in Tabelle 8.13 gefundenen Signifikanzen werden auch durch den Mann-Whitney Test bestätigt.

8.65 AS: Die überalterten Jungen erreichen tendenziell signifikant höhere Mittelwerte in der DTK-Dimension „Agitiertes Verhalten“ als die überalterten Mädchen. Dieser Befund wiederholt möglicherweise die bereits gefundenen allgemeinen Unterschiede zwischen Jungen und Mädchen (siehe Kapitel 8.2).

Da innerhalb der durchgeführten Untersuchung eine Teilmenge der beteiligten Lehrkräfte bereit war, die Leistungen der Schüler ihrer Klasse über einen Vergleich mit dem durch die Lehrkräfte wahrgenommenen allgemeinen Leistungsstand der jeweiligen Schulklasse einzustufen, können

auch die so gewonnenen Informationen mit der Gruppe überalterter Schülerinnen und Schüler in Beziehung gesetzt werden.

Tabelle 8.14: Von Lehrkräften gegebene Leistungsbewertungen für die in ihren Schulklassen alterskonforme oder überalterte Schülerinnen und Schüler

			Leistung					Total
			0	1	2	3	4	
ueber2grup	Alterskonform	Count	46	112	194	261	349	962
		Expected Count	47,4	126,5	200,0	257,7	330,3	962,0
		% within ueber2grup	4,8%	11,6%	20,2%	27,1%	36,3%	100,0%
		Adjusted Residual	-,8	-5,3	-1,8	,9	4,8	
	Überaltert	Count	5	24	21	16	6	72
		Expected Count	3,6	9,5	15,0	19,3	24,7	72,0
		% within ueber2grup	6,9%	33,3%	29,2%	22,2%	8,3%	100,0%
		Adjusted Residual	,8	5,3	1,8	-,9	-4,8	
Total	Count	51	136	215	277	355	1034	
	Expected Count	51,0	136,0	215,0	277,0	355,0	1034,0	
	% within ueber2grup	4,9%	13,2%	20,8%	26,8%	34,3%	100,0%	

Zeichenerklärung: alterskonform = der Altersnorm in der Schulklasse entsprechend; überaltert = älter als Altersnorm in der Schulklasse; 0 = keine Information, 1: viel schlechtere Leistung als die durchschnittliche Leistung der Klasse, 2: schlechtere Leistung als die durchschnittliche Leistung der Klasse, 3: durchschnittliche Leistung, 4: bessere als die durchschnittliche Leistung der Klasse.

Chi-Square Tests

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	43,046(a)	4	,000
Likelihood Ratio	42,348	4	,000
Linear-by-Linear Association	34,898	1	,000
N of Valid Cases	1034		

1 cell (10,0%) have expected count less The minimum expected count is 3,55.

Die oben stehende Tabelle erweist sich als insgesamt hoch signifikant. Danach dürfen auch die korrigierten standardisierten Residuen beachtet werden, die Aussagen über einzelne Zellen der Tabelle gestatten. Obschon eine Teilmenge der Lehrkräfte sich nicht bereit fand, die erbetenen Einstufungen über die Leistungen ihrer Schülerinnen und Schüler vorzunehmen, ist die Tabelle aussagekräftig. Aus der Tabelle sind die nachstehenden Aussagen abgeleitet worden.

8.66 AS: Für die Verteilung der von den Lehrkräften in dieser Untersuchung gegebenen und auf die jeweilige Klasse bezogenen Leistungsbeurteilungen und deren Zuordnung zur Gruppe der altersgerecht in der Klasse befindlichen Schüler und Schülerinnen einerseits oder der Zugehörigkeit zu den überalterten in der Klasse bestehen signifikante Abweichungen von Zufallswerten.

8.67 AS: Die alterskonformen Jungen und Mädchen sind in der niedrigsten Leistungsstufe unterrepräsentiert.

8.68 AS: Die alterskonformen Jungen und Mädchen finden sich in der höchsten Leistungsstufe überrepräsentiert.

8.69 AS: Die überalterten Schülerinnen und Schüler sind in der untersten Leistungsstufe signifikant häufiger als erwartet aufzufinden.

8.70 AS: Die überalterten Schülerinnen und Schüler sind in der höchsten Leistungsstufe signifikant unterrepräsentiert und somit seltener als erwartet aufzufinden.

8.71 AS: Etwa 2/3 der überalterten Schülerinnen und Schüler sind in den beiden unteren von 4 Leistungsstufen positioniert (62,5%).

8.72 AS: Die Unterrepräsentation der überalterten Jungen und Mädchen in der höchsten Leistungsstufe entspricht etwa 25% des statistischen Erwartungswertes.

8.73 AS: Für die Teilgruppe der in ihren Schulklassen überalterten Jungen und Mädchen darf deren nach Aussagen ihrer Lehrkräfte bestehende schulische Leistungsschwäche, von individuellen Ausnahmen abgesehen, als erwiesen gelten.

8.74 AS: Mit der unter 8. mitgeteilten Feststellung wird ein deutlicher Zusammenhang zwischen höheren depressiven Belastungen und dem schulischen Misserfolg überalterter Schüler und Schülerinnen als gegeben dargestellt.

Mit den hier eben bekannt werdenden Ergebnissen aus der Analyse von überalterte Schüler und Schülerinnen betreffen Daten bestätigt sich die weiter oben geäußerte Vermutung, dass überalterte und durch schulischen Misserfolg belastete Jungen und Mädchen mit erheblichen Risikofaktoren für die Entwicklung höherer Ausprägungen depressiver Störungen belastet sind. Die Klärung der Verursachungsstruktur für diese Belastung wäre von therapeutischer wie auch von pädagogischer Bedeutung. Falls beispielsweise das Erleben von Hilflosigkeit ein gewichtiger Faktor sein sollte, könnten therapeutische oder pädagogische Interventionen erfolgversprechende Ansätze finden.

8.7 Die Bedeutung des Wohnortes für die Variation von DTK Gesamtpunktwerten

Alle Jungen und Mädchen in der untersuchten griechischen Stichprobe leben in einem Heimatort. Wenn der als Heimatort bezeichnet wird, repräsentiert er im Deutschen eine andere Bedeutung als sie etwa mit der Bezeichnung Wohnort verbunden wird. Indem darauf verwiesen wird, dient das der Verdeutlichung der Schwierigkeiten bei der Bestimmung von Kennungen für Untergruppen der untersuchten Stichprobe. Indem der jeweilige Wohnort als Kennwert für die Gruppierung aller in diesem Ort lebenden Mädchen und Jungen benutzt wird, ergibt sich daraus ein impliziter Hinweis auf eine psychisch und psychologisch bedeutsame Gemeinsamkeit, eben die des Wohnortes. Ob der Wohnort eine Bedeutung für das Auftreten der individuellen Punktwerte im DTK haben könnte, soll die dafür einzusetzende Datenanalyse prüfen. Fraglich ist nun aber, ob sie den Wohnort prüft oder etwa die jeweilige Schule, denn beide sind lokal identisch. Der Ort hat eine Schule, und die Schule ist in diesem Ort. In Orten mit zwei Schulen verändert sich die Sachlage, dort scheint entscheidbar, ob Variationen von Daten mit dem Ort oder mit der Schule verbunden sind. Vielleicht haben beide Anteil an etwaigen Variationen

Im Folgenden wird zunächst der Frage nachgegangen, ob die DTK Messungen von Schulstandort zu Schulstandort, von Wohnort zu Wohnort signifikant unterschiedliche Werte erbringen.

Sollten mit den Variationen der Wohnorte signifikante Variationen der DTK Messwerte auftreten, würde das möglicherweise auch zu unterschiedlichen Einschätzungen der örtlich nachweisbaren depressionsspezifischen Belastungen führen. Als Indikatoren für solche Belastungen werden in dieser Untersuchung, wie zuvor ausführlich dargestellt und kommentiert, die nach DTK-Gesamtpunktwerten vollzogenen Zuweisungen der Mädchen und Jungen aus der untersuchten Stichprobe 2034GR zu definierten Risikogruppen depBel 1 und depBel 2 benutzt. Den zur Teilstichprobe depBell zusammengefassten Jungen und Mädchen wird eine geringere Wahrscheinlichkeit für das aktuelle Bestehen einer klinisch ausgeprägten depressiven Störungen

zugesprochen als denen aus der Teilstichprobe depBel 2. Die günstigere Prognose für die Teilstichprobe depBel1 beruht auf Einschätzungen, denen entsprechende Messdaten aus verschiedenen Untersuchungen zu Grunde liegen. Diese Prognosen für die beiden Teilstichproben ersetzen keine Diagnosen. Sie sind ausschließlich Wahrscheinlichkeiten für das Auftreten klinischer Ausprägungen depressiver Störungen und allesamt über andere Untersuchungen empirisch bestimmt worden. Damit haben die mit der Zuweisung zu den beiden definierten Belastungsgruppen getroffenen Prognosen einen empirisch fundierten Aussagewert, der mit den in der Untersuchung 2034GR gegebenen Bedingungen übereinstimmen könnte, dies aber keineswegs muss.

Die Zugehörigkeit zu den Gruppen depBel 1 und depBel 2 trennt Gruppen, in denen mit an Sicherheit grenzender Wahrscheinlichkeit auf die Gruppengrößen bezogene und signifikant unterschiedliche Prozentsätze im Anteil von Kindern mit und ohne depressiven Störungen in klinischer Ausprägung beobachtbar sein werden. Der jeweils gewünschte Nachweis wäre über hinreichend valide Diagnosen zu erbringen. Die nach DTK Gesamtpunktwerten vollzogene Zuweisung zu den Risikogruppen hat einen innerhalb bestimmbarer Fehlergrenzen gültigen prognostischen Wert. Er orientiert gegenwärtig und das heißt nach dem gegenwärtigen Informationsstand über depressive Belastungen griechischer Schulkinder verlässlicher als jede andere Vorhersage.

Da die Zuweisung zu den Belastungsgruppen die von der Untersuchung angestrebte Information ist, wird die nachfolgende auf die Variation der DTK Messwerte mit den Wohnorten gerichtete statistische Prüfung unter Aussparung der dahin führenden Zwischenstufen an Analysearbeit auf die definierten Belastungsgruppen bezogen.

Mit der nächsten Tabelle werden die mit den Wohnorten auftretenden Variationen in der Verteilung der Jungen und Mädchen auf DTK Belastungsgruppen mitgeteilt und auf Signifikanz geprüft.

Tabelle 8.15: Mit den Untersuchungsorten variierende Verteilung der Mädchen und Jungen auf die Risikogruppen depBel 1 und depBel 2

Stadt		depBel 1	depBel 2	Gesamt
Veria	Count	146	44	190
	Expected Count	140,7	49,3	190,0
	% within city	76,8%	23,2%	100,0%
	% within dtk2gr17	9,8%	8,5%	9,5%
	% of Total	7,3%	2,2%	9,5%
	Adjusted Residual	,9	-,9	
Thessaloniki	Count	181	74	255
	Expected Count	188,9	66,1	255,0
	% within city	71,0%	29,0%	100,0%
	% within dtk2gr17	12,2%	14,2%	12,7%
	% of Total	9,0%	3,7%	12,7%
	Adjusted Residual	-1,2	1,2	
Galatista	Count	99	33	132
	Expected Count	97,8	34,2	132,0
	% within city	75,0%	25,0%	100,0%
	% within dtk2gr17	6,7%	6,3%	6,6%
	% of Total	4,9%	1,6%	6,6%
	Adjusted Residual	,3	-,3	

(Fortführung Tabelle 8.15)

Stadt		depBel 1	depBel 2	Gesamt
Neo Agioneri	Count	60	32	92
	Expected Count	68,1	23,9	92,0
	% within city	65,2%	34,8%	100,0%
	% within dtk2gr17	4,0%	6,2%	4,6%
	% of Total	3,0%	1,6%	4,6%
Athen	Adjusted Residual	-2,0	2,0	
	Count	321	89	410
	Expected Count	303,7	106,3	410,0
	% within city	78,3%	21,7%	100,0%
	% within dtk2gr17	21,6%	17,1%	20,4%
Ksanthi	% of Total	16,0%	4,4%	20,4%
	Adjusted Residual	2,2	-2,2	
	Count	141	41	182
	Expected Count	134,8	47,2	182,0
	% within city	77,5%	22,5%	100,0%
Stavroupoli	% within dtk2gr17	9,5%	7,9%	9,1%
	% of Total	7,0%	2,0%	9,1%
	Adjusted Residual	1,1	-1,1	
	Count	52	34	86
	Expected Count	63,7	22,3	86,0
Rethymno	% within city	60,5%	39,5%	100,0%
	% within dtk2gr17	3,5%	6,5%	4,3%
	% of Total	2,6%	1,7%	4,3%
	Adjusted Residual	-2,9	2,9	
	Count	99	38	137
Episkopi	Expected Count	101,5	35,5	137,0
	% within city	72,3%	27,7%	100,0%
	% within dtk2gr17	6,7%	7,3%	6,8%
	% of Total	4,9%	1,9%	6,8%
	Adjusted Residual	-,5	,5	
Heraklion	Count	112	36	148
	Expected Count	109,6	38,4	148,0
	% within city	75,7%	24,3%	100,0%
	% within dtk2gr17	7,5%	6,9%	7,4%
	% of Total	5,6%	1,8%	7,4%
Karditsa	Adjusted Residual	,5	-,5	
	Count	79	54	133
	Expected Count	98,5	34,5	133,0
	% within city	59,4%	40,6%	100,0%
	% within dtk2gr17	5,3%	10,4%	6,6%
Itea	% of Total	3,9%	2,7%	6,6%
	Adjusted Residual	-4,0	4,0	
	Count	109	21	130
	Expected Count	96,3	33,7	130,0
	% within city	83,8%	16,2%	100,0%
Gesamt	% within dtk2gr17	7,3%	4,0%	6,5%
	% of Total	5,4%	1,0%	6,5%
	Adjusted Residual	2,6	-2,6	
	Count	86	24	110
	Expected Count	81,5	28,5	110,0
Gesamt	% within city	78,2%	21,8%	100,0%
	% within dtk2gr17	5,8%	4,6%	5,5%
	% of Total	4,3%	1,2%	5,5%
	Adjusted Residual	1,0	-1,0	
	Count	1485	520	2005
Gesamt	Expected Count	1485,0	520,0	2005,0
	% within city	74,1%	25,9%	100,0%
	% within dtk2gr17	100,0%	100,0%	100,0%
	% of Total	74,1%	25,9%	100,0%

(Fortführung Tabelle 8.15)

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	41,802(a)	11	,000
Likelihood Ratio	40,397	11	,000
Linear-by-Linear Association	,003	1	,956
N of Valid Cases	2005		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 22,30.

Symmetric Measures

	Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Nominal by Nominal				
Phi	,144			,000
Cramer's V	,144			,000
Contingency Coefficient	,143			,000
Interval by Interval				
Pearson's R	-,001	,022	-,055	,956(c)
Ordinal by Ordinal				
Spearman Correlation	,000	,022	,008	,993(c)
N of Valid Cases	2005			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

c Based on normal approximation.

Die Tabelle der mit den Wohnorten variierenden Zuweisung von Schulkindern zu den definierten Risikogruppen ist hoch signifikant. Damit sind die nachstehenden Aussagen zu formulieren.

8.75 AS: Die mit den Wohnorten der untersuchten Schulkinder auftretende Variation in der Zuweisung von Kindern zu DTK Belastungsgruppen (depBel 1 und depBel 2) weist hoch signifikante Unterschiede auf. Die Unterschiede sollten nicht als zufällig auftretend angesehen werden.

8.76 AS: In einigen Wohnorten entsprechen die beobachteten Anteile an Schulkindern in den beiden Belastungsgruppen den nach Randsummen bestimmten Erwartungswerten. Die Risiken zu einer der beiden Belastungsgruppen zu gehören, sind in diesen Orten so, wie nach den Werten für die Gesamtstichprobe zu erwarten. Diese Orte sind Veria, Thessaloniki, Galatista, Athen, Ksanthi, Rethymno, Episkopi, Itea.

8.77 AS: In anderen Wohnorten ist die Wahrscheinlichkeit der Zugehörigkeit zu einer von beiden Belastungsgruppen signifikant von der für die Gesamtheit geltenden Wahrscheinlichkeit hoch signifikant unterschieden. Diese Orte sind Athen, Neo Agioneri, Stavroupoli, Heraklion, Karditsa.

8.78 AS: In Neo Agioneri, Stavroupoli und Heraklion ist die Wahrscheinlichkeit in die höhere Belastungsgruppe depBel 2 zu kommen, signifikant größer als nach den Werten für die Gesamtstichprobe zu erwarten. Für diese Schulstandorte könnten ortsgebundene Bedingungen gelten, mit denen sich die Wahrscheinlichkeit für die Ausbildung depressiver Störungen erhöht.

8.79 AS: In Athen und Karditsa ist die Wahrscheinlichkeit zur Gruppe der niedriger belasteten in depBel 1 zu gehören, signifikant höher als nach den Werten für die Gesamtstichprobe zu erwarten wäre.

8.80 AS: Die gefundenen Abweichungen lassen sich auch durchgehend auf die für den Ort geltenden Erwartungswerte an Schulkindern für die Belastungsgruppe depBel 2 beziehen. So geordnet besagen die jeweils zu errechnenden Prozentsätze, um wie viel Prozent in dem betreffenden Ort die Erwartungswerte für die Zugehörigkeit zur Gruppe depBel 2 überschritten

oder unterschritten werden. Da diese Belastungsgruppe die mit dem höher eingeschätzten Risiko für das Auftreten klinischer Ausprägungen depressiver Störungen ist, lässt sich der mitzuteilende Wert auch darauf beziehen. Der Prozentsatz gibt an, um wie viel Prozent die Erwartungswerte für Kinder mit sehr wahrscheinlich klinisch beachtenswerten Ausprägungen depressiver Störungen überschritten oder unterschritten werden. Für die nachstehend aufgeführten Wohnorte gelten die errechneten Prozentsätze: Überschritten werden die Erwartungswerte in Neo Agioneri um 33,9%, in Stavroupoli um 52,5% und in Heraklion um 56,5%. Unter den Erwartungswerten bleiben die Werte für Athen um 16,3% und für Karditsa um 37,7%.

8.81 AS: Die Gründe für die signifikante Abweichung der Wahrscheinlichkeiten für die Zuweisung eines Schulkindes zur Gruppe mit den geringeren oder höheren Belastungsrisiken für das Auftreten depressiver Störungen mit klinischen Ausprägungen werden über die dargestellte Tabelle nicht erfasst.

8.82 AS: Nach den bereits bekannt gewordenen Ergebnissen über die Variation der DTK Gesamtpunktwerte mit Alter und Geschlecht könnten von der Gesamtverteilung abweichende Wahrscheinlichkeiten von örtlichen Abweichungen in der Zusammensetzung der Schülerschaft nach Alter oder Geschlecht oder nach beidem herrühren. Ebenso wäre möglich, dass sich die psycho-sozialen Lebensbedingungen der Orte oder die depressionsspezifischen Qualitäten der Schulen voneinander unterscheiden. Weitere Bedingungen für das Auftreten der Abweichungen sind denkbar.

Ob die mit der Aussage 8.82 als mögliche Ursachen für die Variation der Anteile an Kinder der Stichprobe in den Risikogruppen depBel 1 und depBel 2 genannten Bedingungen weiterhin verdächtig werden sollten, prüft der nächste Analyseschritt. Mit ihm wird festzustellen versucht, ob die am Wohnort zu beobachtenden signifikanten Abweichungen in der beobachteten Größenordnung für alle geschlechtsspezifischen Altersgruppen gelten.

Tabelle 8.16: Geschlechtsspezifische Verteilung der jüngeren Kinder der Stichprobe (Grundschul Kinder) auf die Risikogruppen depBel 1 und depBel 2.

Stadt			depBel 1	depBel 2	Gesamt
Veria	Jungen	Count	50	13	63
		Expected Count	51,2	11,8	63,0
		% within sexRpoleis	79,4%	20,6%	100,0%
		% within dtk2gr17	6,3%	7,1%	6,4%
		% of Total	5,1%	1,3%	6,4%
		Adjusted Residual	-,4	,4	
	Mädchen	Count	41	9	50
		Expected Count	40,6	9,4	50,0
		% within sexRpoleis	82,0%	18,0%	100,0%
		% within dtk2gr17	5,2%	4,9%	5,1%
% of Total		4,2%	,9%	5,1%	
	Adjusted Residual	,1	-,1		
Thessaloniki	Jungen	Count	49	17	66
		Expected Count	53,6	12,4	66,0
		% within sexRpoleis	74,2%	25,8%	100,0%
		% within dtk2gr17	6,2%	9,2%	6,7%
		% of Total	5,0%	1,7%	6,7%
		Adjusted Residual	-1,5	1,5	
	Mädchen	Count	55	12	67
		Expected Count	54,4	12,6	67,0
		% within sexRpoleis	82,1%	17,9%	100,0%
		% within dtk2gr17	6,9%	6,5%	6,8%
% of Total		5,6%	1,2%	6,8%	
	Adjusted Residual	,2	-,2		

(Fortführung Tabelle 8.16)

Stadt			depBel 1	depBel 2	Gesamt
Galatista	Jungen	Count	25	8	33
		Expected Count	26,8	6,2	33,0
		% within sexRpoleis	75,8%	24,2%	100,0%
		% within dtk2gr17	3,1%	4,3%	3,4%
		% of Total	2,6%	,8%	3,4%
	Adjusted Residual	-,8	,8		
	Mädchen	Count	24	5	29
		Expected Count	23,6	5,4	29,0
		% within sexRpoleis	82,8%	17,2%	100,0%
		% within dtk2gr17	3,0%	2,7%	3,0%
% of Total		2,4%	,5%	3,0%	
Adjusted Residual	,2	-,2			
Neo Agioneri	Jungen	Count	19	5	24
		Expected Count	19,5	4,5	24,0
		% within sexRpoleis	79,2%	20,8%	100,0%
		% within dtk2gr17	2,4%	2,7%	2,4%
		% of Total	1,9%	,5%	2,4%
	Adjusted Residual	-,3	,3		
	Mädchen	Count	4	6	10
		Expected Count	8,1	1,9	10,0
		% within sexRpoleis	40,0%	60,0%	100,0%
		% within dtk2gr17	,5%	3,3%	1,0%
% of Total		,4%	,6%	1,0%	
Adjusted Residual	-3,4	3,4			
Athen	Jungen	Count	92	29	121
		Expected Count	98,3	22,7	121,0
		% within sexRpoleis	76,0%	24,0%	100,0%
		% within dtk2gr17	11,6%	15,8%	12,3%
		% of Total	9,4%	3,0%	12,3%
	Adjusted Residual	-1,6	1,6		
	Mädchen	Count	93	17	110
		Expected Count	89,3	20,7	110,0
		% within sexRpoleis	84,5%	15,5%	100,0%
		% within dtk2gr17	11,7%	9,2%	11,2%
% of Total		9,5%	1,7%	11,2%	
Adjusted Residual	,9	-,9			
Ksanthi	Jungen	Count	39	10	49
		Expected Count	39,8	9,2	49,0
		% within sexRpoleis	79,6%	20,4%	100,0%
		% within dtk2gr17	4,9%	5,4%	5,0%
		% of Total	4,0%	1,0%	5,0%
	Adjusted Residual	-,3	,3		
	Mädchen	Count	42	7	49
		Expected Count	39,8	9,2	49,0
		% within sexRpoleis	85,7%	14,3%	100,0%
		% within dtk2gr17	5,3%	3,8%	5,0%
% of Total		4,3%	,7%	5,0%	
Adjusted Residual	,8	-,8			
Stavroupoli	Jungen	Count	21	3	24
		Expected Count	19,5	4,5	24,0
		% within sexRpoleis	87,5%	12,5%	100,0%
		% within dtk2gr17	2,6%	1,6%	2,4%
		% of Total	2,1%	,3%	2,4%
	Adjusted Residual	,8	-,8		
	Mädchen	Count	10	6	16
		Expected Count	13,0	3,0	16,0
		% within sexRpoleis	62,5%	37,5%	100,0%
		% within dtk2gr17	1,3%	3,3%	1,6%
% of Total		1,0%	,6%	1,6%	
Adjusted Residual	-1,9	1,9			

(Fortführung Tabelle 8.16)

Stadt			depBel 1	depBel 2	Gesamt
Rethymno	Jungen	Count	28	6	34
		Expected Count	27,6	6,4	34,0
		% within sexRpoleis	82,4%	17,6%	100,0%
		% within dtk2gr17	3,5%	3,3%	3,5%
		% of Total	2,9%	,6%	3,5%
	Adjusted Residual	,2	-,2		
	Mädchen	Count	22	2	24
		Expected Count	19,5	4,5	24,0
		% within sexRpoleis	91,7%	8,3%	100,0%
		% within dtk2gr17	2,8%	1,1%	2,4%
% of Total		2,2%	,2%	2,4%	
Adjusted Residual	1,3	-1,3			
Episkopi	Jungen	Count	26	5	31
		Expected Count	25,2	5,8	31,0
		% within sexRpoleis	83,9%	16,1%	100,0%
		% within dtk2gr17	3,3%	2,7%	3,2%
		% of Total	2,7%	,5%	3,2%
	Adjusted Residual	,4	-,4		
	Mädchen	Count	29	2	31
		Expected Count	25,2	5,8	31,0
		% within sexRpoleis	93,5%	6,5%	100,0%
		% within dtk2gr17	3,6%	1,1%	3,2%
% of Total		3,0%	,2%	3,2%	
Adjusted Residual	1,8	-1,8			
Heraklion	Jungen	Count	23	5	28
		Expected Count	22,7	5,3	28,0
		% within sexRpoleis	82,1%	17,9%	100,0%
		% within dtk2gr17	2,9%	2,7%	2,9%
		% of Total	2,3%	,5%	2,9%
	Adjusted Residual	,1	-,1		
	Mädchen	Count	15	10	25
		Expected Count	20,3	4,7	25,0
		% within sexRpoleis	60,0%	40,0%	100,0%
		% within dtk2gr17	1,9%	5,4%	2,6%
% of Total		1,5%	1,0%	2,6%	
Adjusted Residual	-2,8	2,8			
Karditsa	Jungen	Count	29	0	29
		Expected Count	23,6	5,4	29,0
		% within sexRpoleis	100,0%	,0%	100,0%
		% within dtk2gr17	3,6%	,0%	3,0%
		% of Total	3,0%	,0%	3,0%
	Adjusted Residual	2,6	-2,6		
	Mädchen	Count	34	2	36
		Expected Count	29,2	6,8	36,0
		% within sexRpoleis	94,4%	5,6%	100,0%
		% within dtk2gr17	4,3%	1,1%	3,7%
% of Total		3,5%	,2%	3,7%	
Adjusted Residual	2,1	-2,1			
Itea	Jungen	Count	12	3	15
		Expected Count	12,2	2,8	15,0
		% within sexRpoleis	80,0%	20,0%	100,0%
		% within dtk2gr17	1,5%	1,6%	1,5%
		% of Total	1,2%	,3%	1,5%
	Adjusted Residual	-,1	,1		
	Mädchen	Count	14	2	16
		Expected Count	13,0	3,0	16,0
		% within sexRpoleis	87,5%	12,5%	100,0%
		% within dtk2gr17	1,8%	1,1%	1,6%
% of Total		1,4%	,2%	1,6%	
Adjusted Residual	,6	-,6			
Gesamt	Count	796	184	980	
	Expected Count	796,0	184,0	980,0	
	% within sexRpoleis	81,2%	18,8%	100,0%	
	% within dtk2gr17	100,0%	100,0%	100,0%	
	% of Total	81,2%	18,8%	100,0%	

(Fortführung Tabelle 8.16)

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	45,804(a)	23	,003
Likelihood Ratio	48,629	23	,001
Linear-by-Linear Association	5,286	1	,022
N of Valid Cases	980		

a 8 cells (16,7%) have expected count less than 5. The minimum expected count is 1,88.

Symmetric Measures					
		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Nominal by Nominal	Phi	,216			,003
	Cramer's V	,216			,003
	Contingency Coefficient	,211			,003
Interval by Interval	Pearson's R	-,073	,030	-2,304	,021(c)
Ordinal by Ordinal	Spearman Correlation	-,078	,031	-2,451	,014(c)
N of Valid Cases		980			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

c Based on normal approximation.

Tabelle 8.17: Übersicht über die an den Wohnorten beobachtete Über- und Unterschreitung der Erwartungswerte für Kinder aus Grundschulen in der Risikogruppe depBel2

Stadt	Erwartungswerte für depBel 2	
	über-/unterschritten oder eingestellt (+/-/=)	
Veria	Jungen	=
	Mädchen	=
Thessaloniki	Jungen	=
	Mädchen	=
Galatista	Jungen	=
	Mädchen	=
Neo Agioneri	Jungen	=
	Mädchen	+ 216%
Athen	Jungen	=
	Mädchen	=
Ksanthi	Jungen	=
	Mädchen	=
Stavroupoli	Jungen	=
	Mädchen	=
Rethymno	Jungen	=
	Mädchen	=
Episkopi	Jungen	=
	Mädchen	=
Heraklion	Jungen	=
	Mädchen	+112,8%
Karditsa	Jungen	-100%
	Mädchen	-70,8%
Itea	Jungen	=
	Mädchen	=

Zeichenerklärung: = Einstellung der Erwartungswerte, keine signifikante Abweichung; + = signifikante Überschreitung der Erwartungswerte; - = signifikante Unterschreitung der Erwartungswerte

Die Übersichtstabelle 8.17 zeigt an, dass in mehreren Orten bei den Kindern aus Grundschulen signifikante geschlechtsspezifische Unterschiede in der Einstellung der Erwartungswerte und deren Über- oder Unterschreitung auftreten. Nach dieser Tabelle treffen die nachstehenden Aussagen zu.

8.83 AS: In drei Städten finden sich bei den Grundschulern in deren geschlechtsspezifischer Verteilung auf die Risikogruppen depBel 1 und depBel 2 signifikante Abweichungen von den nach den Daten der Gesamtstichprobe berechneten Erwartungswerten.

8.84 AS: Die festgestellten Abweichungen könnten im Zusammenhang mit den für Neo Agioneri, Heraklion und Karditsa schon zuvor nachgewiesenen Abweichungen der nach Wohnorten geordneten Verteilung aller Mitglieder der Stichprobe 2034GR auf die Risikogruppen stehen.

8.85 AS: In Heraklion und Neo Agioneri besteht für Mädchen im Grundschulalter eine beachtenswerte Erhöhung des Risikos einen hohen DTK Gesamtpunkt看wert zu erreichen, der die Zuweisung zur Risikogruppe depBel 2 nach sich zieht.

8.86 AS: Der Wohnort Karditsa ist eine hoch signifikante Ausnahme von den für die Gesamtstichprobe geltenden Wahrscheinlichkeiten für die Verteilung der Grundschul Kinder auf die Risikogruppen depBel 1 und depBel 2. Die beobachteten Werte für die Mädchen liegen 7ß% unter den Erwartungswerten. Für die Jungen der Grundschule besteht in Karditsa entgegen den Erwartungswerten nach den beobachteten Werten für die Zuweisung zur Belastungsgruppe depBel 2 eine Wahrscheinlichkeit von 0.

Die erfolgte Überprüfung der nach Wohnorten bestimmten geschlechtsspezifischen Verteilung der Grundschüler auf die definierten Belastungsgruppen wird nachfolgend durch die entsprechende Prüfung der Verteilung aller Jungen und Mädchen aus den Gymnasien der Stichprobe ergänzt.

Tabelle 8.18: Geschlechtsspezifische Verteilung der älteren Jungen und Mädchen der Stichprobe (Gymnasien) auf die Risikogruppen depBel 1 und depBel 2

Stadt			depBel 1	depBel 2	Gesamt
Veria	Jungen	Count	25	5	30
		Expected Count	20,2	9,8	30,0
		% within sexRpoleis	83,3%	16,7%	100,0%
		% within dtk2gr17	3,6%	1,5%	2,9%
		% of Total	2,4%	,5%	2,9%
	Adjusted Residual	1,9	-1,9		
	Mädchen	Count	30	17	47
		Expected Count	31,6	15,4	47,0
		% within sexRpoleis	63,8%	36,2%	100,0%
		% within dtk2gr17	4,4%	5,1%	4,6%
% of Total		2,9%	1,7%	4,6%	
Adjusted Residual	-,5	,5			
Thessaloniki	Jungen	Count	36	16	52
		Expected Count	35,0	17,0	52,0
		% within sexRpoleis	69,2%	30,8%	100,0%
		% within dtk2gr17	5,2%	4,8%	5,1%
		% of Total	3,5%	1,6%	5,1%
	Adjusted Residual	,3	-,3		
	Mädchen	Count	41	29	70
		Expected Count	47,1	22,9	70,0
		% within sexRpoleis	58,6%	41,4%	100,0%
		% within dtk2gr17	6,0%	8,6%	6,8%
% of Total		4,0%	2,8%	6,8%	
Adjusted Residual	-1,6	1,6			

(Fortführung Tabelle 8.18)

Stadt			depBel 1	depBel 2	Gesamt
Galatista	Jungen	Count	33	8	41
		Expected Count	27,6	13,4	41,0
		% within sexRpoleis	80,5%	19,5%	100,0%
		% within dtk2gr17	4,8%	2,4%	4,0%
		% of Total	3,2%	,8%	4,0%
	Adjusted Residual	1,8	-1,8		
	Mädchen	Count	17	12	29
		Expected Count	19,5	9,5	29,0
		% within sexRpoleis	58,6%	41,4%	100,0%
		% within dtk2gr17	2,5%	3,6%	2,8%
% of Total		1,7%	1,2%	2,8%	
Adjusted Residual	-1,0	1,0			
Neo Agioneri	Jungen	Count	16	15	31
		Expected Count	20,8	10,2	31,0
		% within sexRpoleis	51,6%	48,4%	100,0%
		% within dtk2gr17	2,3%	4,5%	3,0%
		% of Total	1,6%	1,5%	3,0%
	Adjusted Residual	-1,9	1,9		
	Mädchen	Count	21	6	27
		Expected Count	18,1	8,9	27,0
		% within sexRpoleis	77,8%	22,2%	100,0%
		% within dtk2gr17	3,0%	1,8%	2,6%
% of Total		2,0%	,6%	2,6%	
Adjusted Residual	1,2	-1,2			
Athen	Jungen	Count	86	20	106
		Expected Count	71,3	34,7	106,0
		% within sexRpoleis	81,1%	18,9%	100,0%
		% within dtk2gr17	12,5%	6,0%	10,3%
		% of Total	8,4%	2,0%	10,3%
	Adjusted Residual	3,2	-3,2		
	Mädchen	Count	50	23	73
		Expected Count	49,1	23,9	73,0
		% within sexRpoleis	68,5%	31,5%	100,0%
		% within dtk2gr17	7,3%	6,8%	7,1%
% of Total		4,9%	2,2%	7,1%	
Adjusted Residual	,2	-,2			
Ksanthi	Jungen	Count	37	11	48
		Expected Count	32,3	15,7	48,0
		% within sexRpoleis	77,1%	22,9%	100,0%
		% within dtk2gr17	5,4%	3,3%	4,7%
		% of Total	3,6%	1,1%	4,7%
	Adjusted Residual	1,5	-1,5		
	Mädchen	Count	23	13	36
		Expected Count	24,2	11,8	36,0
		% within sexRpoleis	63,9%	36,1%	100,0%
		% within dtk2gr17	3,3%	3,9%	3,5%
% of Total		2,2%	1,3%	3,5%	
Adjusted Residual	-,4	,4			
Stavroupoli	Jungen	Count	12	11	23
		Expected Count	15,5	7,5	23,0
		% within sexRpoleis	52,2%	47,8%	100,0%
		% within dtk2gr17	1,7%	3,3%	2,2%
		% of Total	1,2%	1,1%	2,2%
	Adjusted Residual	-1,6	1,6		
	Mädchen	Count	9	14	23
		Expected Count	15,5	7,5	23,0
		% within sexRpoleis	39,1%	60,9%	100,0%
		% within dtk2gr17	1,3%	4,2%	2,2%
% of Total		,9%	1,4%	2,2%	
Adjusted Residual	-2,9	2,9			

(Fortführung Tabelle 8.18)

Stadt			depBel 1	depBel 2	Gesamt
Rethymno	Jungen	Count	20	10	30
		Expected Count	20,2	9,8	30,0
		% within sexRpoleis	66,7%	33,3%	100,0%
		% within dtk2gr17	2,9%	3,0%	2,9%
		% of Total	2,0%	1,0%	2,9%
	Adjusted Residual	-,1	,1		
	Mädchen	Count	29	20	49
		Expected Count	32,9	16,1	49,0
		% within sexRpoleis	59,2%	40,8%	100,0%
		% within dtk2gr17	4,2%	6,0%	4,8%
% of Total		2,8%	2,0%	4,8%	
Adjusted Residual	-1,2	1,2			
Episkopi	Jungen	Count	34	14	48
		Expected Count	32,3	15,7	48,0
		% within sexRpoleis	70,8%	29,2%	100,0%
		% within dtk2gr17	4,9%	4,2%	4,7%
		% of Total	3,3%	1,4%	4,7%
	Adjusted Residual	,5	-,5		
	Mädchen	Count	23	15	38
		Expected Count	25,5	12,5	38,0
		% within sexRpoleis	60,5%	39,5%	100,0%
		% within dtk2gr17	3,3%	4,5%	3,7%
% of Total		2,2%	1,5%	3,7%	
Adjusted Residual	-,9	,9			
Heraklion	Jungen	Count	24	14	38
		Expected Count	25,5	12,5	38,0
		% within sexRpoleis	63,2%	36,8%	100,0%
		% within dtk2gr17	3,5%	4,2%	3,7%
		% of Total	2,3%	1,4%	3,7%
	Adjusted Residual	-,5	,5		
	Mädchen	Count	17	25	42
		Expected Count	28,2	13,8	42,0
		% within sexRpoleis	40,5%	59,5%	100,0%
		% within dtk2gr17	2,5%	7,4%	4,1%
% of Total		1,7%	2,4%	4,1%	
Adjusted Residual	-3,8	3,8			
Karditsa	Jungen	Count	25	11	36
		Expected Count	24,2	11,8	36,0
		% within sexRpoleis	69,4%	30,6%	100,0%
		% within dtk2gr17	3,6%	3,3%	3,5%
		% of Total	2,4%	1,1%	3,5%
	Adjusted Residual	,3	-,3		
	Mädchen	Count	21	8	29
		Expected Count	19,5	9,5	29,0
		% within sexRpoleis	72,4%	27,6%	100,0%
		% within dtk2gr17	3,0%	2,4%	2,8%
% of Total		2,0%	,8%	2,8%	
Adjusted Residual	,6	-,6			
Itea	Jungen	Count	29	11	40
		Expected Count	26,9	13,1	40,0
		% within sexRpoleis	72,5%	27,5%	100,0%
		% within dtk2gr17	4,2%	3,3%	3,9%
		% of Total	2,8%	1,1%	3,9%
	Adjusted Residual	,7	-,7		
	Mädchen	Count	31	8	39
		Expected Count	26,2	12,8	39,0
		% within sexRpoleis	79,5%	20,5%	100,0%
		% within dtk2gr17	4,5%	2,4%	3,8%
% of Total		3,0%	,8%	3,8%	
Adjusted Residual	1,7	-1,7			
Gesamt	Count	689	336	1025	
	Expected Count	689,0	336,0	1025,0	
	% within sexRpoleis	67,2%	32,8%	100,0%	
	% within dtk2gr17	100,0%	100,0%	100,0%	
	% of Total	67,2%	32,8%	100,0%	

(Fortführung Tabelle 8.18)

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	57,577(a)	23	,000
Likelihood Ratio	57,542	23	,000
Linear-by-Linear Association	,485	1	,486
N of Valid Cases	1025		

a 0 cells (,0%) have expected count less than 5. The minimum expected count is 7,54.

Symmetric Measures					
		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Nominal by Nominal	Phi	,237			,000
	Cramer's V	,237			,000
	Contingency Coefficient	,231			,000
Interval by Interval	Pearson's R	,022	,031	,696	,486(c)
Ordinal by Ordinal	Spearman Correlation	,032	,031	1,026	,305(c)
N of Valid Cases		1025			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

c Based on normal approximation.

Tabelle 8.19: Übersicht über die an den Wohnorten beobachtete Über- und Unterschreitung der Erwartungswerte für Jungen und Mädchen aus den Gymnasien der Stichprobe in der Risikogruppe depBel2

Stadt		Erwartungswerte für depBel 2 über-/unterschritten oder eingestellt (+/-/=)
Veria	Jungen	=
	Mädchen	=
Thessaloniki	Jungen	=
	Mädchen	=
Galatista	Jungen	=
	Mädchen	=
Neo Agioneri	Jungen	=
	Mädchen	=
Athen	Jungen	- 42,4%
	Mädchen	=
Ksanthi	Jungen	=
	Mädchen	=
Stavroupoli	Jungen	=
	Mädchen	+86,7%
Rethymno	Jungen	=
	Mädchen	=
Episkopi	Jungen	=
	Mädchen	=
Heraklion	Jungen	=
	Mädchen	+ 81,2%
Karditsa	Jungen	=
	Mädchen	=
Itea	Jungen	=
	Mädchen	=

Zeichenerklärung: = Einstellung der Erwartungswerte, keine signifikante Abweichung; + = signifikante Überschreitung der Erwartungswerte; - = signifikante Unterschreitung der Erwartungswerte

Wie aus der Übersicht abzulesen ist, traten bei den Schülerinnen und Schülern an drei Wohnorten signifikante Abweichungen von den auf die Gesamtstichprobe bezogenen örtlichen Erwartungswerten für die geschlechtsspezifische Verteilung auf die Risikogruppen depBel 1 und depBel 2 auf.

Die nachstehenden Aussagen fassen die Ergebnisse zusammen.

8.87 AS: Die geschlechtsspezifische Verteilung der Schülerinnen und Schüler aus Gymnasien auf die definierten Risikogruppen (nach DTK) weicht signifikant von einer Zufallsverteilung ab.

8.88 AS: In drei Städten treten signifikante geschlechtsspezifische Abweichungen von den Erwartungswerten auf.

8.89 AS: In Athen blieben für Jungen die beobachteten Frequenzen in der Risikogruppe depBel 2 um 42,4% unter den Erwartungswerten.

8.90 AS: In Heraklion und in Stavroupoli sind die Mädchen in der Belastungsgruppe depBel 2 stärker vertreten als die Erwartungswerte anzeigten. In Heraklion betrug die Überbesetzung 81,2% und in Stavroupoli 86,7%.

8.91 AS: Die an einigen Wohnorten beobachteten geschlechtsspezifischen Abweichungen bei den Schülerinnen und Schülern aus den Gymnasien in deren Verteilung auf die nach DTK definierten Risikogruppen kann dazu beigetragen haben, dass die zuvor in der Gesamtstichprobe beobachteten signifikanten Abweichungen aufgetreten sind.

8.92 AS: Die Beteiligung weiterer Bedingungen an dem unterschiedlichen Aufkommen von nach DTK als wenig und als stark mit depressionsspezifischen Risiken als belastet bezeichneten Teilgruppen der gesamten Stichprobe ist wahrscheinlich. Dies dürfte insbesondere deswegen zutreffen, weil das Auftreten der signifikanten Variationen mit den Wohnorten zustande gekommen sein dürfte, weil die Orte als Stellvertreter für nicht entdeckte Bedingungen (latente Variablen) auftreten. Solche Bedingungen könnten auch mit den Lebensumständen in den Orten und den Schulen zusammenhängen.

Die Aussagen bekräftigen die mit den Analysen der Daten auftretenden Feststellungen. Sie geben aber keine hinreichenden Informationen über die zu den festgestellten Fakten hinführenden Bedingungen.

Die getrennt nach Jungen und Mädchen geführten Nachforschungen über die an den Wohnorten zu beobachtende Verteilung auf die definierten Risikogruppen hat zu Ergebnissen geführt, die nachstehend für die gesamte Stichprobe zusammengefasst werden. In dieser Tabelle werden die Ergebnisse aus drei Überprüfungen verallgemeinernd wiederholt. Die erste Prüfung betraf die nach Wohnorten geordnete Verteilung der Gesamtstichprobe auf die Risikogruppen, die zweite die geschlechtsspezifische Verteilung für das Grundschulalter und die dritte ergänzte das durch die auf die Gymnasien bezogene Analyse.

Tabelle 8.20: Zusammenfassende Darstellung der mit den Wohnorten variierenden Zuweisungen der Jungen und Mädchen aus der untersuchten Stichprobe auf nach DTK Gesamtpunktwerten definierten Risikogruppen

Ort	Gesamtstichprobe	<u>Grundschul Kinder</u>		<u>Gymnasiasten</u>	
		Jungen	Mädchen	Jungen	Mädchen
Neo Agioneri	Plus 33,89%	<i>Erwartungswerte</i>	Plus 215,79%	<i>Erwartungswerte</i>	<i>Erwartungswerte</i>
Athen	Minus 16,27%	<i>Erwartungswerte</i>	<i>Erwartungswerte</i>	Minus 42,36%	<i>Erwartungswerte</i>
Stavroupoli	Plus 52,47%	<i>Erwartungswerte</i>	<i>Erwartungswerte</i>	<i>Erwartungswerte</i>	Plus 86,67%
Heraklion	Plus 56,52%	<i>Erwartungswerte</i>	Plus 112,7%	<i>Erwartungswerte</i>	Plus 81,16%
Karditsa	Minus 37,69%	Minus 100%	Minus 70,59%	<i>Erwartungswerte</i>	<i>Erwartungswerte</i>

Die nach dieser Aufstellung zu gebenden Aussagen sind zu einem Großteil bereits weiter oben aufgeführt worden und dort nachzulesen. Die Aufstellung gibt aber auch Gelegenheit zu zusammenfassenden Feststellungen.

8.93 AS: Im Vergleich von 12 griechischen Städten stimmen die für den Anteil der Jungen an der nach DTK definierten Risikogruppe depBel 2 (klinisch belastet) beobachteten Frequenzen mit den berechneten Erwartungswerte für 10 Städte überein. In 2 Städten werden signifikante Unterschreitungen der Erwartungswerte festgestellt.

8.94 AS: Die mit den Wohnorten und zwei Altersstufen (Grundschulalter und Gymnasialalter) variierenden Frequenzen der Anteile der Jungen an den Zuweisungen zur Belastungsgruppe depBel 2 übersteigen für alle 12 Städte niemals in signifikanter Weise die Erwartungswerte. In der untersuchten Stichprobe treten keine die Anteile an der Risikogruppe depBel 2 signifikant über Erwartungswerte erhöhenden Bedingungen auf.

8.95 AS: Im Gegensatz zu der absoluten Resistenz der Jungen gegenüber potentiellen Bedingungen zur Erhöhung der Anteile an der Belastungsgruppe depBel 2 treten bei dem Vergleich zwischen 12 griechischen Städten für 4 orte signifikante Überschreitungen der Erwartungswerte auf und eine signifikante Unterbesetzung.

8.96 AS: Nach der Feststellung in Aussage 3 wird vorgeschlagen, die folgende Hypothese neu einzuführen: „Mädchen sind in Grundschulen und Gymnasien sensibel gegenüber mit dem Wohnort verbundenen Bedingungen, die zu signifikanten Abweichungen von den Erwartungswerten für Anteile an der Besetzung der nach DTK bestimmten Belastungsgruppe depBel 2 führen.“

8.97 AS: Die für Jungen und Mädchen im Grundschulalter für den Ort Karditsa beobachteten Werte folgen mit ihrer gegenüber Erwartungswerten signifikanten Unterbesetzung keinem der mit den bisherigen Ergebnissen bekannt gewordenen Determinanten für DTK Gesamtpunktwerte. An dieser Stelle wird angemerkt, dass die mit den beobachteten Ergebnissen entstehende Ungewissheit über deren Bedeutung, durch eine Wiederholung der Untersuchung mit einer neuen örtlichen Stichprobe reduziert werden könnte.

Die mit den Wohnorten und den Altersstufen variierende geschlechtsspezifische Verteilung der Jungen und Mädchen auf die nach DTK Gesamtpunktwerten definierten depressionsspezifischen Belastungsgruppen könnte von der Einwohnerzahl der jeweiligen Städte beeinflusst werden. Ob dies so ist, wird nachfolgend zu prüfen versucht.

Tabelle 8.21: Nach drei über die Einwohnerzahl definierten Stadtgrößen geordnete Verteilung der Kinder und Jugendlichen auf die definierten Belastungsgruppen

		depBel 1	depBel 2	Gesamt
Großstadt	Anzahl	581	217	798
	Prozent	72,8%	27,2%	100%
Mittelstadt	Anzahl	495	144	639
	Prozent	77,5%	22,5%	100%
Kleinstadt	Anzahl	409	159	568
	Prozent	72,0%	28,0%	100%
Gesamt	Anzahl	1485	520	2005
	Prozent	74,1%	25,9%	100%

Die in Tabelle 8.21 dargestellte Verteilung ist nicht signifikant [$\chi^2(2) = 5,755$; $p = 0,056$].

8.98 AS: Die Verteilung der Kinder und Jugendlichen aus der untersuchten Stichprobe auf nach DTK Gesamtpunktwerten definierte Belastungsgruppen variiert nur zufällig mit der Unterscheidung der Orte nach Groß-, Mittel- und Kleinstädten.

Da die Sensibilität gegenüber den potentiell wirkenden Bedingungen für eine signifikante Variation mit den drei Stadttypen altersabhängig sein könnte, werden die entsprechenden Daten nach Grundschule und Gymnasium getrennt verglichen.

Tabelle 8.22: Nach drei Stadtgrößen geordnete Verteilung der Kinder aus den Grundschulen auf die definierten Belastungsgruppen

Wohnort	Geschlecht		depBel 1	depBel 2	Gesamt
Großstadt	Jungen	Anzahl	164	51	215
		Prozent	76,3%	23,7%	100%
	Mädchen	Anzahl	163	39	202
		Prozent	80,7%	19,3%	100%
Mittelstadt	Jungen	Anzahl	146	29	175
		Prozent	83,4%	16,6%	100%
	Mädchen	Anzahl	139	20	159
		Prozent	87,4%	12,6%	100%
Kleinstadt	Jungen	Anzahl	103	24	127
		Prozent	81,1%	18,9%	100%
	Mädchen	Anzahl	81	21	102
		Prozent	79,4%	20,6%	100%
Gesamt	Anzahl	796	184	980	
	Prozent	81,2%	18,8%	100%	

Die in Tabelle 8.22 dargestellte Verteilung ist nicht signifikant [$\chi^2(5) = 8,268$; $p = 0,142$].

8.99 AS: Die geschlechtsspezifische Verteilung von Grundschulkindern aus nach Einwohnerzahl unterschiedenen drei Ortstypen zeigt keine signifikanten Abweichungen. Die Grundschul Kinder entsprechen in ihren Werten der Gesamtverteilung aller Teilnehmer.

Tabelle 8.23: Nach drei Stadtgrößen geordnete geschlechtsspezifische Verteilung der Kinder und Jugendlichen aus den Gymnasien auf die definierten Belastungsgruppen

Wohnort	Geschlecht		depBel 1	depBel 2	Gesamt
Großstadt	Jungen	Anzahl	146 **	50 **	196
		Erwartete Anzahl	131,8	64,2	196
		Prozent	74,5%	25,5%	100%
	Mädchen	Anzahl	108 **	77 **	185
		Erwartete Anzahl	124,4	60,6	185
		Prozent	58,4%	41,6%	100%
Mittelstadt	Jungen	Anzahl	107 **	37 **	144
		Erwartete Anzahl	96,8	47,2	144
		Prozent	74,3%	25,7%	100%
	Mädchen	Anzahl	103	58	161
		Erwartete Anzahl	108,2	52,8	161
		Prozent	64,0%	36,0%	100%
Kleinstadt	Jungen	Anzahl	124	59	183
		Erwartete Anzahl	123,0	60,0	183
		Prozent	67,8%	32,2%	100%
	Mädchen	Anzahl	101	55	156
		Erwartete Anzahl	104,9	51,1	156
		Prozent	64,7%	35,3%	100%
Gesamt	Anzahl	689	336	1025	
	Prozent	67,2%	32,8%	100%	

Zeichenerklärung: ** = sehr signifikante Abweichung; * = signifikante Abweichung.

Die in Tabelle 8.23 dargestellte Verteilung ist signifikant [$\chi^2(5) = 15,773$; $p = 0,008$].

Die bisher gegebenen Aussagen werden danach ergänzt.

8.100 AS: Die geschlechtsspezifische Verteilung von Gymnasiasten aus Orten unterschiedlicher Einwohnerzahl zeigt statistisch signifikanten Abweichungen.

8.101 AS: Die in dieser Stichprobe befindlichen Jungen, die das Gymnasium einer Großstadt besuchen, sind in der niedrigen Risikogruppe depBel 1 überrepräsentiert und in der hohen Risikogruppe depBel 2 unterrepräsentiert.

8.102 AS: Die Überrepräsentation der Jungen aus großstädtischen Gymnasien in der Gruppe depBel 1 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten +10,77%.

8.103 AS: Die Unterrepräsentation der Jungen aus großstädtischen Gymnasien in der Gruppe depBel 2 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten -22,12%.

8.104 AS: Die in dieser Stichprobe befindlichen Mädchen, die das Gymnasium einer Großstadt besuchen, sind in der niedrigen Risikogruppe depBel 1 unterrepräsentiert und in der hohen Risikogruppe depBel 2 überrepräsentiert.

8.105 AS: Die Überrepräsentation der Mädchen aus großstädtischen Gymnasien in der Gruppe depBel 1 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten -13,18%.

8.106 AS: Die Überrepräsentation der Mädchen aus Gymnasien in Großstädten in der Gruppe depBel 2 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten +27,06%.

8.107 AS: Die in dieser Stichprobe befindlichen Jungen, die das Gymnasium einer Mittelstadt besuchen, sind in der niedrigen Risikogruppe depBel 1 überrepräsentiert und in der hohen Risikogruppe depBel 2 unterrepräsentiert.

8.109 AS: Die Überrepräsentation der Gymnasiasten aus Mittelstädten in der Gruppe depBel 1 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten +10,54%.

8.110 AS: Die Unterrepräsentation der Gymnasiasten aus Mittelstädten in der Gruppe depBel 2 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten -21,61%.

8.111 AS: Die in dieser Stichprobe befindlichen Mädchen, die das Gymnasium einer Mittelstadt besuchen, zeigen mit ihren Frequenzen in den DTK-Risikogruppen depBel 1 und depBel 2 nur zufällige Abweichungen.

8.112 AS: Die in dieser Stichprobe befindlichen Jungen, die das Gymnasium einer Kleinstadt besuchen, zeigen mit ihren Frequenzen in den DTK-Risikogruppen depBel 1 und depBel 2 nur zufällige Abweichungen.

8.113 AS: Die in dieser Stichprobe befindlichen Mädchen, die das Gymnasium einer Kleinstadt besuchen, zeigen mit ihren Frequenzen in den DTK-Risikogruppen depBel 1 und depBel 2 nur zufällige Abweichungen.

Kapitel 9

Analyse der mit dem CDI aufgenommenen Daten

Aus den vorangegangenen Darstellungen ist bekannt, dass zur Erfassung und Einschätzung von depressiven Belastungen in der untersuchten und hier zu beschreibenden Stichprobe von 2034 griechischen Schulkindern neben dem DTK auch das CDI als ein zweites Verfahren zum Erreichen der Untersuchungsziele eingesetzt worden ist. Beide Verfahren waren zuvor nicht in griechischer Sprache verfügbar. Die Einführung zweier in der griechischen Sprache neuer depressionsspezifischer Messinstrumente in ein und derselben Stichprobe, dies in Anlehnung an den in vielfältige Weise nützlichen Multitrait-Multimethod-Ansatz, öffnet neben anderem die Möglichkeit zu einer wechselseitigen Kontrolle der Messinstrumente und der mit ihnen erarbeiteten Befunde. Ein solcher Vergleich ist aber nur möglich, wenn zuvor bekannt gegeben wird, welche Ergebnisse die Anwendung der eingesetzten Verfahren im Hinblick auf die in dieser Arbeit gestellten Fragen erbrachte. Nachdem das in den beiden voran gegangenen Kapiteln für den DTK erfolgte, sind nun die Ergebnisse aus dem Einsatz des CDI in einer griechischen Stichprobe zu betrachten. Den dazu notwendigen Analysen wird in diesem Arbeitsteil nachgegangen.

Die Sichtweise, in der die aus der Durchführung dieser empirischen Studie gesammelten Daten betrachtet werden, ist in der Einführung über die Analyse der mit dem DTK aufgenommenen Daten explizit dargestellt und erläutert worden (Kapitel 7). Die dort beschriebenen und diskutierten Annahmen gelten in den ihnen zu Grunde liegenden Überlegungen auch für die Daten aus dem CDI und aus allen anderen in dieser Arbeit eingesetzten Messverfahren. Deshalb erübrigt sich die Wiederholung der im Zusammenhang mit der Analyse des DTK dargestellten und so auch für die Analyse des CDI geltenden Arbeitsperspektive. Für die Daten aus jedem der eingesetzten Verfahren erweist sich aber als notwendig, die jeweils geltende spezielle Theorie bekannt zu machen, in der die erarbeiteten Daten eingebunden sind. Dem dient die nun folgende Darstellung über das Verfahren CDI, das auf Informationen aus unterschiedlichen anderen Studien basiert.

9.1 Das CDI als Messverfahren zur Erfassung depressiver Störungen bei Kindern und Jugendlichen

Die Entwicklung des „Children’s Depression Inventory“ (CDI, Kovacs, 1985, 2004) fällt in eine Zeit, in der die kontroversen wissenschaftlicher Diskussionen um die Existenz von Depressionen bei Kindern und Jugendlichen abgeklungen waren (siehe dazu Kapitel 3.1 und 3.2). In der Einleitung zu einer der frühen Studien über Depressionen bei Kindern und Jugendlichen, mit der auch spezifisches Wissen über die psychometrischen Eigenschaften des CDI erarbeitet worden war, beschreiben Smucker, Craighead, Craighead und Green (1986) den wissenschaftshistorischen Rahmen der Entwicklung eines neuen Verständnisses depressiver Störungen in der Kindheit, mit dem dann auch das Messinstrument CDI entwickelt worden ist: „Data supporting the existence of childhood depression (Puig-Antich & Gittelman, 1982) and indicating that masked depression is at worst ‚thinly veiled‘ when adequately assessed (Carlson & Cantwell, 1980b), have led to the development of a number of assessment instruments (Kazdin, 1981; Kazdin & Petti, 1982). The one that has been most widely used is the Childhood Depression Inventory (Kovacs & Beck, 1977), a downward extension of the adult Beck Depression Inventory (Beck, Ward, Mendelson, Mock & Erbaugh, 1961). Its development was based on three then-controversial but increasingly accepted assumptions: (1) that childhood depression does exist, (2)

that it can be observed and measured, and (3) that its characteristics are comparable to those of adult depression” (Smucker und Mit., 1986, S. 26).

Die von Smucker et al. (s.o.) vollzogene Kennzeichnung des CDI als meistgenutztes Verfahren, “... The one that has been most widely used ...“, in der Erforschung von Depressionen bei Kindern und Jugendlichen, wird von zahlreichen Veröffentlichungen gestützt (u.a. Carey, Faulstich, Gresham, Ruggiero & Enyart, 1987; Craighead und Mit., 1998; Fristad, Emery und Beck, 1997; Kazdin, 1981; Kazdin und Petti, 1982; Kendall, Cantwell und Kazdin, 1989; Saylor, Finch, Spirito und Bennett, 1984; Twenge & Nolen-Hoeksema, 2002; Maikranz, Steele und Forehand, 2003; Worchel, Hughes, Hall et al., 1990). Die Tatsache, dass das CDI eines der ersten verfügbaren Verfahren zur Erfassung depressiver Symptome bei Kindern ist, bietet eine erste und nachvollziehbare Erklärung für die beachtenswert große Zahl von CDI-Studien. Zusätzlich bedeutsam mag sein, das CDI auf dem wohlbekannten und bis heute noch, auch in der klinischen Praxis, am häufigsten verwendete Depressionsinventar für Erwachsenen, dem von A. Beck, basiert. Darüber hinaus ermöglicht die original in englischer Sprache publizierte Fassung des Verfahrens einen leichten Zugang in der Anwendung für alle daran interessierten Forscher innerhalb und außerhalb englischsprachiger Länder. Das beweist die internationale Streuung der Anwendungen des CDI auch in nicht englischsprachigen Ländern, so beispielsweise in Bulgarien, Tschechien, Dänemark, Norwegen, Polen, Russland, Ukraine, Frankreich, Italien, Spanien und in der Türkei. Im deutschsprachigen Raum gibt es sogar unterschiedliche CDI-Versionen: eine deutsche Übersetzung von Bowi und Krampe (1987), eine deutsche CDI-Version von Stiensmeier (1988) und noch eine von Nevermann (1989).

Unabhängig von den Gründen, die zu dieser Phase einer bevorzugten Anwendungen des CDI innerhalb der internationalen Depressionsforschung geführt haben, brachte die Fülle von Untersuchungen auch eine Flut von Daten und Einsichten, mehr als bei vielen anderen Messinstrumente. Mit der Generierung von Wissen über den CDI erhöhte sich auch die Zahl identifizierter Anwendungsprobleme. Auf zwei dieser Problembereiche ist hier einzugehen, weil sie implizit auch die Analysearbeit an den Daten aus der griechischen Anwendung des CDI in der Stichprobe 2034GR berühren. Problematisiert werden in der Literatur 1. seine psychometrischen Eigenschaften und 2. die Widersprüchlichkeit der unter Anwendung des CDI gewonnenen Untersuchungsergebnisse. Letzteres lässt sich mit der Gegenfrage konfrontieren, ob die Widersprüchlichkeit der mit einem Verfahren in unterschiedlichen Stichproben eingeholten unterschiedlichen Ergebnisse eine Eigenheit des Verfahrens oder eher eine der untersuchten Stichproben ist. Widersprüchliche in den Ergebnissen finden sich im Hinblick auf die Beziehungen zwischen den depressionsspezifischen individuellen CDI Messwerten einerseits und Alter, Geschlecht, Wohnort und sozioökonomischem Status der Familien von getesteten Personen andererseits.

Im Hinblick auf psychometrische Eigenheiten des CDI werden in der Literatur insbesondere Aspekte der Reliabilität und der Validität des CDI aufgegriffen. Ein Großteil der damit befassten Studien erreicht dazu relative Übereinstimmung. So sprechen die Ergebnisse solcher Studien sowohl in klinischen als auch in nicht klinischen Stichproben mit Kindern von einer guten bis sehr guten Reliabilität des CDI (Hesel und Matson, 1984; Fundudis und Mit., 1991; Nelson, Politano, Finch, Wendel und Mayhall, 1987; Saylor und Mit., 1984; Smucker und Mit., 1986). Auch die Test-Retest Reliabilität wird für gut befunden (Kaslow, Rehm & Siegel, 1984; Nolen-Hoeksema, Seligman & Girgus, 1986; Seligman, Kaslow, Alloy, Peterson, Tananbaum & Abramson, 1984; Smucker, Craighead, Craighead & Green, 1986). Weiterhin findet die Konstruktvalidität des Verfahrens Anerkennung (u.a. Asarnow und Carlson, 1985; Carey, Faulstich, Gresham, Ruggiero

& Enyart, 1987; Costello und Angold, 1988; Hershberg, Carlson, Cantwell & Strober, 1982; Saylor et al., 1984; Hodges, 1990; Kazdin, French, Unis & Esveldt-Dawson, 1983; Kovacs, 1985; Lobovits & Handal, 1985; Worchel et al., 1990).

Da die auf diese Weise in der Literatur als positiv bewerteten Eigenheiten vor allem messtechnische Qualitäten darstellen, ist die Frage nach der Validität und Nützlichkeit des Verfahrens naheliegend. Aus Studien in klinischen Populationen von Kindern hat sich ergeben, dass das CDI zwischen Kindern mit einer Diagnose der Major Depression und Kindern mit anderen psychiatrischen Störungen gut zu diskriminieren vermag (u.a. Carey und Mit., 1987; Carlson & Cantwell, 1980a; Craighead und Mit., 1995; Domènech und Mit., 1992; Haley, Fine, Marriage, Moretti & Freeman, 1985; Hogdes, 1990; Knight, Hensley und Waters, 1988; Kovacs, 1985; Lobovits & Handal, 1985; Moretti, Fine, Haley & Marriage, 1985; Rotundo und Hensley, 1985; Saylor, Finch, Spirito und Bennet, 1984; Smucker und Mit, 1986). Gut bedeutete indessen aber keineswegs fehlerfrei. Beachtenswerte Diskriminationen zwischen klinischen und nicht klinischen Populationen gelten auch als nachgewiesen (Carey, Faulstich, Gresham, Ruggiero & Enyart, 1987; Saylor et al., 1984;).

Die vorstehende Auflistung von Publikationen betrifft ausnahmslos Beiträge, die positive Einschätzungen der Messwerte des CDI geben. Die Zahl solcher Beiträge ist höher als die Liste der aufgeführten Untersuchungsberichte. Problematisiert werden die positiven Einschätzungen, weil zugleich auch Studien, vor allem mit klinischen Stichproben zu finden sind, die Hinweise auf Fehleinschätzungen zur diskriminativen Leistungsfähigkeit des CDI geben. So habe sich gezeigt, dass CDI-Mittelwerte von depressiven und nicht depressiven Kindern keinen signifikante Unterschiede aufgewiesen hätten (Kazdin, 1987; Nelson, Politano, Finch, Wendel & Mayall, 1987; Wendel, Nelson, Politano, Mayhall und Finch, 1988).

Nach Carey und seinen Mitarbeitern (1987) könnten die in ihren Untersuchungen mit dem CDI erreichten schwachen Differenzierungen zwischen depressiv oder verhaltensgestörten Kindern, aus unterschiedlichen Gründen aufgetreten sein: „....This lack of differentiation could have occurred because the overall level of dysfunction with inpatient psychiatric populations is elevated across multiple indices of emotional distress (e.g. depression, anxiety, behavior problems) or because of the comorbidity of affective and conduct-disorders diagnoses... Because the CDI was originally intended to be a quick-screening devise for depressive symptoms, another explanation may be that the CDI is assessing a global versus specific distress. Also, because diagnostic reliability was not formally assessed, the possibility exists that unreliable diagnoses accounted for some of the inability of the CDI to discriminate clinical groups...“ (Carey und Mit., 1987, S. 761).

Eine weitere Annahme betrifft möglicherweise bestehende Überlappungen zwischen den vom CDI erfassten Störungsbildern. Indem das D in der Bezeichnung des Tests als CDI für Depression steht, ist das Messobjekt eindeutig bezeichnet. Eine Bezeichnung sichert aber nicht gegenüber dem denkbaren Sachverhalt einer ungewollten Erfassung weiterer, nicht gewollter, aber doch überlappender Störungsbilder. In diesem Sinne äußern sich Saylor et. al., 1984, indem sie ausführen:.... Our data suggested that the CDI measures a multidimensional construct that overlaps with other childhood disorders, particularly anxiety (siehe auch Kresanov und Mit., 1998; Almqvist und Mit., 1999). Saylor und seine Mitarbeiter betrachten das CDI als ein Verfahren zur Erfassung “of general distress”. Ähnlich ist die Einschätzung, mit dem CDI würden Informationen über das Auftreten von „...conduct disorders...“ aufgenommen werden (Nelson und Mit., 1987).

In diese Problematisierung der Validität des CDI geht auch die Diskussion über die Komorbidität depressiver Störungen ein. Wenn depressive Störungen von Kindern und Jugendlichen in stark ausgeprägter Weise mit dem gleichzeitigen Auftreten von Angststörungen verbunden sein sollten, würde auch mit dem gleichzeitigen oder interagierenden Auftreten der Symptome parallel bestehender Störungen gerechnet werden müssen. Damit aber hätte jedes Instrument zur Identifizierung von Störungen so auf Interferenzen zu reagieren, dass differentielle Diagnosen dennoch möglich werden, was sich als schwierig, wenn nicht unmöglich, erweisen dürfte.. Die ausgeprägte Komorbidität psychischer Störungen bei Kindern wird nun aber allgemein als gegeben angenommen. Durch die vielen Nachweise über Komorbidität angeregt, wird auch diskutiert, ob Depression und Angststörungen bei Kindern zwei unterschiedliche psychische Zustände seien, oder ob sie beide doch zu einem umfassenderen Störungsbild gehörten, für das in der einschlägigen Literatur der Begriff „negative affectivity“ Verwendung findet (Hodges, 1990). Nach Doerfler und Mit. (1988) ist die schon so häufig und dann auch in ihrer eigenen Studie auftretende Gleichzeitigkeit des Auftretens von Depressionen und Angst bei Kindern ein Grund, um Fragen nach den psychischen Qualitäten von Emotionen in Schulkindern zu stellen. Nach ihrer Meinung sollte dies bedacht werden: „Poor discriminant validity of measures of anxiety and depression may reflect the inadequacies of available assessment tools or the genuine coexistence of these emotions in human experience“ (S. 771-772). Auch Kovacs und Devlin äussern sich in diesem Sinne: „Yet it should be noted that depressive and anxiety symptoms are highly comorbid in children and adolescents, whereas pure depression is relatively rare“ (Kovacs und Devlin, 1998)

Die Hinweise auf die aus der Komorbidität von Depression und Angst hervorgehenden messtheoretischen Probleme verweisen auf die bestehende allgemeine Problematik. Die Vermutungen darüber, wie die so häufig beobachtete Komorbidität erklärt werden könnte, öffnen Denkperspektiven. Problemlösungen bieten sich nicht an. Das Problem besteht weiterhin und so eben auch für die Anwendung des CDI in der Untersuchung 2034GR.

Da der Umgang mit dem CDI als Messinstrument unteilbar ist, berührt jedes Anwendungsproblem für sich auch die anderen. So könnte einerseits die Problematisierung des Cut-offs im CDI, weil lediglich Punktwerte zu erhöhen oder herab zu setzen wären, beiläufig als technisches Problem in seiner Wichtigkeit herabgestuft werden. Andererseits ist eine Testskala ohne verbindliche oder relativ verbindliche Wertungskriterien beliebig. Die Beliebigkeit ist inkompatibel mit jedem Anspruch auf Zuerkennung höherer Messqualitäten. Das Auffinden verlässlicher Kriterien, so auch die auf einen verlässlichen Cut-off Wert sind an verlässliche Vorstellungen über das Messobjekt und seine Auftretensbedingungen gebunden. So müsste entschieden werden, ob das Messobjekt des CDI die depressiven Störungen von Kindern, deren Angststörungen, „general distress“ (s.o.), „negative affectivity“ oder eben doch eher „...conduct disorders...“ (s.o.) sind. Ein Cut-off Wert ohne Hinweis darauf, welche Kinder einer Stichprobe unter seiner Anwendung aus welchen Gründen von anderen getrennt werden sollten, wäre ein messtheoretischer Widersinn.

Die theoretischen Hintergründe der Diskussionen um einen für den CDI gültigen Cut-off Wert sind mitunter nicht eindeutig zu bestimmen. So wird ein individueller CDI-Gesamtwert von 19 oder mehr Punkten als Cutoff-Wert vorgeschlagen, weil eine solche Punktzahl mit dem oberen 10% der Verteilung von CDI-Punkten aus vielen Studien mit nicht klinischen Stichproben zusammenfällt (u.a. Kovacs, 1980/1981; Smucker, Craighead, Craighead & Green, 1986). Die Bestimmung eines solchen Cut-off Wertes wäre aber ohne Angaben zu den damit verbundenen Identifizierungsleistungen überflüssig und unbefriedigend. Als überflüssig ist dieser 10%-Wert anzusehen, weil ein solcher Wert ohnehin in jeder Verteilung zu finden ist, er braucht nicht

bestimmt zu werden. Bestimmte man ihn aber doch als Trennwert, so bliebe unbefriedigend, dass nicht hinreichend verbindlich erklärt wird, welche kleinere Teilgruppe der Stichprobe aus welchen Gründen von der größeren Teilgruppe getrennt werden soll. Eine solche Begründung bietet dann aber die Information, dass bei diesem Grenzwert in der Studie von Lobovits und Handal (1985) etwa 88% der beteiligten Kinder richtig identifiziert werden konnten. Nun verweisen aber andere Studien auf andere Punktwerte mit anderen Erkennensleistungen. Die jeweilige Sensitivität oder Spezifität von Cut-off Werten ist keine Eigenheit der Trennwerte, sondern viel mehr eine Eigenheit der jeweils untersuchten Stichprobe. Innerhalb einer Studie von Asarnow und Carlson (1985) wurde der CDI in einer kinderpsychiatrischen Abteilung bei 82 Kindern, davon 22 Mädchen und 60 Jungen, insgesamt mit einem durchschnittlichen Alter von 10.76 Jahren, angewendet. Die Autoren berichten aus dieser Studie für einen CDI-Cutoff von 15 Punkten die folgenden kritische Werte: Sensitivität 54%, Spezifität 91% und eine insgesamt richtige Klassifikation von 71%. Die als Kriterium für die korrekte Identifizierung dienende Diagnose beruhte für die Mehrheit der Kinder auf einem modifizierten klinischen Interview, dem K-SADS („Schedule for Affective Disorders and Schizophrenia for School Age Children“, Puig-Antich, Orvaschel, Tabrizi und Chambers, 1983). Die Übereinstimmung zwischen den zwei unabhängigen Diagnosen ist für diejenige Kinder, die nicht mit dem K-SADS untersucht worden waren ($n = 16$) von den Autoren als „relatively high“ eingestuft worden ($\kappa = .81$). Die Diagnosen für die untersuchten Kinder lauteten unter Angabe der Zahl der jeweils Betroffenen wie folgt: 16 Kinder „Major Depression“, 18 Kinder „dysthymic disorder“, 5 Kinder „adjustment disorder with depressed mood“, 5 Kinder „anxiety disorder“, 8 Kinder „attention deficit disorder“, 21 Kinder „conduct disorder“, 9 Kinder „other mental disorder“. Die Reliabilität dieser Diagnosen wurde nur für die Teilstichprobe der depressiv diagnostizierten Kinder (33% der Stichprobe) berechnet und erreichte einen Wert von $\kappa = 0,82$. Die Teilstichprobe der als depressiv diagnostizierten Kinder hatte darüber hinaus die folgenden komorbiden Merkmale: 12 dieser Kinder wiesen eine zusätzliche Verhaltensstörung auf, 6 Kinder litten unter Angststörung, 2 Kinder wiesen Verhaltens- und Angststörungen auf und nur für 16 wurde eine „reine“ depressive Störung festgestellt (Asarnow und Carlson, 1985). Die mitgeteilten Zahlen veranschaulichen das gegebene Validitätsproblem.

Eine andere Studie, die zum Wissen über die konvergenten und diskriminanten Validität des CDI beiträgt, ist die von Hodges (1990). Die untersuchte Stichprobe umfasste 70 psychiatrische Kinderpatienten des Duke University Medical Center. Zu ihnen gehörten 45 Jungen und 25 Mädchen zwischen 6 und 13 Jahren mit einem durchschnittlichen Alter von 10 Jahren. Die Diagnosen für diese Kinder basierte auf dem klinische Interview CAS („Child Assessment Schedule“, Hodges et al., 1982). Diagnostisch identifiziert wurden 13 Kindern mit einer depressiven Störung (Major Depression oder Dysthymia), 32 Kindern mit einer Angststörung und 25 Kindern mit einer Verhaltensstörung, wobei auch diese Stichprobe aus Kindern mit komorbiden Störungsbildern besteht. Die Reliabilitätswerte für das CAS waren $k = 0,71$ für die Diagnose „conduct disorder“, $k = 0,83$ für die depressive Störungen und schließlich $k = 0,72$ für die Diagnose „Angststörung“. Die Annahme einer diskriminanten Validität des CDI wird nach Hodges durch das Ergebnis dieser Studie unterstützt. Die als depressiv diagnostizierten Kinder erreichten signifikant höhere Werte im CDI als die verhaltens- oder angstgestörten Kinder. Für einen Wert von 15 CDI-Punkten war die Spezifität des Verfahrens 84% und die Sensitivität 54%. So kommt Hodges zur Feststellung, dass das CDI eher für die Identifizierung „gesunder“ als für die gestörter Kinder geeignet sei. Nach seiner Meinung und unter Berufung auf Costello und Angold (1988) ist dies nicht als Nachteil zu bewerten. Im Rahmen eines „screening“ innerhalb epidemiologischer Untersuchungen relativiere der ökonomische Vorteil solcher Vorgehensweisen das Problem der Missklassifikationen. In der Gesamtbewertung der Befunde seiner Studie betont

Hodges seine Annahmen: „These conclusions are based on the assumption that the method used to generate diagnoses in the present study was sufficiently reliable and valid, which may be not the case.“ (Hodges, 1990, S. 380).

Eine weitere Studie, die unter anderem auch Angaben über die diskriminante Validität des CDI bietet, ist die von Craighead, Curry und Ilardi (1995). Die Stichprobe umfasst 107 Jugendliche zwischen 12 und 18 Jahren. Unter ihnen waren 58 Jungen und 49 Mädchen mit einem durchschnittlicher Alter von 15.4 Jahren bei einer SD von 1.5. Alle Jugendlichen waren „psychiatric inpatients“, in der Psychiatrie stationär behandelte Patienten. Das Hauptziel dieser Studie war die Untersuchung der Beziehungen zwischen den CDI-Gesamtscores, den Gesamtpunktwerten von 5 aus einer früheren Studie mit dem CDI extrahierten Faktoren und von klinischen Diagnosen. Als Kriterium für die diagnostische Leistung des CDI wurde in dieser Studie das halbstrukturierte Interview K-SADS-III-R verwendet. Die psychometrische Eigenschaften dieses Verfahrens sind in einer Veröffentlichung von Ambrosini näher beschrieben. Nach dieser Studie sind die Messeigenschaften allerdings eher problematisch zu nennen, wenn die Anmerkungen dazu zur Kenntnis genommen werden: „...discrepancies in diagnoses between the interviewer and the observer were resolved by discussion of all available data...“ (Craighead und Mit., 1995, S. 173). In dieser Studie wurden unter Verwendung eines Cutoffscores von 17 CDI-Punkten 81% aller mit einer Major Depression diagnostizierten Jugendlichen und 84% aller anderen Beteiligten richtig klassifiziert. Bei einem Grenzwert von 19 war die Spezifität und Sensitivität des CDI entsprechend 91,4% und 69,2%. Im Hinblick auf die verwendeten CDI-Faktoren hat sich aus dieser Studie ergeben, dass alle 5 Faktoren hoch mit dem klinischen Konstrukt der „Major Depression“ korrelieren. Das traf auch auf den vieldiskutierten „externalizing factor“ (CDI-Items 5, 15, 23, 26, 27) zu. Der Faktor „Dysphoric Mood“ (CDI-Items 1, 7, 10, 11, 20) war für sich allein fast genauso treffsicher wie der CDI-Gesamtwert bei der Aufgabe, Kinder mit einer Major Depression richtig zu identifizieren.

Kresanov, Tuominen, Piha und Almqvist (1998) haben unter Einbeziehung des CDI eine epidemiologische Untersuchung, als „two-stage“ Untersuchung gekennzeichnet, mit einer Stichprobe von 5664 Kindern aus Finnland im Alter zwischen 8 und 9 Jahren durchgeführt. Ziel war unter anderem die Überprüfung der diskriminanten Validität des CDI, aber weiterhin auch die Prüfung des Parent Questionnaire (scale A2) und Teacher Questionnaire (scale B2) von Rutter. In Bezug auf die diagnostische Fähigkeit des CDI hat sich in dieser Studie eine Sensitivität des CDI von 48% bei einem Cutoff von 17 CDI-Gesamtpunktwerten ergeben. Die diagnostische Leistung des CDI war aber bei einem Cutoff-Wert von 13 CDI-Punkten mit einer Sensitivität von 67% deutlich höher als die Leistungen eines verwendet klinische Interviews. Die psychometrischen Eigenschaften des klinischen Interviews, des hochstrukturierten klinischen Interviews DISC, das als Validitätskriterium für den CDI benutzt worden war, sind in der Publikation über die finnische Studie aber nicht angegeben worden. In der Diskussion ihrer Ergebnisse teilen Kresanov und Mitarbeiter einige methodologische Aspekte ihrer Studie mit, die einen Einfluss auf die diagnostische Leistung des CDI gehabt haben könnten: „Apparently children of 8-9 years of age respond differently to written questionnaires than to oral ones, or the time (3-5 months) from the screening procedure to the interview is too extended for a child this age“ (Kresanov und Mit., 1998, S. 91).

In einer Studie von Fundudis und Mitarbeitern (1991) unter Einbeziehung des CDI bei der Untersuchung einer Stichprobe von 93 klinisch diagnostizierten depressiven Kindern ergab sich für den CDI eine Rate fehlerhafter Zuweisungen von 25% bei einem Cut-off von 15 CDI-Punkten. Solche Angaben lassen nicht erkennen, wie hoch die spezifischen Erkennensleistungen für depressive Kinder oder für andere waren. Eine richtige Zuweisung von 75% der untersuchten Kinder wird auch aus anderen Untersuchungen und auch über andere Verfahren bekannt, sagt aber

nicht hinreichend viel über die Bewältigung der Aufgabe aus, nach Möglichkeit die depressiven Kinder fehlerfrei zu identifizieren.

Die Ergebnisse aus den aufgeführten Untersuchungen über eine mehr oder minder erfolgreiche Suche nach einem idealen Grenzwert zur Identifizierung von klinisch depressiv gestörten Kindern bestätigen in ihrer Gesamtheit am ehesten die Aussagen von Koot und Verhulst (s.o.). Sie beklagten zum einen das Fehlen eines „gold standard“ und hoben zum anderen den Effekt der Stichprobemerkmale auf die Effizienz des CDI gegenüber allen anderen Erklärungen hervor.

Zumindest mit den zitierten unterschiedlichen Studien wird eindeutig eigentlich nur bestätigt, was im Voraus bekannt war, nämlich, dass das CDI nicht das zu erreichen ermöglicht, wofür es konzipiert worden ist, die nach diagnostischen Kriterien sichere Identifizierung von depressiven Kindern. Mit der Anwendung des Screening Verfahrens CDI, ist die Identifizierung von Kindern möglich, für die eine beschreibbare Wahrscheinlichkeit vorliegt, dass sie zu denen gehören, die von einer klinisch ausgeprägten depressiven Belastung betroffen sind. Das CDI hilft, die Wahrscheinlichkeit für das Auftreten gelungener Identifizierungen zu erhöhen. Ob das für die jeweils verfolgte Zielsetzung hinreichend ist, muss im Hinblick auf die Aufgaben entschieden werden, die dem CDI jeweils zugedacht werden. Einen Zweck von dimensional Verfahren, denen einige diagnostischen Qualitäten fehlen, wie eben auch dem CDI, hat Kovacs schon in einer früheren Veröffentlichung explizit betont: „...rating scales should be utilized with awareness that they are severity measures and not diagnostic tools“ (Kovacs, 1980/1981, p. 313). Unter Berücksichtigung dieser methodologischen Limitierung könnte die über den Einsatz des CDI, je nach Stichprobe, erreichbare Identifizierung von 50% bis 88% der Kinder mit einer Major Depression, auch als eine befriedigende Leistung bewertet werden. Wichtiger als die Vergabe von Prädikaten ist die empirisch exakte Bestimmung der Leistungen eines Verfahrens, weil nur damit die Auswahl des für das Erreichen der jeweils verfolgten Zielsetzung potentiell hilfreichen Verfahrens ermöglicht wird.

Als Belege für eine potentielle Nützlichkeit des CDI sind Ergebnisse aus vielen unterschiedlichen Studien heranzuziehen. Auch wenn andere Studien zu kritischen Einschätzungen der diagnostischen Potenzen des CDI kommen, da das Inventar beispielsweise Kinder mit einer Major Depression nicht zureichend häufig und damit nur fehlerhaft identifizierte, in der Perspektive zahlreicher Aufgabenstellungen, so auch bei pädopsychologischen Zielsetzungen, kann das CDI helfen, Ziele zu erreichen. Zahlreiche empirische Untersuchungen führten zum Nachweis von höheren Korrelationen zwischen CDI-Punkten und klinisch relevanten individuellen Beeinträchtigungen auf unterschiedlichen psychischen Ebenen. Zu betonen ist, dass solche Befunde aus zahlreichen Studien mit sehr unterschiedlichen Stichproben immer wieder mit sehr ähnlichen Werten und somit dann auch mit nachweisbarer Stabilität auftreten. In der Studie von Doerfler und Mitarb. (1988) hatten die untersuchten Kinder mit individuell hohen CDI-Werten (ab 19 Punkte) auch signifikant höhere Werte in parallel vorgenommenen Messungen von Angst, negativem Selbstbild und einem „external focus of control“ als Kinder mit niedrigeren CDI-Scores. Für die erste Gruppe von Kindern haben ihre Lehrkräfte darüber hinaus berichtet, dass diese Kinder signifikant höhere Frequenzen von Verhaltensproblemen aufweisen und auch deutlich häufiger in der Schule fehlen. In ähnlicher Weise hatte eine nach CDI „depressive Gruppe“ von Kindern in der Studie von Kazdin (1989) signifikant höhere Werte in weiteren Messungen von Depression und von Hilflosigkeit als die Mitschüler mit niedrigeren CDI-Werten. Durch hohe CDI-Werte belastete Kinder zeigten wiederum signifikant niedrigeres Selbstwertgefühl und berichteten im Vergleich zu anderen Kindern von geringerer Freude an Aktivitäten. Zwischen den zwei Gruppen mit höheren und niedrigeren CDI-Punkten wurden auch Unterschiede in kognitiven Prozessen beobachtet. Die Gruppe der Kinder mit höheren CDI-Werten gab höhere Frequenzen an negativen Aussagen über sich selbst als die Vergleichsgruppe

ab, äußerte weniger persönliche Wünsche und Zielsetzungen sowie niedrigere Erwartungen an positiven Konsequenzen aus der Umwelt, reagierte gegenüber negativen Ereignissen stärker mit internen Attributionen und ließ des weiteren stärkere emotionale Belastungen durch solche Ereignisse erkennen. Alle diese Befunde belegen, dass das CDI hilft, Kinder mit unterschiedlichen psychischen Eigenheiten zu unterscheiden und dies auch - mit Fehlern – im Hinblick auf depressionsspezifischen Eigenheiten.

In diesem Sinne sind auch Studien zu würdigen, die dem Verlauf depressiver Entwicklungen bei Kindern und Jugendlichen folgten und denen zufolge ein CDI Gesamtwert von 13 Punkten oder höher einen hohen prädiktiven Wert für die Entwicklung einer klinischen depressiven Störung oder für ernste psychosoziale Beeinträchtigungen aufweist (Handford und Mitarbeiter, 1986; Garrison und Mitarbeiter, 1990; Mattison und Mitarbeiter, 1990). Ein solches prädiktives Potential spricht für die Anwendung des CDI in epidemiologischen Studien. Der CDI könnte helfen, besonders gefährdete Teilgruppen in Stichproben von Kindern und Jugendlichen zu entdecken.

Canals, Domènech-Llaberia, Fernández-Ballart und Martí-Henneberg (2002) haben über 4 Jahre hinweg den Verlauf der Entwicklung depressiver Belastungen, wie sie wie sie mit dem CDI erfasst werden, bei 199 Kindern beobachtet. Die Kinder waren zu Beginn der Studie zwischen 11 und 12 Jahren alt. Im Alter von 14 bis 18 Jahren, wurden die Angehörigen dieser Stichprobe zusätzlich mit einem klinischen Interview untersucht (Kriterien des ICD-10). Die Untersuchungsteilnehmer, die im Alter von 18 Jahren als klinisch depressiv diagnostiziert wurden, hatten zu 30% auch schon im Alter von 14 Jahren die Diagnose einer Major Depression. Aber 80% der Teilnehmer, die mit 18 Jahren als durch eine Major Depression belastet diagnostiziert werden mussten, erreichten im Alter zwischen 11 und 14 Jahren einen CDI-Gesamtwert, der 13 Punkte erreichte oder überschritt. Die Autoren unterstützen aufgrund dieser Ergebnisse die These, dass „subklinische“ CDI-Werte einen signifikanten Indikator für ein langfristiges Risiko zur Entwicklung einer depressiven Störung darstellen. Der Übergang von subklinischen zu klinischen depressiven Belastungen verlaufe kontinuierlich.

Diese Studie ermutigt dazu, das Auftreten von höheren CDI-Gesamtpunktwerten in Stichproben von jüngeren Schulkindern als das Auftreten von Indikatoren für eine sich entwickelnde depressive Störung zu sehen. Der Prädiktionswert dieser Indikatoren lässt sich aus der Divergenz zwischen ihrer Prävalenz und der Prävalenz der im Alter von 18 Jahren nachweisbaren Fälle von Major Depression abschätzen. Die Frequenz für das Auftreten der CDI-Gesamtpunktwerte 13 und höher ist größer als die Prävalenz der nachfolgend im Alter von 18 Jahren nachweisbaren Fälle von Major Depression. Demnach ließen sich die im jüngeren Alter auftretenden höheren CDI-Gesamtpunktwerte (13 und höher) als persönlichkeitspezifische Marker für eine im subklinischen Bereich bestehende Entwicklung einer depressiven Belastung verstehen. Die Ergebnisse der Studie von Canals, Domènech-Llaberia, Fernández-Ballart und Martí-Henneberg (2002) können als verwertbare Hinweise auf die jeweils weitere Entwicklung einer subklinischen Belastung verstanden werden. Die Untersuchung kann nicht die latent verlaufende Entwicklung der subklinischen depressiven Belastung deutlich machen, aber sie findet die Repräsentationen der zu klinischen Ausprägungen entwickelten Störungsbilder. In der über höhere CDI-Werte (13 und höher) seinerzeit erkennbar gewordenen Risikogruppe wird durch neue diagnostische Maßnahmen eine Teilgruppe erkannt, die aus Persönlichkeiten besteht, in denen sich die ursprünglich subklinische Ausprägung zu einer voll ausgeprägten Major Depression ausgebildet hat. Die anderen Angehörigen der seinerzeit identifizierten größeren Risikogruppe mögen drei weitere Untergruppen aufweisen. Eine Untergruppe könnte von Persönlichkeiten gebildet werden, die zu einem späteren Zeitpunkt eine Major Depression entwickelt. Eine andere Teilgruppe ließe sich denken, in der Persönlichkeiten sind, die eine Dysthymie entwickeln. Schließlich wäre eine

weitere Teilgruppe von Persönlichkeiten zu erwarten, die aus nicht bekannten aber wirkungsvollen Gründen die seinerzeit bestehende subklinische depressive Belastung reduziert oder überstanden hätte. Falls diese Auffassung Erklärungswert haben sollte, was nach den Ergebnissen der Untersuchung von Canals, Domènech-Llaberia, Fernández-Ballart und Martí-Henneberg (2002) angenommen werden dürfte, sollte davon ausgegangen werden, dass in einer entsprechenden Längsschnittstudie beim Alter von 18 bis 20 Jahren die beschriebenen Untergruppen der ehemals identifizierten subklinischen Risikogruppen empirisch nachweisbar sein werden.

Unter der Voraussetzung, dass keine Untersuchungen bekannt sind oder bekannt werden, die mit ihren Ergebnissen zu einer kritischen Prüfung der von Canals et al. mitgeteilten Befunde zwingen, sollte der nach Canals et al. (s.o.) sich aufdrängenden Hypothese von der prädiktiven Potenz der CDI-Gesamtpunkte in Forschung und Praxis verstärkt gefolgt werden. Die sich mit ihr ergebende Perspektive scheint in eine stringente und auf ihre Nützlichkeit relativ schnell zu prüfende und somit auch entwickelbare Praxis zu führen. Diesem speziellen präventiven Ansatz der Verwendung von Screening Instrumenten wie dem CDI oder auch dem DTK wird an weiteren Stellen der nachfolgenden Ausführungen einzugehen sein.

Eine Studie von DuBois, Felner, Bartels und Silverman (1995) wendet sich ebenfalls der Entwicklung depressiver Störungen zu und dies in Anwendung des Konzeptes der Stabilität. „Stability of self-reported depressive symptoms in a community sample of children and adolescents“ (s.o.) würde die Stabilität und Kontinuität depressiver Symptome bestätigen. Nach den Ergebnissen dieser Untersuchung hatten 33% der Kinder, die bei der ersten Messung mit dem CDI, wegen ihres relativ hohen CDI-Punktwertes, als wahrscheinlich „depressiv“ klassifiziert worden waren, zwei Jahre später erneut einen Gesamtscore von 19 CDI-Punkten oder höher. Die von diesen Kindern weiterhin selbstberichteten hohen Werte von Angst und von niedrigem Selbstwert blieben für diese Gruppe von Messung zu Messung stabil. Im Gegensatz dazu haben sowohl die Werte der anfangs, wegen ihrer relativ niedrigen CDI-Punktwerte, als „nicht depressiv“ eingestuften Kinder und die Werte derjenigen, die zwar bei der ersten Messung der „depressiven“ Gruppe zugerechnet worden waren, dann aber nach den Ergebnissen der zweiten Messung und der Gruppe der „nicht depressiven“ Kinder zugewiesen werden mussten, in begleitenden Messungen von Angst und Selbstwert jeweils Werte unterhalb klinischer Ausprägungen gezeigt. Damit konstituierten die Messungen zwei spezielle Teilgruppen, eine mit stabil unterhalb kritischer Werte bleibende Gruppe „nicht depressiv“ belasteter Kinder und eine zweite Teilgruppe von ehemals „depressiv belasteten“ Kindern, die offensichtlich eine Phase der signifikanten Besserung ihrer Befindlichkeit erfahren hatten. Die Autoren diskutieren die sich aus ihrer Studie ergebenden Hinweise im Hinblick auf deren Bedeutung für die Praxis: „In the absence of intervention, it appears that both clinical and preclinical levels of depressive symptoms may often develop into chronic and more global sources of distress.... Our results also have implications for concerns related to false positives and false negatives in the assessment of depression among youth. For youths identified as depressed using either the clinical or more liberal cutoff score, these were consistent significant associations with other forms of psychological distress. Further, when employing the slightly lower cutoff score, a notably greater percentage of youths were classified as stable depressed. These patterns have implications for false decisions in at least two respects. First, even when youths exhibiting elevated levels of depressive symptoms possibly do not satisfy criteria for a “true” depressive disorder, it still may be accurate to view them as experiencing levels of overall psychopathology that are worthy of concern and possible intervention. Second, it also appears that rigid adherence to clinical cutoff scores to identify stably depressed youths may result in a significant number of false negatives. Such youths may have slipped slightly below the criterion but nonetheless still exhibit levels of

depressive symptoms that are associated with a problematic developmental course and predictive of movement back above the cutoff threshold. “

Sowohl die Ergebnisse dieser Studie als auch die Ausführungen der Autoren sind kompatibel mit dem hier weiter oben beschriebenen präventiven Ansatz der Anwendung von Screening Verfahren. In der Diskussion von Validitätsproblemen psychologischer Messinstrumente dürfte eine Position der expliziten Relativierung sogenannter wahrer Werte nützlich sein. Als wahr werden in der Regel die Werte bezeichnet, die zu einem bestimmten Zeitpunkt, eben dem Zeitpunkt der Messung, dem Stratus der Persönlichkeit entsprechend hätten gemessen werden müssen. Der sogenannte wahre Wert kann dann der zum Messzeitpunkt leider verfehlt Wert sein. Nicht nur in der klinischen Psychologie, aber dort eben auch, ist von wissenschaftlichem Nutzen, die fortlaufende dynamische Entwicklung von psychischen Störungen als Basisannahme zu setzen. Selbst die stabil bestehende psychische Störung könnte einer, ihre Ausprägung verändernde, Entwicklung unterworfen sein. Sie ist in einem Organismus eingepasst, der sich seine gesamte Stabilität innerhalb seiner Lebenszeit durch fortgesetzte Produktion der ihn erhaltenden Ressourcen sichert. Indem die eine bestehende psychische Störung potentiell destabilisierenden organismischen Auswirkungen eine sie selbst stärkende Entwicklung durchlaufen, werden sie die Destabilisierung der depressiven Störung erreichen und somit ihre Überwindung möglich werden lassen.

Der Erörterung von Messproblemen unterliegen immer Theorien, auch wenn sie nicht explizit genannt werden. Indem etwa für den CDI beklagt werden sollte, dass seine prädiktiven Leistungen bezogen auf die Existenz depressiver Störungen über einen Zeitraum von zwei Jahren hinweg nicht hinreichend sein sollten, wäre eine solche Argumentation an die Überzeugung gebunden, die Entwicklung psychischer Störungen sei zu einem gegebenen Messzeitpunkt fixiert. Auch ob das in manchen Fällen durchaus so sein könnte, wird hier nicht wirklich gefragt. Veranschaulicht werden sollte, dass die jeweils zu Grunde liegenden impliziten oder expliziten Überzeugungen auch die Bewertung sogenannter harter Daten beeinflussen.

Des Weiteren wird für nützlich gehalten, in die Erörterungen über die Messleistungen von psychologischen Instrumenten explizit die in der Zeit potentiell gegebenen Veränderungsmöglichkeiten psychischer Störungen einzubeziehen.

Die Nützlichkeit des CDI als Screening-Instrument wird in der Literatur unter anderem auch von Almqvist et al. (1999), Carey und Mitarbeitern (1987), Costello und Angold (1988), Hodges (1990) und Nelson, Politano und Mitarbeitern (1987) vertreten. Aber auch eine veränderte Strategie der Anwendung des CDI und anderer Messinstrumente wird vorgeschlagen. Die Kombination von Testverfahren zu einer „multi-instrument battery“ wird von Kazdin, Colbus & Rodgers, 1986 empfohlen. In ähnlicher Weise sprechen sich auch andere Autoren dafür aus, so Almqvist et al. (1999), Carey et al. (1987), Doerfler et al. (1988), Fundudis et al. (1991), Kresanov und Mit. (1998), Mattisson et al. (1990) und Wendel et al. (1988).

Allgemein zu begrüßen ist im fachlichen Interesse die inhaltlich ausdifferenzierte Erörterung der Messqualitäten des CDI und deren internationale Breite.

Aus deren Rezeption werden für die hier zu führende Erörterung der Untersuchung 2034GR mehrere Feststellungen gezogen, die als wissenschaftlich tragfähig und in diesem Sinne auch nutzbringend angesehen werden:

Der CDI ist ein gegenüber aktuellen psychischen Befindlichkeiten sensitives Messinstrument.

Die mit dem CDI aufgenommenen Informationen können helfen, die Notwendigkeit zu pädagogischen und klinischen Interventionen zu erkennen.

Der CDI ist ein potentielles Hilfsmittel bei der Durchführung von Interventionen.

Hohe CDI-Werte geben spezifische Hinweise auf einen individuellen psychischen Zustand, der hoch wahrscheinlich mit depressiven Symptomen verbunden ist und in jedem Fall nachgehende und problemzentrierte Hilfeleistungen verlangt.

In dieser Perspektive ist der CDI mit den für die vorliegende griechische CDI-Version gefundenen Messqualitäten (siehe Kapitel 6.3), ein geeignetes Messinstrument zur Suche nach Hinweisen auf das wahrscheinliche Bestehen von psychischen und speziell von depressiven Belastungen in der untersuchten Stichprobe griechischer Schulkinder.

Unter speziellem Verweis auf die Untersuchungen von Canals, Domènech-Llaberia, Fernández-Ballart und Martí-Henneberg (2002) und die von weiteren Autoren werden auch die über die griechische Anwendung des CDI in der Untersuchung 2034GR gewonnenen Informationen auf die mit ihnen potentiell gegebenen Hinweise auf die Notwendigkeit der Realisierung präventiver oder kurativer Interventionen geprüft werden.

9.2 Die Datenbasis der nachfolgenden Analysearbeit zum CDI

Die derzeit genutzte Version des CDI von Kovacs (1985) umfasst 27 Items, die wesentliche Symptome einer depressiven Störung nach DSM-III, und damit auch, agitiertes Verhalten ausgenommen, nach DSM-IV, thematisieren. Gefragt wird neben anderem nach Beeinträchtigungen der emotionalen und somatischen Befindlichkeit und der Selbstwerteinschätzung. Für jedes angesprochene Verbalisierung eines Symptoms gibt es im CDI drei Aussagen, in denen das jeweils genannte Phänomen entsprechend seiner vom Befragten selbst wahrgenommenen Häufigkeit durch Selbsteinstufung einer Skala von drei Antwortalternativen, „selten“ – „manchmal“ – „sehr oft“, zugeordnet werden soll. Das jeweilige untersuchte Kind ist gebeten, unter den allgemein vorgegebenen Antworten die auszuwählen, die der eigenen Befindlichkeit in den letzten zwei Wochen besser entspricht andere. Je nach dem individuellen Antwortverhalten auf die CDI-Fragen ergibt sich über die Addition der für jede Frage zu vergebenden Punktwerte von 0 bis 2 ein Gesamtpunktwert, der von 0 bis zu 54 Punkten betragen kann. Der Gesamtpunktwert 0 wird erreicht, wenn alle 27 abgefragten Items nach Angaben des Befragten nur selten zu beobachten sind. Der Höchstwert 54 ergibt sich als Summe, wenn alle 27 vom CDI präsentierten Items von einem Kind als sehr oft erlebt berichtet werden.

Die Berechnung des Gesamtpunktwertes ist eine, und wohl auch die meistgenutzte Möglichkeit zur Differenzierung des Antwortverhaltens der befragten Kinder und Jugendlichen. Andere Analysen des Antwortverhaltens sind denkbar. So führen die mehrstufigen Möglichkeiten der Antworten, die den Kindern mit dem CDI angeboten werden, auch zu einer entsprechenden Variabilität der eigenen Summenbildung. Ein Gesamtpunktwert 12 könnte sich etwa aus 12×1 oder auch aus 6×2 zusammensetzen, oder auch aus weiteren Kombinationen von Antworten mit den Werten 0, 1 oder 2. Anzunehmen wäre, dass die Kombinatorik der Gesamtpunktwerte, ihre jeweilige Zusammensetzung aus unterschiedlichen Mengen an Antworten mit unterschiedlichen Punktwerten, spezifische diagnostisch verwertbare Informationen bietet.

Auf die Existenz dieser und weiterer, jeweils keinesfalls unbedeutender Aspekte der Analyse des Antwortverhaltens im CDI wird an dieser Stelle hingewiesen, ohne ihnen explizit zu folgen. Eine solche Frage betrifft auch den Entscheidungsprozess, dem sich die Kinder beim Ausfüllen des CDI aussetzen. Der mit jedem Item verlangte Entscheid über die relative Häufigkeit des Auftretens eines psychischen Zustandes oder seiner Verhaltenskorrelate könnte in Zusammenhang mit subjektiv wahrgenommenen und interindividuell variierenden Problemen stehen. Problematisch, nicht nur für Kinder, dürfte beispielsweise die Unterscheidung zwischen „manchmal“ und „häufig“ sein. Zumindest ist die Vermutung darüber naheliegend.

Interessanterweise waren in der gesamten Untersuchung 2034GR keinerlei Hinweise auf solche Schwierigkeiten bei den individuellen Entscheidungen wahrzunehmen. Für die 2034 an dieser Untersuchung beteiligten Kinder, die ja immer nur in gut beobachtbaren Gruppen von deutlich weniger als 30 Kindern zur Beantwortung der Fragebögen zusammentrafen, schien die Selbsteinstufung relativ unproblematisch zu sein. Die als teilnehmende Beobachterin tätige Untersuchungsleiterin fühlte sich zu der Feststellung verleitet, die Beantwortung des CDI sei für die Kinder so gut wie immer problemlos gewesen. Eine sehr kleine Minderheit unter den Kindern der Stichprobe, die während des Ablaufs der jeweiligen Untersuchung persönliche Schwierigkeiten zur Sprache gebracht haben, bestätigen, wegen des für sie extrem niedrigen prozentualen Anteils und ihrem daraus hervorgehenden Status als hervorzuhebende Ausnahme einerseits die Regelmäßigkeit der Feststellung eines unproblematischen Antwortverhaltens und andererseits immerhin die randständige Existenz des Entscheidungsproblems. Die empirisch belegte Antwort auf die Frage, warum eigentlich nur eine so extrem kleine Zahl von Kindern die vermutete Schwierigkeit bei der Wahl zwischen den Antwortmöglichkeiten erlebt, könnte einen Beitrag zum psychologischen Verständnis des Antwortverhaltens geben. An Prüfhypothesen dazu bestünde kein Mangel. Aus der zitierten Beobachtung eines problemlosen Ausfüllens des CDI-Fragebogens durch die untersuchten Kinder lassen sich andererseits und aus sehr nahe liegenden methodischen Gründen keine weiterreichenden Schlüsse für eine Erklärung des Antwortverhaltens ableiten. Unbekannt bleibt weiterhin auch, ob die Qualität der Antworten als über allgemeine Zeitmaße definierte Angaben von ebenso unbestimmten aber in Relation zueinander stehenden Mengen, so beispielsweise die „manchmal“-Antwort, aussagekräftige interpersonale Übereinstimmung erreichen. Falls dies nicht so wäre, bedeutete das aber auch nicht zwangsläufig eine Einbuße an Messqualitäten. Da die Relationen der Begriffe zueinander sehr wahrscheinlich von allen Kindern in übereinstimmender Weise verstanden werden und sich so eine interpersonal übereinstimmende Rangreihe mit unterschiedlichen Abständen zwischen den Rangplätzen ergäbe.

Die primäre Datenbasis für die Analyse der aus dem Einsatz des CDI in der Untersuchung 2034GR aufgenommenen Informationen besteht aus der Matrix der für die Untersuchungsteilnehmer festgestellten und kodierten Antworten auf die Fragen des CDI.

Als ein erster Arbeitsschritt zur Analyse der vorliegenden Daten wird auch für den CDI, wie schon zuvor bei den DTK-Daten und mit der gleichen Begründung, die Untersuchung des Antwortverhaltens der Kinder aus der Perspektive der individuell erbrachten Gesamtpunktwerte begonnen.

An dieser Stelle sollte zunächst auf einen inhaltlichen Unterschied hingewiesen werden, der zwischen dem originalen CDI von Kovacs und der hier vorliegenden griechischen Übersetzung besteht.

Dieser Unterschied betrifft das Item 9 des originalen Verfahrens, das aus Aussagen zu Todesgedanken besteht (CDI-Item 9: *„I do not think about killing my self“*, *„I think about killing myself but I would not do it“* und *„I want to kill myself“*). Die Anwendung dieser Frage in Untersuchungen mit Kindern wurde vielfach aus nahe liegenden und nachvollziehbaren Gründen als problematisch betrachtet und demzufolge unterlassen. So arbeiteten in entsprechenden Studien neben anderen auch Hoffmann et al. (2000), Kresanov et al. (1998), Leon et al., (1980), Nevermann, (1992), Reicher und Rossmann (1991), Smucker et al. (1986), Stiensmeier (1988), Stiensmeier et al. (1989), Weiss und Weisz (1988) und Weisz et al. (1991). In der hier zu beschreibenden Untersuchung 2034GR wurde das Suiziditem wegen ansonsten denkbarer Belastungen zumindest einzelner Kinder aus der Itemliste entfernt.

Die Elimination des Suiziditems 9 aus dem CDI hat nach Almqvist und Mitarbeiter (1999) keinen signifikanten Einfluss auf die sich in Abwesenheit des Items ergebenden Gesamtergebnisse. In früheren CDI-Studien hat sich ergeben, dass der Prozentsatz von Schulkindern, die von Suizidgedanken berichten, gering ist, was auf in dieser Hinsicht aus eher unauffällige Stichproben schließen lässt. Einer solchen Interpretation kann aber auch begründet widersprochen werden. Aus der Elimination einer der Fragen des CDI entsteht als Folge aber ein anderes Problem. Der höchste erreichte Gesamtpunktwert eines CDI mit 27-1 Item ist nicht mehr 54 sondern 52. Weiterhin könnten auch gestaltpsychologisch zu begründende Veränderungen vermutet werden. Das aus der verminderten Gesamtzahl an Items entstehende statistische oder messtheoretische Problem wurde von unterschiedlichen Autoren in unterschiedlicher Weise behandelt. In der einschlägigen Literatur sind Studien zu finden, in denen die Ergebnisse aus der um ein Item verkürzten Skala ohne weitere Korrekturen berechnet und verwendet werden. Ein Argument für diese Berechnungsweise könnte sein, dass der theoretisch höchste zu erreichende CDI-Gesamtpunktwert von 54 Punkten bisher in keiner Stichprobe erbracht worden ist. Andere Autoren haben die fehlende Frage des CDI durch eine neu formulierte weitere Frage ersetzt. Stiensmeier (1988) beispielsweise hat in seiner deutschen CDI-Version eine Frage, die sich mit Problemlösungsfähigkeiten befasst, beigefügt. Auch Nevermann (1989, 1992) hat in einer frühen Version das Item 9 mit einer neuen Frage ersetzt, die den jeweils eigenen Schulbesuch thematisierte: „*Ich gehe gerne in die Schule*“, „*Ich fehle in der Schule ohne ausreichenden Grund*“, „*Ich will nicht in die Schule gehen*“. Diese Fragen wurden für die Untersuchung HE2934 übernommen und ins Griechische übersetzt.

Die Einbeziehung eines Items, das nach dem Schulbesuch fragt, kann sowohl theoretisch als auch empirisch unterstützt werden. Theoretisch ließe sich in Übereinstimmung mit dem depressiven Störungsbild die Abwesenheit von der Schule etwa als Resultat reduzierter Antriebskraft, als Folge emotionaler Verstimmungen oder auch als durch psychosomatische Beschwerden erzwungen denken. Das betroffene Kind klagte über Kopfschmerzen, Bauchschmerzen oder Übermüdung wegen massiver Schlafstörungen. Lobert (1990) schreibt beispielsweise dazu: „Neben den typischen so genannten ‚passiven‘ Depressionsmerkmalen tritt auch soziales Fehlverhalten, wie z.B. ‚Schuleschwänzen‘, unter den so genannten ‚aktiven‘ Symptomen einer klinisch sicheren Depressivität auf“. Entsprechendes Verhalten wird sowieso auch bei depressiven Erwachsenen beobachtet. Für sie wird geschrieben, dass sie signifikant häufiger als ihr „gesunder“ Mitarbeiter an ihrem Arbeitsplatz fehlen. In der Studie von Doerfler und Mitarbeitern (1988) haben die Lehrkräfte der beteiligten Kinder berichtet, dass die dort gefundene „depressive Gruppe“ von Kindern deutlich häufiger in der Schule fehlte als die Gruppe ihrer Mitschüler. Obwohl es also theoretische und empirische Gründe dafür gibt, die Frage nach dem Schulbesuch potentiell auch als depressionsrelevant zu verstehen, stellt die Einbeziehung dieser Frage in der vorliegenden griechischen CDI-Übersetzung eine inhaltliche Abweichung vom originalen Verfahren dar. Der Ersatz des 9. Items des CDI durch diese neue, aber immerhin begründet gestellte Frage führt dazu, dass der höchste erreichte Gesamtpunktwert zwar mit 54 unverändert scheint, jedoch qualitativ unterschiedlich ist. Aus eben diesem Grund werden in der hier zu beschreibenden Analyse der griechischen CDI-Daten die Antworten der Kinder auf dieses eingefügte Item bei der Berechnung des individuellen Gesamtpunktwertes nicht mitberechnet. Stattdessen wird eine Berechnungsvariante ausgewählt, die ebenfalls schon sehr häufig in der Literatur zu finden ist. Der CDI-Gesamtscore wird nach dem folgenden Algorithmus gebildet: Individueller Gesamtpunktwert = Summe aus 26 Antworten + Mittelwert aus dieser Summe. Im Hintergrund dieser Berechnung steht das zweifellos zutreffende Argument, dass mit der Addition des jeweiligen Mittelwertes aus 26 Antworten lediglich die gemessene Antworttendenz verstärkt wird. Dadurch bleibt der theoretisch höchste erreichte CDI-Gesamtpunktwert wie bei der

originalen CDI-Version mit 54 Punkten erhalten, was die Vergleichbarkeit zwischen CDI-Studien erleichtert. Nevermann (1992) und Casanovas (2003) nennen dieses errechnete Item ein „virtuelles Item“. Diese Berechnungsweise haben auch Twenge und Nolen-Hoeksema (2002) in einer „meta-analysis“ über die Ergebnisse aus 310 in U.S. durchgeführten CDI-Untersuchungen verwendet. Bei der Beschreibung der von ihnen verwendeten Methodik erläutern diese Autoren die gegebene Problematik und die von ihnen ausgewählte Lösung: „In some cases, it was necessary to adjust the means. Some studies used only 26 items of the CDI because school officials or review boards would not allow them to include to item about suicide. These datapoints were recalculated to conform to the usual 27-item mean (the 26-item mean divided by 26, and the result added to the 26-item mean). Controls for number of items (26 vs. 27) did not produce any changes in the results.“ (Twenge und Nolen-Hoeksema, 2002, S. 581).

Die beschriebene Berechnungsweise wird bei den nun folgenden Analysen der aus der griechischen Untersuchung 2034GR gewonnenen CDI-Daten verwendet. Wie auch bei anderen Autoren (s.o) dient diese Entscheidung vor allem der Vergleichbarkeit der Ergebnisse mit anderen Untersuchungen.

9.3 Verteilung der CDI-Gesamtpunktwerte

In der Untersuchung 2034GR wurde eine Stichprobe von 2034 griechischen Schulkindern untersucht. Alle teilnehmenden Kinder sind der Aufgabe gefolgt, die Itemlisten der einzelnen Verfahren zu beantworten. So waren auch 2034 Kinder damit beschäftigt, die mit dem CDI gestellten Fragen zu beantworten. Die Zahl der dann in die Analyse der CDI-Daten einzubeziehenden t Kinder ist dennoch kleiner als 2034. Da die selbst gesetzte Vorschrift befolgt wird, nur lückenlos ausgefüllte Fragebögen in die Analyse einzubeziehen, ergibt sich bei jedem eingesetzten Verfahren, so auch beim CDI, dass einige Kinder jeweils einzelne Items nicht beantwortet haben. Die Größe der für die Analyse der CDI-Daten verbleibenden Stichprobe ist demzufolge 2034 – Zahl nicht vollständig ausgefüllter Antwortbögen.

Die nachstehende Tabelle zeigt die Verteilung der durch die beteiligten Kinder im CDI erreichten individuellen Gesamtpunktwerte. Die Berechnung dieser Punktwerte folgte der Auswertungsvariante, die hier zuvor als die mit dem „virtuellen Item“ bezeichnet worden war.

Die Frequenzen in dieser Tabelle basieren auf den Daten der Kinder, die alle nach Eliminierung des Suizid-Items verbleibende 26 Fragen des CDI lückenlos bearbeitet haben. Ihre Zahl, $n = 2013$, entspricht 99% der Gesamtstichprobe. Die für eine Verteilung von individuellen Gesamtpunkten aus dem CDI ungewöhnlichen Wertangaben mit zwei Stellen hinter dem Komma ergeben sich aus der weiter oben beschriebenen Verfahrensweise, bei der zu der zunächst berechneten Summe x aus 26 Items das arithmetische Mittel mit $x/26$ hinzu gegeben wird.

Tabelle 9.1: Verteilung der CDI-Gesamtpunktwerte (CDI-GPW)

CDI-GPW	Frequenz	Prozent	Kumul. Prozent	CDI-GPW	Frequenz	Prozent	Kumul. Prozent
0	23	1,1	1,1	20,77	34	1,7	92,4
1,04	48	2,4	3,5	21,81	35	1,7	94,1
2,08	69	3,4	7,0	22,85	33	1,6	95,8
3,12	90	4,5	11,4	23,88	16	,8	96,6
4,15	108	5,4	16,8	24,92	21	1,0	97,6
5,19	99	4,9	21,7	25,96	10	,5	98,1
6,23	108	5,4	27,1	27,00	12	,6	98,7
7,27	147	7,3	34,4	28,04	3	,1	98,9
8,31	133	6,6	41,0	29,08	3	,1	99,0
9,35	141	7,0	48,0	30,12	7	,3	99,4
10,38	145	7,2	55,2	31,15	1	,0	99,4
11,42	122	6,1	61,3	32,19	2	,1	99,5
12,46	115	5,7	67,0	33,23	2	,1	99,6
13,50	90	4,5	71,4	35,31	2	,1	99,7
14,54	82	4,1	75,5	36,35	1	,0	99,8
15,58	79	3,9	79,4	37,38	1	,0	99,8
16,62	76	3,8	83,2	38,42	1	,0	99,9
17,65	42	2,1	85,3	39,46	2	,1	100,0
18,69	59	2,9	88,2	44,65	1	,0	100,0
19,73	50	2,5	90,7	Gesamt	2013	100,0	

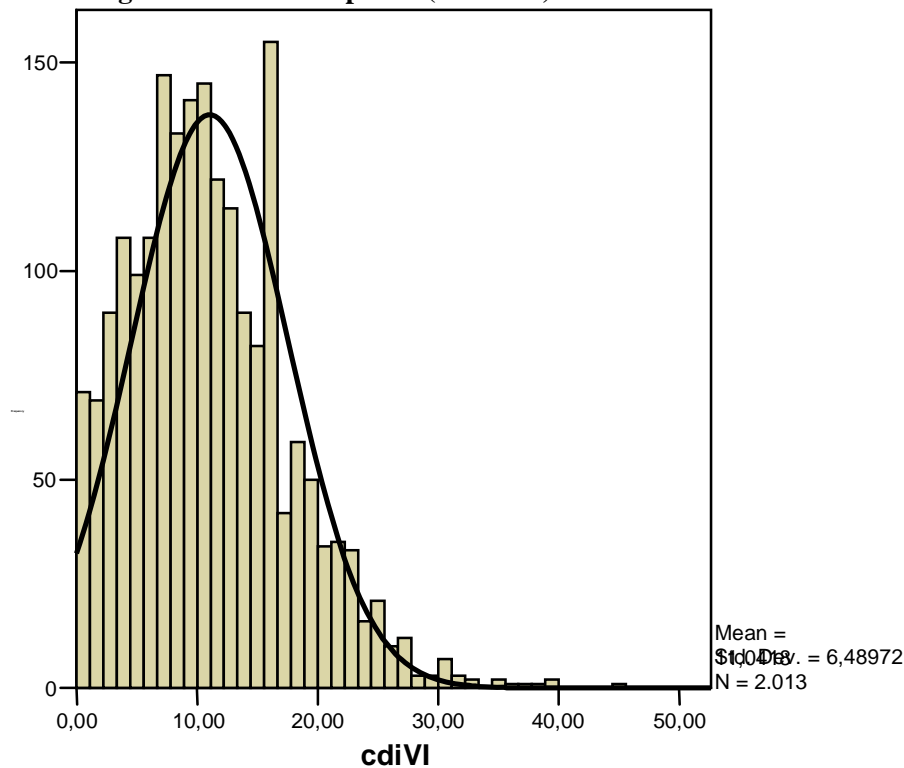
Zeichenerklärung: CDI-GPW= CDI-Gesamtpunktwerte, Kumul. Prozent = kumulativer Prozent

Zur dieser Häufigkeitsverteilung der CDI-Gesamtwerte gehören die nachfolgend aufgeführten Kennwerte.

Mean	11,0418	Variance	42,116	Skewness	,807
Std. Error of Mean	,14465	Range	44,65	Std. Error of Skewness	,055
Median	10,3846	Minimum	,00	Kurtosis	,914
Mode	7,27	Maximum	44,65	Std. Error of Kurtosis	,109
Std. Deviation	6,48972	Sum	22227,23		

Die durch die nachstehende Abbildung (Abb. 9.1) veranschaulichte Verteilung der CDI-Gesamtpunktwerte lässt eine durch die Modalitäten der individuellen Produktion von CDI-Werten entstandene und auch zu erwartende Abweichung von der Normalverteilung vermuten. Die Durchführung des Kolmogorov-Smirnov-Tests bestätigt die Hypothese einer signifikanter Abweichung von der Normalverteilung (Kolmogorov-Smirnov Z = 4,139, p = 0,000) Aus den Positionen des arithmetischen Mittels (AM= 11,04), des Medianwertes (Md = 10,38) und des Modalwertes (Mo = 7,27) darf diese Verteilung von CDI-Punktwerten als linksteil bezeichnet werden, worauf auch die nach Pearson berechnete positive Schiefe (Skewness = 0,807) verweist (Bortz, 1999, 5. Aufl., S. 46). Die gefundene linkssteile CDI-Verteilung ist durch den relativ großen Exzeß (Kurtosis = 0,914) zusätzlich als breitgipflig zu charakterisieren.

Abb. 9.1: Verteilung der CDI-Gesamtpunktwerte in der griechischen Stichprobe (n = 2013)



Die gefundene Verteilungsform der CDI-Gesamtpunktwerte aus der untersuchten griechischen Stichprobe führt zur Formulierung der folgenden Informationen als erste Analyseergebnisse.

9.1 AS: Die Verteilung der CDI-Gesamtpunktwerte in der Untersuchung 2034GR erweist sich als linksasymmetrisch. Dadurch stimmt sie sowohl mit den theoretischen Annahmen des originalen Messverfahrens als auch mit allen aus anderen CDI-Studien bisher erbrachten korrespondierenden Ergebnissen überein.

9.2 AS: Die gefundene linksasymmetrische Verteilung der von den beteiligten Kindern erreichten CDI-Werte weist darauf hin, dass die Mehrheit der Kinder vom eher geringen Aufkommen der sie betreffenden Symptome aus der Liste der CDI-Items berichten. Das darf nach dem gegebenen Erkenntnisstand als ein durch die Anwendung des CDI gewonnener Hinweis auf eine in der untersuchten Stichprobe sehr wahrscheinlich vorgefundenen relativ geringe depressiver Belastung angesehen werden. Von einer Randgruppe mit einem Anteil von 1,1% der untersuchten Kinder wird sogar angegeben, dass sie zum Zeitpunkt der Untersuchung überhaupt keine Belastung von depressiven Symptomen erlebten.

9.3 AS: Die von den griechischen Kindern erreichten individuellen CDI-Gesamtpunktwerte nehmen von einer theoretisch verfügbaren Variationsbreite von 0 bis zu 54 Gesamtpunkten nur den Teilbereich von 0 bis 45, genau bis 44,65 Punkten, in Anspruch. In der erbrachten Punktbreite werden nicht alle einnehmbaren Punktwerte besetzt. So fehlen beispielsweise Frequenzwerte für einen Gesamtscore von 34,27 Punkten wie auch für die Punktwerte von 39,46 bis 44,65. Der höchste erbrachte Score (44,65) ist nur von einem Kind erreicht worden.

9.4 AS: Aus den aufgeführten Divergenzen zwischen theoretisch erreichbaren und real erreichten Gesamtpunkten im CDI ergibt sich für die griechische Stichprobe wie auch schon in fast allen anderen bekannt gewordenen Untersuchungen, dass die schwersten vom CDI erfassbaren Belastungen, laut Selbstrating, bei den untersuchten Kindern nicht auftreten. Der höchste in dieser Untersuchung beobachtete Gesamtwert von 44,65 Punkten entspricht 82,69% der im CDI maximal repräsentierbaren Hinweise auf mögliche depressive Belastungen. In einer theoretisch verfügbaren Breite von 0 bis 54 Punkten besetzen 98,7% der Teilnehmer dieser Untersuchung einen Wertebereich von 0 bis 27 Punkten und erreichen einen Höchstwert, der in der Mitte der definierten CDI-Skala liegt. In der von den griechischen Kindern tatsächlich eingenommenen Breite des Wertebereiches von 0 bis 44,65 Punkten ist der Prozentsatz der Kinder, die Gesamtpunktwerte bis zur Skalenmitte (22,85) erreicht hatten, 95,8%. Die andere Hälfte der CDI-Skala wird dementsprechend nur von 4,2% der Teilnehmer besetzt. Beim CDI gilt demnach auch das, was in den DTK- Skalen erwartet und auch beobachtet wurde. In der griechischen Stichprobe 2034GR ist das Erreichen höherer CDI-Gesamtpunktwerte ein seltenes Ereignis. Umso aussagekräftiger ist das Einstellen solcher Werte durch eine Teilgruppe der untersuchten Schulkinder.

Das in dieser Untersuchung ebenso wie in vielen anderen aufgetretene Phänomen der Nichtbesetzung höherer Gesamtpunktwerte zum Ende der CDI-Skala dürfte, da fast schon die Regel, nicht zufällig auftreten. Eine Antwort auf die Frage nach den Ursachen könnte die Feststellung sein, dass die Wahrscheinlichkeit der fortlaufenden Wahl von Antworten mit dem höchsten Punktwert mit steigender Antwortzahl sehr deutlich abnimmt. Die in ihrem rechten Kurventeil signifikant unter dem Verlauf der Normalverteilung abgeflachten Verteilungen von Gesamtpunktwerten aus Verfahren wie CDI oder DTK belegen die Annahme. Warum die Wahrscheinlichkeiten für das Erreichen hoher Gesamtpunktwerte in diesem Maße sinken, wäre damit allerdings nicht erklärt, sondern lediglich erneut bestätigt worden. Denken ließe sich neben anderem, dass die insgesamt von den CDI-Items thematisierten 27 depressogenen Phänomene von jedem Kind nur zu einem Teil verstanden werden. Wenn aber jedes Kind subjektiv meint, nur auf $27 - y$ Items antworten zu können und für y Items aus eben diesem Grund der Feststellung zustimmen, die mit 0 Punkten verrechnet wird, wäre der CDI-Maximalwert 54 nicht mehr erreichbar. Die teilnehmenden Kinder hätten die mit 27 Items vorgegebene Liste jeweils willkürlich auf $27 - y$ gekürzt und insgesamt, interindividuell betrachtet, keine übereinstimmend lange Itemliste bearbeitet, sondern immer nur gekürzte Listen.

Der Messvorgang kann, obwohl die Ergebnisse eindeutig festliegen, durchaus in unterschiedlicher Weise abgelaufen sein. In der Verlaufsvariante 1 stünde die Antwortenvariante, die mit 0 Punkten bewertet wird, für die jeweils abgegebene Auskunft, das im jeweiligen Item genannte Phänomen zwar zu kennen, aber von dem nicht betroffen zu sein. In der Variante 2 wäre die Wahl der mit 0 verrechneten Antwort vollzogen worden, weil die Bedeutung des genannten Phänomens nicht hinreichend bekannt war und deshalb die gegebene Antwort eher näher lag als andere. Denkbar wäre dann auch, dass einige Antwortalternativen, so auch diejenigen, die auf eine schwere Belastung (2 Punkte) hinweisen, aus soziokulturellen Gründen den griechischen Teilnehmern unbekannt waren oder aber auch als nicht wählbar erschienen. Was zunächst nur ein Spiel mit Hypothesen scheint, könnte aber auch durch erfassbare und prüfbare Daten belegt sein. Sollten interkulturell voneinander abweichende Wahlbedingungen vorhanden sein und die individuellen Entscheidungen wirklich determinieren, müsste das in Frequenzlisten über die Entscheidung in jeder der 27 mal 3 Entscheidungspositionen numerisch fassbar sein. Die Kinder können bei einem CDI mit 27 Items wahlweise an insgesamt 81 Positionen ihre Wahl durch Ankreuzen bekunden. Für die Untersuchung 2034GR demonstrierte eine solche Frequenzliste zunächst, welche

Verteilung der Antworten auf die gegebenen Wahlmöglichkeiten der in der Stichprobe ablaufende Entscheidungsprozess bewirkt hatte. Ob solche Frequenzliste dann innerhalb der hier geführten Arbeit auch als spezielles Analyseobjekt herangezogen werden sollte, bliebe nachfolgend zu entscheiden.

Die folgende Tabelle (9.2) stellt die auf den Grundwert aller Teilnehmer bezogenen Prozentsätze der Kinder dar, die für das jeweils betrachtete CDI-Item entweder die mit 0 bewertete Antwortmöglichkeit, oder aber die dem Punktwert 1 oder 2 zugewiesenen Antworten angekreuzt haben.

Tabelle 9.2: Frequenzen der auf die wählbaren 78 Wahlpositionenim CDI entfallenden Antworten (auf 26 Items verkürzte Itemliste)

	0 Punkte	1 Punkt	2 Punkte
1 "Sadness"	85,7	12,8	1,5
2 "Pessimism"	55,6	37,7	6,7
3 "Self-depreciation"	65,2	33,2	1,7
4 "Anhedonia"	77,2	21,3	1,5
5 "Misbehavior"	82,5	12,9	4,6
6 "Pessimistic worrying"	74,1	19,1	6,8
7 "Self-hate"	76,9	19,2	3,8
8 "Self-blame"	70,3	22,9	6,8
10 "Crying spells"	84,0	14,2	1,8
11 "Irritability"	45,1	44,7	10,2
12 "Reduced social interest"	88,9	9,9	1,3
13 "Indecisiveness"	68,9	25,5	5,7
14 "Negative body image"	50,6	41,3	8,0
15 "Schoolwork difficulty"	52,3	28,5	19,2
16 "Sleep disturbance"	67,0	29,1	3,9
17 "Fatigue"	50,6	42,7	6,7
18 "Reduced appetite"	59,0	38,0	3,0
19 "Somatic concerns"	53,5	39,4	7,2
20 "Loneliness"	72,9	22,5	4,6
21 "School dislike"	77,3	18,9	3,7
22 "Lack of friends"	77,4	20,6	2,0
23 "School performance decrement"	59,7	31,2	9,1
24 "Self-depreciation via peer comparison"	56,6	30,9	12,4
25 "Feeling unloved"	75,5	20,5	4,0
26 "Disobedience"	54,8	43,5	1,7
27 "Fighting"	81,9	16,4	1,7

Aus der Tabelle ist zu ersehen, dass die für jedes CDI-Item wählbare Antwort, von eben diesem Phänomen nicht betroffen zu sein, mit Prozentsätzen von 45,1% bis zu 88,9% durch die teilnehmenden Schulkinder genutzt wird. von einer Nichtbetroffenheit im Hinblick auf die CDI-Symptome berichtet. Die vom CDI angebotene nächst höhere Ausprägungsstufe, nach der man bisweilen betroffen sei, findet, ebenfalls von Item zu Item variierend, zwischen 9,9% bis 44,7% Zuspruch. Den höchsten Ausprägungsgrad der Betroffenheit durch eines der Items wählen 1,3% bis 19,2% der Teilnehmer.

Insoweit als den Angaben der Kinder zu vertrauen ist, verweisen die Prozentsätze in Spalte „0 Punkte“ auf Kinder, die angeben, von dem in der Zeile genannten Item aktuell nicht betroffen zu

sein. Demgegenüber ergibt die Summe der Prozentsätze aus der Spalte „1 Punkt“ + Spalte „2 Punkte“ den Prozentsatz der Schulkinder, die über eine aktuelle Betroffenheit berichten. Die nachfolgende Tabelle 9.3 zeigt diese Werte.

Tabelle 9.3a: Prozentuale Verteilung der nach Punktwerten pro Item gegebenen Antworten

Prozentualer Anteil von Antworttypen pro Item	% 0 Punkte	% 1 + 2 Punkte	Rang	Summe
1. „Sadness“	85,7	14,3	25.	100
2. „Pessimism“	55,6	44,4	7.	100
3. „Self-depreciation“	65,2	34,9	11.	100,1
4. „Anhedonia“	77,2	22,8	19.	100
5. „Misbehavior“	82,5	17,5	23.	100
6. „Pessimistic worrying“	74,1	25,9	16.	100
7. „Self-hate“	76,9	23	18.	99,9
8. „Self-blame“	70,3	29,7	14.	100
10. „Crying spells“	84	16	24.	100
11. „Irritability“	45,1	54,9	1.	100
12. „Reduced social interest“	88,9	11,2	26.	100,1
13. „Indecisiveness“	68,9	31,2	13.	100,1
14. „Negative body image“	50,6	49,3	3.	99,9
15. „Schoolwork difficulty“	52,3	47,7	4.	100
16. „Sleep disturbance“	67	33	12.	100
17. „Fatigue“	50,6	49,4	2.	100
18. „Reduced appetite“	59	41	9.	100
19. „Somatic concerns“	53,5	46,6	5.	100,1
20. „Loneliness“	72,9	27,1	15.	100
21. „School dislike“	77,3	22,6	20,5	99,9
22. „Lack of friends“	77,4	22,6	20,5	100
23. „School performance decrement“	59,7	40,3	10	100
24. „Self-depreciation via peer comparison“	56,6	43,3	8.	99,9
25. „Feeling unloved“	75,5	24,5	17.	100
26. „Disobedience“	54,8	45,2	6.	100
27. „Fighting“	81,9	18,1	22.	100

Zeichenerklärung: Rang = Rangplatz des Prozentsatzes der „1+2 Punkte“- Antworten

Die Spalte „Rang“ lässt erkennen, dass Item 11, in der englischen und griechischen Version "Irritability", den höchsten Bekanntheitsgrad in der griechischen Stichprobe 2034GR hat. In der Rangreihe des Bekanntheitsgrades folgen die Items 17, "Fatigue", 14, "Negative body image" und 15 "Schoolwork difficulty". Anzunehmen ist, dass eventuell stichprobenspezifische oder sogar nationale Unterschiede in solchen Rangreihen des Bekanntheitsgrades der nach Punktwerten zusammengestellten Antworten auf die einzelnen Items zu beobachten wären. Vereinfacht ließe sich eine relative Bevorzugung und eine relative Ablehnung der jeweiligen Antworten im Vergleich zu anderen Antworten feststellen. So erweist sich in der griechischen Stichprobe 2034GR das Item 12, "Reduced social interest" als zu fast 89% abgelehnt und dabei in der umgepolten Rangreihe mit gut 85% gefolgt von Item 1, "Sadness". Mit 84% Ablehnung liegt Item 10, "Crying spells", unmittelbar dahinter. Danach folgen die Items 5, "Misbehavior" und 27, "Fighting". Diese fünf Items sind die unpopulärsten CDI-Antworten in der griechischen Stichprobe von Schulkindern. Dieser Feststellung kommt zumindest zweifache Bedeutung zu. Zum einen besagt sie, dass einige Befindlichkeiten von mindestens 80% der teilnehmenden Schulkinder in der griechischen Untersuchung 2034GR abgelehnt werden. Wahrscheinlich träfe

die Feststellung, diese fünf Befindlichkeiten seien „unpopulär“ nicht nur als statistische Aussage zu. Bedeutsam ist aber eben deswegen auch jede Antwort eines der beteiligten Kinder, die trotz der gruppenspezifischen Unpopularität dieser fünf Befindlichkeiten eine 1- oder 2- Punkte Antwort bei diesen Items geben. Solchen Antwort müsste dann jeweils ein aussagekräftiger Indikator für eine von der psychischen Befindlichkeit der Mehrheit aller Kinder der Stichprobe abgesetzte eigene Befindlichkeit sein. Das erweist sich in der Nachprüfung auch so. Diese Antworten zeigen die höchstn Korrelationen zum CDI-Gesamtpunktwert, der selbst ein Indikator für zu vermutende depressive Belastungen ist.

Die nachstehende Tabelle zeigt die Items des CDI mit der in dieser griechischen Stichprobe durch das Antwortverhalten der Schulkinder bekannt gewordenen Rangreihe der bevorzugten Wahlen von nach Punktwerten geordneten Antworten zu jedem CDI- Item. In dieser Tabelle ist die zuvor dargestellte Rangreihe der Ablehnungen umgepolt worden.

Tabelle 9.3b: Prozentuale Verteilung der nach Punktwerten pro Item gegebenen Antworten

Prozentualer Anteil pro Item	0 Punkte	1+2 Punkte	Rang
11 "Irritability"	45,1	54,9	1
14 "Negative body image"	50,6	49,4	2
17 "Fatigue"	50,6	49,4	3
15 "Schoolwork difficulty"	52,3	47,7	4
19 "Somatic concerns"	53,5	46,5	5
26 "Disobedience"	54,8	45,2	6
2 "Pessimism"	55,6	44,4	7
24 "Self-depreciation via peer comparison"	56,6	43,4	8
18 "Reduced appetite"	59,0	41,0	9
23 "School performance decrement"	59,7	40,3	10
3 "Self-depreciation"	65,2	34,8	11
16 "Sleep disturbance"	67,0	33,0	12
13 "Indecisiveness"	68,9	31,1	13
8 "Self-blame"	70,3	29,7	14
20 "Loneliness"	72,9	27,1	15
6 "Pessimistic worrying"	74,1	25,9	16
25 "Feeling unloved"	75,5	24,5	17
7 "Self-hate"	76,9	23,1	18
4 "Anhedonia"	77,2	22,8	19
21 "School dislike"	77,3	22,7	20
22 "Lack of friends"	77,4	22,6	21
27 "Fighting"	81,9	18,1	22
5 "Misbehavior"	82,5	17,5	23
10 "Crying spells"	84,0	16,0	24
1 "Sadness"	85,7	14,3	25
12 "Reduced social interest"	88,9	11,1	26

Obschon die Originalversion des CDI 27 Items hat, finden sich hier nur 16 Rangplätze. Sowohl das Suiciditem als auch das benutzte Ersatzitem fehlen in diesen Auflistungen.

Den zuvor mitgeteilten vier Aussagen zur Verteilung der CDI Gesamtpunktwerte sind drei weitere hinzu zu fügen.

9.5 AS: Mit ihrem Antwortverhalten auf die Fragen des CDI demonstrieren die griechischen Schulkinder über die relative Bevorzugung oder Ablehnung von Items in gruppenspezifisch unterschiedliche Gewichtigkeit der möglichen Antworten.

9.6 AS: Von 26 der benutzten CDI Items werden 11 mit mindestens 40% zustimmenden Antworten relativ bevorzugt gewählt. (Item 11, "Irritability"; 14, "Negative body image"; 17, "Fatigue"; 15, "Schoolwork difficulty"; 19, "Somatic concerns"; 26, "Disobedience"; 2, "Pessimism"; 24, "Self-depreciation via peer comparison"; 18, "Reduced appetite"; 23, "School performance decrement"). Diese Items dürfen als unter den Schulkindern der Stichprobe relative weit verbreitet angesehen werden.

9.7 AS: Von 26 der benutzten CDI Items werden 5 mit mindestens 80 abweisenden Antworten relativ deutlich als nicht auf die eigene Situation zutreffend abgelehnt. (Item 12, "Reduced social interest"; 1, "Sadness"; 10, "Crying spells"; 5, "Misbehavior"; 27, "Fighting") Mit dem Antwortverhalten der griechischen Schulkinder dieser Stichprobe erweist sich, dass 5 der CDI Items eine deutlich reduzierte Verbreitung unter den Kindern dieser Stichprobe haben.

Die Tabellen 9.2 und 9.3 geben die Antworten aller an der Durchführung des CDI teilnehmenden Kinder der Gesamtstichprobe wieder. Demnach ist dann noch nicht bekannt, welchen Beitrag die individuelle Bearbeitung jedes einzelnen CDI-Symptoms zum individuell erreichten CDI-Gesamtpunktwert geleistet hat.

Obschon die hier zuvor wiedergegebenen Tabellen mit den in ihnen zu findenden jeweiligen Werten vornehmlich deskriptiven Zwecken dienen, geben sie doch Anlass zu zwei Anmerkungen. Eine erste bezieht sich auf die selbstberichtete Betroffenheit der Kinder durch die Antworten auf die Items „sleep disturbance“ (Item 16), „fatigue“ (Items 17), „reduced appetite“ (Item 18) und „somatic concerns“ (Item 19). Die Prozentsätze der Kinder, die gegenüber diesen Symptomen berichten, „manchmal“ oder „häufig“ durch sie belastet zu sein, sind relativ hoch (33%, 49,4%, 41% und 46,5%). Alle diese Items entsprechen den im DTK zu findenden Symptomen der DTK-Skala 3, die psychosomatische Symptome innerhalb einer depressiven Störung zu messen vermag. Aus den Analysen der Daten dieser Skala wurde nach Einführung des von Pötz vorgeschlagenen Cutoffs ein relativ hoher Prozentsatz für schwer belastete griechische Kinder gefunden (siehe 7.5.2.3). Zumindest auf einer solchen numerischen Ebene stehen die hier zu betrachtenden und zu vergleichenden Ergebnisse aus CDI und DTK in Übereinstimmung. Wenn nur die Prozentsätze der Kinder zur Kenntnis genommen werden, die einerseits für diese CDI-Items angaben, selbst belastet zu sein und andererseits die Prozentsätze der Kinder betrachtet werden, die sich als durch die von DTK-Items repräsentierten psychosomatischen Symptome betroffen bezeichneten, ergeben sich übereinstimmende Prozentsätze. Zweifellos bliebe zu prüfen, ob nicht nur die Prozentsätze, sondern eben auch die Kinder identisch wären. Die Beobachtung ist hier aber erwähnt, weil sie entsprechende Hinweise auf den im Folgenden dann zu vollziehenden Vergleich der Informationen von DTK und CDI gibt. Die zweite Bemerkung bezieht sich auf die zwei CDI-Symptome, bei der die Zahl der Kinder, die als Antwort den höchsten Belastungsgrad wählen, die 2-Punkte Antwort, den höchsten Prozentsatz der Zustimmung erreicht.

Das ist bei den Fragen nach „schoolwork difficulty“ (Item 15) und „self-depreciation via peer comparison“ (Item 24) zu beobachten. Die relativ hohen Prozentsätze der Kinder, die sich von diesen, durch den CDI angesprochenen Problemen, „manchmal“ (Item 15 = 28,5% und Item 24 = 30,9%) oder sogar „sehr häufig“ (Item 15 = 19,2% und Item 24 = 12,4%) betroffen scheinen, könnten ein Hinweis darauf sein, dass das Erreichen eines von der Schule gesetzten Hauptzieles, das der Wissensvermittlung, mit beachtenswerten psychischen Problemen verbunden ist. Ein solcher Zusammenhang wäre, auch unabhängig von seiner Bedeutung im Hinblick auf die Existenz einer depressiven Störung, Anlass zum Nachdenken über Schule. Das wird als

hinreichender Anlass für einen entsprechenden Vermerk durch die nachstehende Aussage gewertet.

9.8 AS: Mit 47,7% zum einen 43,1% zum anderen sind erstens geminderter Selbstwert im Vergleich mit Gleichaltrigen (CDI-Item 24) und zweitens allgemeine Schulprobleme (CDI-Item 15) die Items, die in der griechischen Stichprobe 2034GR die beiden Probleminhalte, die den höchsten Grad an Erklärungen der eigenen Betroffenheit auf sich gezogen haben. Dieser Sachverhalt wird, unabhängig von der Bedeutung der Items innerhalb einer depressiven Störung, als eigenes Problem gewertet. Die jeweilige Betroffenheitsquote ist ungleich höher als alle für Schulkinder bekannt gewordenen depressionsspezifischen Prävalenzraten. Eben das wird hier als Indiz für eine über depressive Belastungen hinausreichende eigene Problematik gewertet.

9.4 Die CDI-Gesamtpunktwerte als Informationen über das Vorliegen von Hinweisen auf das Bestehen subklinischer und klinischer Ausprägungen an depressiven Störungen in der untersuchten Stichprobe griechischer Schulkinder

Eine schon mit seiner theoretischen Fundierung eingeführte Annahme unterstellt, dass höhere Gesamtpunktwerte im CDI auch Hinweise auf mutmaßlich bestehende unterschiedliche Grade an Belastungen durch depressive Symptome sind (u.a. Kovacs, 2004). Nahezu jede Anwendung des CDI wird sehr wahrscheinlich mit der Überzeugung geführt, dass höhere CDI-Gesamtpunktwerte zumindest relative Steigerungen von mutmaßlichen Belastungen anzeigen. Inzwischen darf diese Annahme als durch die empirische Praxis und durch die ihr unterliegenden Daten als bislang nicht widerlegt, sondern ganz im Gegenteil, als in ihrem Erklärungswert gefestigt angesehen werden. Die in diesen Erörterungen hier zuvor referierten Studien belegen diese Aussage. Insbesondere die Längsschnittstudien von Canals, Domènech-Llaberia, Fernández-Ballart und Martí-Henneberg (2002) dürften diese Feststellungen rechtfertigen.

Auszugehen ist dementsprechend davon, dass die CDI-Gesamtpunktwerte spezifische Hinweise auf das mutmaßliche Vorliegen depressiver Störungen bieten. Die CDI-Gesamtpunktwerte haben in ihrer Hinweisfunktion ein Ordinalskalenniveau. Höhere Werte zeigen unter Beachtung der geltenden statistischen Verfahrensregeln auch höhere Risiken für das Vorliegen und für die Entwicklung subklinischer und klinischer Ausprägungen von depressiven Störungen an.

Eingehend begründbar ist die Einschätzung, dass diese Leistung des CDI unter Einbeziehung der bekannt gewordenen Fehlerquoten funktional hinreichend sei für die Erfassung von depressionsspezifischen Risikogruppen. Als solche wären alle Gruppen von Kindern und Jugendlichen anzusehen, für die präventive oder kurative Interventionen als mutmaßlich notwendig eingeschätzt werden sollten.

Eine damit eng verbundene aber nicht identische Problematik ist die der diskriminativen Validität des CDI für die Erkennung depressiver Störungen nach DSM oder ICD. Das Erfassen von Risikogruppen verlangt ein möglichst sensitives Erkennen aller subklinischen und klinischen Ausprägungen depressiver Belastungen. Das Erkennen von depressiven Störungen ist demgegenüber eine spezielle diagnostische Leistung, die anderen Qualitätskriterien zu unterwerfen ist. Für diagnostische Tests wird beispielsweise ein Alpha Wert von mindestens .95 verlangt. Der CDI bleibt darunter. Der CDI ist kein diagnostisches Messinstrument. Demgegenüber sind die für den CDI geltenden Messdaten für dessen Verwendung als Hilfsmittel zur ökonomischen Bestimmung von depressionsspezifischen Risikogruppen hinreichend.

Mit diesen Anwendungen korrespondiert die sich durch die Literatur zum CDI hindurch ziehende Suche nach Trennwerten für die Unterscheidung von Kindern und Jugendlichen mit und ohne

klinischen Ausprägungen einer depressiven Störung. Wer einen optimalen Trennwert für CDI Gesamtpunktwerte sucht, ist gut beraten, vorab zu bestimmen, für welche Zwecke die mit dem CDI untersuchten Stichproben durch einen entsprechenden Trennwert nach Teilgruppen unterschieden werden sollen. Ergebnisse aus der in unterschiedlichen CDI-Studien betriebenen Suche nach optimalen Trennwerten (cut-offs) finden sich in einem vorangegangenen Abschnitt dieser Arbeit dargestellt (Kap. 9.1).

Eine solche Suche nach verlässlichen Trennwerten folgt einerseits messtheoretischen Zielen, wird aber andererseits auch aus alltagspraktischen Bedürfnissen notwendig. Für die Sicherung der umfänglichen und unterschiedlichen Versorgung einer Gesellschaft in allen Planungsbereichen des Zusammenlebens unersetzlich sind beispielsweise empirische Informationen über unvorhersehbar auftretende Risiken, aber auch Meldungen über die eventuell beobachtbare Rückbildung von krisenhaften Veränderungen. Eine gesellschaftlich überaus bedeutsame krisenhafte Veränderung wäre auch die Zunahme des Auftretens klinischer Ausprägungen depressiver Störungen (s. dazu Laux, in Laux (Hsg.), 2002). Nutzbare Abschätzungen von Steigerungen oder Rückbildungen in entsprechenden Entwicklungsprozessen sind nur unter Anwendung konstant bleibender Kriterien und Maße möglich. In der Epidemiologie sind Prävalenzen ein für diese Zwecke nutzbare Quantifizierung. Verlässliche Prävalenzraten setzen verlässliche Messungen voraus. Die Messpraxis benötigt dabei fast regelhaft Trennwerte für die von Fall zu Fall vergleichbare Trennung der Mitglieder einer Stichprobe in Teilgruppen, so beispielsweise in die von Kindern mit und ohne depressive Störungen. Da für die Untersuchung 2034GR eingangs die Aufgabe gestellt worden ist, Hinweise auf das mutmaßliche Bestehen depressiver Störungen unter griechischen Schulkindern zu erfassen, zu quantifizieren und bekannt zu geben, steht auch diese Untersuchung unter dem Zwang, nutzbare Trennwerte zu bestimmen oder zu übernehmen. Bei der Analyse der DTK-Daten in Kapitel 7 bestand Anlass, eingehend die speziellen Bedingungen der Bestimmung von Trennwerten für den DTK zu erörtern. Auch für die CDI-Daten sind Entscheidungen über die Art und Weise der Bestimmung von Trennwerten unumgänglich.

Der genannten Untersuchungsaufgabe entspricht die zur Beantwortung anstehende Frage, welche Schulkinder in dieser griechischen Untersuchung durch ihre subjektiven Selbsteinschätzungen über den Einsatz des CDI als mutmaßlich depressiv belastet zu identifizieren sind. Die Beantwortung dieser Frage verlangt den Einsatz von Trennwerten, die im Aufkommen von CDI-Gesamtpunktwerten die Punktbereiche unterscheiden lassen, die für das Auftreten unterschiedlicher Ausprägungsgrade von depressiven Belastungen relevante Indizien bieten. In dem Technischen Manual des CDI (2004) steht für die Anwendung des Verfahrens in nicht klinischen Stichproben die Empfehlung, einen CDI-Gesamtwert von 19 Punkten als kritischen Grenzwert für die Identifizierung von Kindern zu benutzen, die wahrscheinlich in einem klinischen Ausmaß depressiv belastet sind. Diese Empfehlung wird durch die den Ergebnissen unterschiedlicher Studien unterstützt (u.a. Kovacs, 1983, im Technischen Manual, S. 107; Craighead, Carry und Ilardi, 1995; Doerfler und Mit., 1988; Hodges, 1990; Kazdin, 1989; Lobovits und Handal, 1985; Smucker und Mit., 1986). Tabelle 9.4 zeigt die von Kovacs übernommenen Kategorien kritischer Werte, die für die Identifizierung unterschiedlicher Ausprägungsgrade von depressiven Belastungen auch auf die CDI Gesamtpunktwerte der in dieser Studie untersuchten Kinder angewendet wird.

Tabelle 9.4: Kritische CDI-Gesamtpunktwerte (nach Kovacs, 2002) und die durch sie zu identifizierenden Belastungsgruppen

CDI-GPW	Belastungsgruppe	Bezeichnung (*)
0 - 12	keine bis leichte depressive Belastung	CDI-Dep 1
13 - 18	moderate depressive Belastung	CDI-Dep 2
19 - 54	schwere depressive Belastung	CDI-Dep 3

(*): innerhalb dieser Arbeit benutzte Kennung für die nach CDI bestimmte Belastungsgruppe

Für diese Grenzwerte ist aus mehreren, hier zuvor bereits dargestellten CDI-Studien bekannt, dass mit ihnen nicht fehlerfrei zu arbeiten ist. Das gilt sowohl im Hinblick auf die Zahl der jeweils zu identifizierenden Kinder als auch im Bezug auf den mit dem CDI erfassten Ausprägungsgrad einer psychischen Belastung. Dieser Sachverhalt beeinträchtigt die Nutzung der mit dieser Arbeit erfassten Hinweise. Dieser Sachverhalt zwingt zur Einbeziehung der empirisch bekannt gewordenen Fehlerquoten. Die wird hier mit 25% angesetzt. Dieser Wert relativiert sich beträchtlich, wenn empirische Befunde zu den fehlerhaft zugeordneten Kindern beachtet werden. Zur Verdeutlichung dieser Position wird hier noch einmal und ausdrücklich auf die Längsschnittstudie von Canals, Domènech-Llaberia, Fernández-Ballart und Martí-Henneberg (2002) verwiesen. Sie demonstriert die potentielle Nützlichkeit von CDI-Messungen überaus nachdrücklich. Auch nach Ergebnissen aus früheren CDI-Studien wird durch die Anwendung der von Kovacs übernommenen Trennwerte eine hohe Zahl von etwa 75% bis 90%, der relativ unbelasteten Kinder richtig charakterisiert, weil sie sich im weiteren Entwicklungsverlauf als durchaus depressiv belastet erweisen. Ebenfalls zutreffend werden zwischen 55-80% derjenigen Kinder identifiziert, die nach einer klinischen Untersuchung die psychiatrische Diagnose einer reinen depressiven Störung bekommen hätten. In der Gruppe der nach Anwendung des vorgeschlagenen Cutoffs als in klinischem Ausmaß belasteten Kinder werden aber auch Kinder zu finden sein, die eher der Diagnose einer komorbiden emotionaler Störung entsprechen. Vor allem trifft dies auf Angst- oder Verhaltensstörungen zu. Weiterhin werden auch Kinder zu finden erfasst, die keiner klinischen Diagnose entsprechen. Geltend gemacht wird auch, dass viele von denen, empirisch nachweisbar, klinisch bedeutsame psychosoziale Beeinträchtigungen aufweisen und entsprechenden Risikofaktoren für die Entwicklung einer emotionalen Störung ausgesetzt sind.

Tabelle 9.5 zeigt die Zahl der an der Untersuchung 2034GR beteiligten Kinder, die nach ihren Selbstzuschreibungen von depressiven Symptomen im CDI als alltäglich, subklinisch oder klinisch belastet gelten sollen.

Tabelle 9.5: Verteilung der Kinder aus Stichprobe 2034GR auf nach Schweregraden unterschiedenen Gruppen mit mutmaßlich unterschiedlicher Belastung durch depressive Symptome

Belastungsgruppe	CDI-Dep 1	CDI-Dep 2	CDI-Dep 3
CDI-GPW	0 - 12	13 - 18	19 - 54
Schülerzahl	1348	369	296
und Prozentsatz	67,0%	18,3%	14,7%

Die in Tabelle 9.5 dargestellten Prozentsätze von Kindern mit unterschiedlicher Ausprägung von depressiven Belastungen nach CDI basieren auf der weiter oben beschriebenen Verteilung der

CDI-Gesamtpunktwerte für die untersuchte Stichprobe. Die an der Stelle der angeführten Grenzwerte dort zu findenden Dezimalzahlen wurden, den Regeln folgend, gerundet.

Das Ergebnis der Durchführung eines nonparametrischen Tests (χ^2) bestätigt statistisch, dass die Anwendung der Cutoffs zu Gruppenbildungen führt, die signifikant von Zufallsverteilungen abweichen. ($\chi^2= 1028,551$; $p= ,000$). Demnach dürfen die in Tabelle 9.5 zu findenden Informationen zur Formulierung des nachfolgenden Arbeitsergebnisses genutzt werden.

9.9 AS: Über den Einsatz des CDI in einer Stichprobe von 2034 griechischen Kindern hat sich unter Anwendung des von Kovacs empfohlenen Cutoff-Wertes von CDI Gesamtpunktwert 19 ergeben, dass 14,7% der an der Untersuchung teilnehmenden Schulkinder mutmaßlich vom Auftreten depressiver Symptome in einem für klinische Ausprägungsgrade charakteristischem Ausmaß berichten und weiterhin etwa 18% der Kinder auf depressive Symptome verweisen, die das Bestehen subklinischer Belastungen annehmen lassen.

In Übereinstimmung mit der Vorgehensweise, die bei der Analyse der DTK-Daten gewählt worden war, wird auch für das CDI ein zweites Kriterium für die Identifizierung der mutmaßlich klinisch belasteten Kinder in der griechischen Stichprobe verwendet. Dieses Kriterium wurde als „Frequenzunterschreitung von $GPW=0$ “ in der Verteilung der Gesamtpunktwerte des jeweiligen Verfahrens und als verteilungsabhängiges Kriterium bezeichnet.

Die Verwendung dieses Kriterium basiert auf der Annahme, dass zumindest ab diesem Punkt eine innerhalb der untersuchten Stichprobe beobachtbare kritische Abweichung vom zuvor beobachteten Verlauf der Verteilung der individuell erbrachten CDI-Gesamtpunktwerte gesehen werden darf. Dieser Trennwert fällt laut seiner Definition für die CDI-Verteilung mit diesem Gesamtpunktwert zusammen, „der erstmals nach dem $CDI-GPW=0$ unter dessen prozentualem Anteil an der gesamten Stichprobe von griechischen Schulkindern bleibt“.

Nach Verwendung dieses Trennwertes (23,88 bzw. 24) sind 4,2% der Teilnehmer an der Untersuchung 2034 an GR als mutmaßlich in klinischem Ausmaß depressiv gestört zu betrachten.

Die nachstehende Tabelle fasst die über die zwei durchgeführten Schätzungen quantifizierten Hinweise auf Kindern mit mutmaßlich bestehenden klinischer Ausprägungsgraden einer depressiven Symptomatik zusammen.

Tabelle 9.6: Nach CDI-Gesamtpunktwerten geschätzte Quoten von Hinweisen auf das Auftreten depressiver Störungen in der untersuchten Stichprobe 2034GR

Nr. der Schätzung	Trennwert	Zahl der Kinder (n= 2013)	Prozentsatz
1. Schätzung	Trennwert 19 nach Kovacs	296	14,7
2. Schätzung	Kriterium der Frequenzunterschreitung nach Kampisiou	85	4,2

Die in der voran stehenden Tabelle zu findenden Analyseergebnisse werden mit der nachstehenden Aussage zusammengefasst.

1. Nach der Analyse der in der Untersuchung 2034GR gewonnenen CDI-Gesamtpunktwerte wird begründet angenommen, dass die Quoten für Hinweise auf das Auftreten klinischer Ausprägungen depressiver Störungen bei den untersuchten griechischen Schulkindern mit hoher Wahrscheinlichkeit zwischen 4,2% und 14,7% der Gesamtzahl der Mitglieder dieser Stichprobe liegt.

2. Die Spanne zwischen den Mengen von 4,2% und 14,7% an Hinweisen dürfte sehr wahrscheinlich durch Kinder mit graduell abnehmenden Risiken besetzt sein.

Kapitel 10

Beziehungen zwischen CDI-Gesamtpunktwerten und verschiedenen Teilstichproben in dieser Studie

Nachdem im vorangegangenen Teil Informationen über das Antwortverhalten der Kinder auf die Fragen des CDI gegeben werden konnten und die Bedeutung ihrer Antworten als Indizien zum Auftreten niedriger oder höherer Frequenzen von Hinweisen auf mutmaßlich vorhandene depressive Störungen erörtert wurde, folgt dem die Untersuchung möglicher Beziehungen zwischen CDI-Gesamtpunktwerten und unterschiedlichen persönlichen Merkmalen der untersuchten Schulkinder.

Die Vorgehensweise innerhalb dieses Teils der Erörterungen folgt der Struktur, die bei den entsprechenden Analysen für die aus der Anwendung des DTK gewonnenen Daten (Kapitel 8). Das dient dem Ziel, direkte Vergleiche zwischen den Ergebnissen aus den beiden eingesetzten und in ihrem Datenaufkommen analysierten Verfahren zur Selbsteinschätzung des Auftretens depressiver Symptome zu ermöglichen. Auf Theorien über das Auftreten depressiver Störungen bei Kindern und Jugendlichen bezogene Erwartungen zum Datenaufkommen waren, insbesondere in Bezug auf die mit dieser Studie untersuchten Risikofaktoren, bereits im Kapitel 8 dargestellt, begründet und erörtert worden. Die den DTK betreffenden Ausführungen gelten gleichermaßen für den CDI und werden deshalb in diesem, erneut auf den CDI eingehenden Teil der Erörterungen nicht wiederholt. Ergänzungen zu den mitzuteilenden Ergebnissen sollen, wo sich das als möglich erweist, durch Hinweise auf entsprechende Befunde aus anderen CDI-Studien gegeben werden.

10.1 Geschlechts- und Alterseffekte in Punktwerten aus der Anwendung des CDI

In der durch Kovacs (2004) zitierten normierenden Stichprobe des CDI (1266 Kinder zwischen 7 und 16 Jahren) haben die Jungen signifikant höhere Mittelwerte in der Summe der CDI-Punkte als die Mädchen. Aber im Hinblick auf das Alter der Kinder wurde kein Haupteffekt gefunden. Ebenfalls hatte die Interaktion von Geschlecht und Alter in dieser Stichprobe keinen signifikanten Einfluss auf die erbrachten CDI-Mittelwerte (Kovacs, 2004, im CDI-Manual).

Twenge und Nolen-Hoeksema (2002) führten mit den Daten aus 310 in den USA durchgeführten CDI-Studien eine Metaanalyse durch. Sie analysierten dabei die Effekte von Geschlecht, Alter, sozioökonomischem Status, Nationalität und von Alterskohorten auf die CDI-Gesamtscores. Im Hinblick auf die hier diskutierten Geschlechts- und Alterseffekte auf die Mittelwerte der CDI-Gesamtpunkte teilten die Autoren, dass „...Girls' depressive symptoms were fairly steady until age 12, when they began to increase. Girls' CDI scores reached a peak at age of 15 and remained high at age 16. Boys' depressive symptoms remained at about the same level regardless of age; the only exception was a slight elevation at age 12" (Twenge & Nolen-Hoeksema, 2002; S. 585). Darüber hinaus wurde in der Teilstichprobe der Kinder bis zu 12 Jahren kein signifikanter Unterschied gefunden, wobei in der Altersgruppe zwischen 13-16 Jahren die Mädchen signifikant höhere Durchschnittswerte als die Jungen aufwiesen. In der Diskussion ihrer Ergebnisse kommentieren die Autoren die gefundenen Limitationen vor einem depressionsspezifischen theoretischen Hintergrund. Wenn nur die jeweils zu berechnenden Mittelwerte als Indizien für alters- und geschlechtsspezifischen Einflüsse auf die im CDI selbstberichteten depressiven Belastungen herangezogen werden, sehen die Autoren die mitgeteilten Werte als aussagekräftig an. Vorgeschlagen wird, in der CDI-Forschung diese Effekte auch auf anderen Belastungsebenen zu überprüfen. Für seine deutschsprachige CDI-Version fand Stiensmeier (1988) keine

signifikanten geschlechtsspezifischen Unterschiede. In der Eichstichprobe von Stiensmeier-Pelster, Schürmann und Duda (1989) wird über Unterschiede für die Mädchen berichtet. Reicher und Rossmann (1991) fanden demgegenüber keine die Gesamtpunktwerte betreffenden Divergenzen, wohl aber signifikante geschlechtsspezifische Unterschiede auf Itemebene. Nach Reinhard et al. (1990) traten in der dort untersuchten Stichprobe keine Geschlechtsunterschiede auf.

Alle für den Vergleich mit den aus einer Metaanalyse hervorgegangenen Daten von Twenge & Nolen-Hoeksema (2002) bestimmten Hinweise auf Abweichungen einzelner Untersuchungen von den amerikanischen Vergleichswerten sind nicht gegen das Argument geschützt, dass die Metaanalyse, basierend auf 310 Einzeluntersuchungen, das Auftreten von Abweichungen implizit vorhersagt. Insofern ist die Diskussion von Einzelergebnissen der Metaanalyse methodisch unterlegen. Jede Einzeluntersuchung kann jedem Abschnitt aus der Verteilung entstammen, die in die Metaanalyse eingegangen ist, auch den vom Mittelwert dieser Verteilung extrem entfernten Abschnitten. Die Diskussion von Einzelergebnissen ist insofern wenig hilfreich und dürfte, für sich genommen, auch endlos sein.

In der Literatur zum CDI wird allgemein die Auffassung vertreten, dass die Mittelwerte der Gesamtpunktwerte des CDI mit zunehmendem Alter signifikant zunehmen. Der von Reicher und Rossmann gegebene Bericht über das Ausbleiben von Alterseffekten für von ihnen untersuchte 5. und 6. Schulklassen steht dazu nicht im Widerspruch.

Innerhalb der hier zu führenden Erörterungen wird den durch Twenge & Nolen-Hoeksema (2002, S. 585) errechneten Ergebnissen als Orientierungsdaten vertraut.

10.2 Geschlechts- und Alterseffekte in den CDI-Gesamtwerten aus der griechischen Untersuchung 2034GR

Zur Untersuchung von Geschlechts- und Alterseffekten in den Gesamtpunktwerten (CDI-GPW) aus der Anwendung des CDI in der hier zu erörternden griechischen Untersuchung wird die selbe statistische Vorgehensweise angewendet, mit dem auch alle andere CDI-Autoren gearbeitet haben, nämlich eine zweifaktorielle Varianzanalyse auf einem Signifikanzniveau von $p < 0,001$.

Aus der durchgeführten Varianzanalyse ergab sich für die Gesamtstichprobe ein signifikanter Haupteffekt für den Faktor Alter [$F(1,2012) = 12,628$; $p = 0,000$; Eta-Quadrat = 4,2%].

Aus der ersten Spalte der nachstehenden Tabelle ist zu entnehmen, dass mit zunehmendem Alter der Kinder die von ihnen erbrachten Mittelwerte im CDI, abgesehen von einem leichten Rückgang bei den 10jährigen Kindern, insgesamt zunehmen.

Diese Analyse hat aber keinen signifikanten Effekt für den Faktor Geschlecht gefunden.

Die zweifache Interaktion zwischen Geschlecht und Alter hat sich mit der Varianzanalyse ebenfalls als statistisch signifikant erwiesen [$F(7,2012) = 12,628$; $p = 0,039$; Eta-Quadrat = 0,7%]. Über separat geführte t-Tests wurde in diesem Zusammenhang gefunden, dass vor allem die 12jährigen und auch die 13jährigen Mädchen signifikant höhere Mittelwerte im CDI erreichen als ihre gleichaltrigen Mitschüler (siehe Tabelle 10.1, Spalte „t-Test“). Die praktische Relevanz dieser Interaktion bleibt jedoch aufgrund des geringen Anteil der Variabilität der Messwerte (Eta-Quadrat 0,7%) fraglich.

Tabelle 10.1: Mittelwerte (M) und Standardabweichung (SA) der CDI-Werte in der Gesamtstichprobe und in geschlechtsspezifischen Altersstufen nebst T-Test

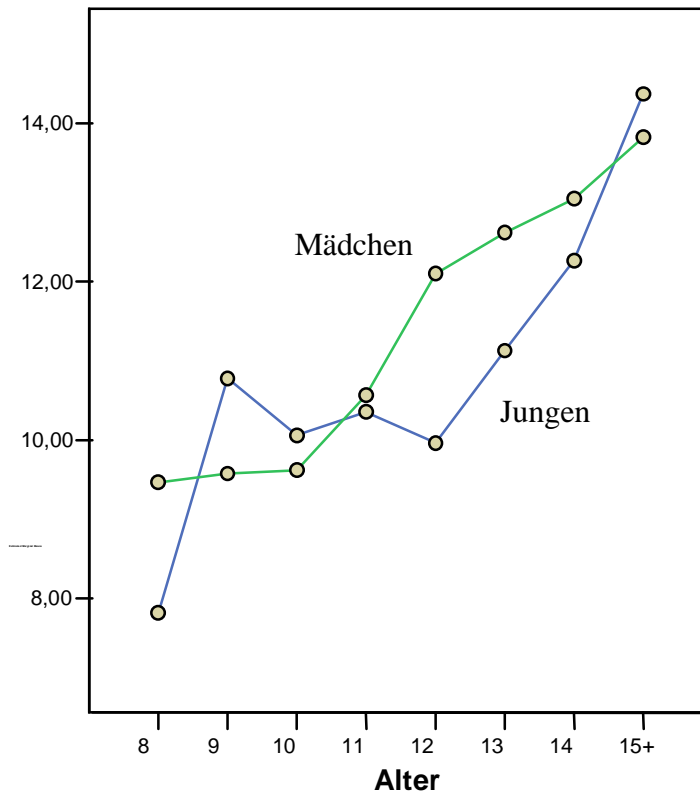
Alter	Gesamtstichprobe			Jungen			Mädchen			T-test	
	n	M	SA	n	M	SA	n	M	SA	t	P
8	171	8,5688	5,88	93	7,8164	5,16	78	9,4660	6,56	-1,840	,068
9	218	10,1560	6,43	105	10,7802	6,40	113	9,5759	6,43	1,385	,168
10	258	9,8614	5,76	143	10,0578	5,68	115	9,6171	5,86	,611	,542
11	283	10,4580	6,13	145	10,3560	6,18	138	10,5652	6,11	-,286	,775
12	329	11,0664	6,77	159	9,9601	6,44	170	12,1011	6,93	-2,897	,004
13	337	11,8668	6,21	170	11,1299	6,28	167	12,6170	6,06	-2,212	,028
14	322	12,6196	6,78	176	12,2668	7,22	146	13,0448	6,20	-1,025	,306
15+	95	14,1121	6,87	50	14,3723	6,65	45	13,8231	7,17	,388	,699
Gesamt	2013	11,0418	6,49	1041	10,7128	6,46	972	11,3942	6,51	-2,357	,019

Zeichenerklärung: n= Zahl der entsprechenden Teilstichprobe, M= Mittelwert, SA= Standardabweichung

In Abbildung 10.1 sind die zuvor mitgeteilten Ergebnisse graphisch dargestellt.

Abbildung 10.1: Scores in der CDI-Skala in Abhängigkeit von Geschlecht und Alter

Estimated Marginal Means of cdiVI



Nach den mitgeteilten Analyseergebnissen sind die nachstehenden Feststellungen zu vermerken.

10.1 AS: In der Produktion der CDI-Gesamtpunktwerte wurden Effekte des Alters und der Interaktion zwischen Alter und Geschlecht gefunden.

10.2 AS: Mit zunehmendem Alter nehmen die erbrachten Durchschnittswerte für die Gesamtstichprobe zu.

10.3 AS: Die 12- und 13jährigen Mädchen erreichen mit ihren Gesamtpunktwerten signifikant höhere Mittelwerte als die gleichaltrigen Jungen.

10.4 AS: Der zuvor erwähnte signifikante Effekt aus der Interaktion Geschlecht-Alter erklärt nur einen geringen Teil der Messvariationen, so könnte dieser Effekt mit dem relativ großen Umfang der Stichprobe zusammenhängen.

10.5 AS: Die hier dargestellten Ergebnisse stimmen mit entsprechenden Befunden aus anderen Untersuchungen nur partiell überein, so beispielsweise mit denen, über die Twenge & Nolen-Hoeksema (2002, S. 585) berichten. Die auftretenden Unterschiede liegen allerdings im Bereich der für Vergleiche zwischen Einzeluntersuchungen und Angaben aus Metaanalysen zu erwartenden Abweichungen.

10.3 Variationen der Gesamtpunktwerte nach dem Geschlecht der Schulkinder

Die aus der Durchführung des CDI bekannt werdenden Gesamtpunktwerte waren die Datenbasis für die Bildung von Risikogruppen für das Auftreten depressiver Störungen. Die verwendete Bezeichnung betraf das für Entscheidungen über das Vorliegen von depressiven Störungen einzugehende Risiko von Fehlurteilen der einen und der anderen Art. Da die gebildeten Risikogruppen DEP1, DEP2 und DEP3 sich durch die von Gruppe zu Gruppe und auch innerhalb der Gruppen ansteigenden CDI Gesamtpunktwerte unterscheiden, nimmt das Risiko, durch die Abweisung des Verdachts eine bestehende Störung zu übersehen mit den Risikogruppen aufsteigend und in ihnen mit dem steigenden Gesamtpunkt看wert zu.

In der nachfolgenden Kreuztabelle wird geprüft, ob das Geschlecht der Kinder auf deren Zuweisung an die gebildeten Risikogruppen, die auf eine unterschiedliche Wahrscheinlichkeit für das Vorliegen einer depressiven Störung hinweisen, einen isolierbaren Einfluss hat. Zur Verbesserung der Übersichtlichkeit wurden die Gruppen Dep1 und Dep2 für diese Tabelle zusammengefasst.

Tabelle 10.2: Geschlechtsspezifische Verteilung der Kinder aus der griechischen Stichprobe 2034GR auf zwei Gruppen mit unterschiedlicher Wahrscheinlichkeit für das Vorliegen depressiver Belastungen nach CDI (Cutoff GPW= 19)

Geschlecht		Dep1+2	Dep3	Gesamt
Jungen	Anzahl	899	142	1041
	Erwartete Anzahl	887,9	153,1	1041
	Prozent	86,4%	13,6%	100%
Mädchen	Anzahl	818	154	972
	Erwartete Anzahl	829,1	142,9	972
	Prozent	84,2%	15,8%	100%
Gesamt	Anzahl	1717	296	2013
	Prozent	85,3%	14,7%	100%

Zeichenerklärung: Dep1+2= Gruppe niedriger oder mittlerer depressiver Belastung nach CDI (bis zu 18 GPW), Dep3= Gruppe hoher depressiver Belastung nach CDI (ab 19 GPW)

Die in Tabelle 10.2 dargestellte geschlechtsspezifische Verteilung der beteiligten Kinder auf die CDI-Belastungsgruppen Dep1 und Dep2 weicht nicht signifikant von den Erwartungswerten ab [$\chi^2(1) = 1,945; p = 0,163$]. Daraus ergibt sich die folgende Aussage:

10.6 AS: Die geschlechtsspezifische Verteilung der beteiligten Kinder auf zwei nach CDI Gesamtpunktwerte gebildeten Risikogruppen (Dep1= niedrige oder mittlere depressive Belastung, Dep2= vermutlich klinische depressive Belastung, Cutoff= 19 CDI-GPW) zeigt keine signifikanten Variationen.

Mit der entsprechenden Analyse an den Daten aus dem DTK wurde gefunden, dass die Mädchen in der DTK-Gruppe höher depressiver Belastung depBel 3 signifikant überrepräsentiert sind (siehe Kapitel 8.3).

Bevor die Ergebnisse über das Geschlecht der Kinder als Risikofaktor für eine erhöhte depressive Belastungen nach den in dieser Studie benutzten zwei depressionsspezifischen Verfahren, DTK und CDI, als widersprüchlich bewertet wird, sollte erneut darauf hingewiesen werden, dass die Gesamtstichprobe eine Altersspanne von 8 bis 18 Jahren deckt. Daraus ergibt sich die Möglichkeit, dass in den Altersgruppen bestehende geschlechtsspezifische Unterschiede für die Gesamtpunktwerte des CDI einander aufheben könnten. An einer nachfolgenden Stelle dieses Arbeitsteiles wird auch für die CDI-Daten untersucht, ob und in welcher Weise das Alter mit dem Geschlecht der Kinder auf deren Verteilung auf die definierten CDI-Belastungsgruppen einwirkt.

10.4 Variationen der CDI Gesamtpunktwerte nach dem Alter der untersuchten Kinder

Die nachstehende Tabelle stellt die altersspezifische Verteilung der an dieser Untersuchung beteiligten Schulkinder auf zwei nach CDI-GPW gebildeten Belastungsgruppen Dep1+2 und Dep3 dar. Zu wiederholen ist, dass die Altersgruppe der sogenannten 15+-jährigen Kinder auch ältere Kinder und dies bis zu 18 Jahren einschließt. Darauf verweist auch das Pluszeichen neben der Altersangabe (15+). Die Zusammenfassung wurde wegen der niedrigen Frequenzen in den Altersklassen 16 bis 18 für notwendig und vertretbar angesehen.

Tabelle 10.3: Altersspezifische Verteilung der untersuchten Kinder auf die Risikogruppen Dep1+2 und Dep3

Alter		Dep1+2	Dep3	Gesamt
8	Anzahl	157 **	14 **	171
	Erwartete Anzahl	145,9	25,1	171,0
	Prozent	91,8%	8,2%	100,0%
9	Anzahl	192	26	218
	Erwartete Anzahl	185,9	32,1	218,0
	Prozent	88,1%	11,9%	100,0%
10	Anzahl	229	29	258
	Erwartete Anzahl	220,1	37,9	258,0
	Prozent	88,8%	11,2%	100,0%
11	Anzahl	250	33	283
	Erwartete Anzahl	241,4	41,6	283,0
	Prozent	88,3%	11,7%	100,0%

(Fortführung Tabelle 10.3)

Alter		Dep1+2	Dep3	Gesamt
12	Anzahl	277	52	329
	Erwartete Anzahl	280,6	48,4	329,0
	Prozent	84,2%	15,8%	100,0%
13	Anzahl	281	56	337
	Erwartete Anzahl	287,4	49,6	337,0
	Prozent	83,4%	16,6%	100,0%
14	Anzahl	259 **	63 **	322
	Erwartete Anzahl	274,7	47,3	322,0
	Prozent	80,4%	19,6%	100,0%
15 +	Anzahl	72 **	23 **	95
	Erwartete Anzahl	81,0	14,0	95,0
	Prozent	75,8%	24,2%	100,0%
Gesamt	Anzahl	1717	296	2013
	Prozent	85,3%	14,7%	100%

Zeichenerklärung: Dep1+2= Gruppe niedriger oder mittlerer depressiver Belastung nach CDI (bis zu 18 GPW), Dep3= Gruppe hoher depressiver Belastung nach CDI (ab 19 GPW), **= Felder, die von den Erwartungen signifikant abweichen

Die in Tabelle 10.3 dargestellte altersspezifische Verteilung der Kinder auf die CDI-Risikogruppen Dep1 +2 und Dep3 weicht insgesamt und dann hochsignifikant von den Erwartungswerten ab [$\chi^2(7) = 25,903$; $p = 0,001$].

Die eingeführten Sternchen verweisen auf Einzelfelder, bei denen die jeweils gerechneten korrigierten und standardisierten Residuen den kritischen Wert von $\pm 1,96$ bzw. nach Aufrundung durch SPSS $\pm 2,0$ erreichen oder übersteigen.

Die nachstehenden Aussagen fassen die in der vorangegangenen Tabelle enthaltenen Ergebnisse zusammen.

10.7 AS: Die altersspezifische Verteilung der Jungen und Mädchen weicht in der Gesamttabelle nach chi-Quadrat hochsignifikant von zufällig auftretenden Werten ab.

10.8 AS: Für die analysierte Tabelle werden in der Signifikanzprüfung allgemein über die Randsummen berechnete Erwartungswerte benutzt. Durch die hochsignifikante Abweichung der in der Tabelle aufzufindenden Variationen wird ausgesagt, dass die über alle Altersstufen hinweg und für alle Kinder berechneten Erwartungswerte von den beobachteten Werten in stärkerem Masse als zufällige Abweichungen das könnten über- oder unterschritten wird.

10.9 AS: Über die Verwendung korrigierter und standardisierter Residuen wird die Regelmäßigkeit dieser für die gesamte Tabelle festgestellten Abweichungen mit den Signifikanzen für einzelne Tabellenfelder erkennbar. Danach wird die über die Berechnung der Randsummen für die Gesamttabelle aufgestellte modellartige Regelmäßigkeit der Verteilung durch die beobachteten Werte an drei benennbaren Stellen durchbrochen.

10.10 AS: Die erste Nichteinhaltung der modellartigen Regelmäßigkeit der Verteilung betrifft die niedrigste Altersstufe in der Tabelle. Die 171 Kinder der Altersstufe 8 Jahre überbieten die Zahl der erwarteten Kinder in Dep1+2 und bleiben zu fast 45% unter der für Dep3 verlangten Präsenz. Beide Abweichungen sind depressionsspezifisch gleichsinnig gerichtet. Die Altersklasse der 8jährigen Kinder bleibt signifikant unter den über mehrere und höhere Altersstufen hinweg berechneten Erwartungswerten.

10.11 AS: Die zweite Verletzung der modellartigen Regel betrifft die Altersstufen der 14- und 15+-jährigen Kinder. Diese Kinder oder Jugendlichen weichen von den über mehrere Altersstufen hinweg berechneten Erwartungen ab.

10.12 AS: Die Angehörigen der Teilgruppe der älteren Kinder und Jugendlichen (14 Jahre und älter) übersteigen ihre Präsenz in der Gruppe Dep3 um 40% (14 Jahre und älter).

10.13 AS: Die mitgeteilten Ergebnisse gehören zu einer über die Altersgruppen hinweg darzustellenden Verteilung der auf die jeweilige Altersgruppe prozentual bezogenen Anteile an Angehörigen der Risikogruppe Dep3. Die Kurve dieser bei den 8jährigen Kindern beginnenden Verteilung hat die nachstehenden Werte (siehe auch die vorangegangene Tabelle): 8,2%**, 11,9%, 11,2%, 11,7%, 15,8%, 16,6%, 19,6%** , 24,2%**.

Da in der durchgeführten griechischen Untersuchung zwei Verfahren zur Selbsteinschätzung des Auftretens von Anzeichen für depressive Symptome verwendet worden sind, bietet sich im gegebenen Zusammenhang an, die über CDI festgestellten altersspezifischen Variationen der Anteile an der jeweils höchsten Risikogruppe des CDI, dep3, mit denen an der entsprechenden Risikogruppe nach Punktwerten im DTK, DepBel3, zu vergleichen. Die nachfolgende Tabelle bietet einen Überblick über die Richtung der auftretenden signifikanten Abweichungen von den für die Tabellen berechneten Erwartungswerten für die Zugehörigkeit von Kindern zu den Risikogruppen. Abweichungen von den Erwartungswerten treten auf. Die Tabellen weichen insgesamt signifikant von Erwartungswerten ab. Die nachstehende Tabelle gibt an, auf welcher Altersstufe diese Abweichungen liegen und in welche Richtung sie gehen, ob sie Unterschreitungen (Hinweis: Minus) oder Überschreitungen (Hinweis: Plus) der Erwartungswerte sind. Die Hinweise sind Kreuztabellen entnommen, für deren Tabellenfelder korrigierte standardisierte Residualwerte berechnet worden sind.

Tabelle 10.4: Richtung der auf den Altersstufen von statistischen Erwartungen abweichenden Zuordnungen von Kindern zu den Risikogruppe Dep3 (CDI) und DepBel3 (DTK)

Alter	Richtung der Abweichung	
	CDI-Ergebnisse	DTK-Ergebnisse
8	Minus	Minus
9	=	Minus
10	=	=
11	=	Minus
12	=	=
13	=	Plus
14	Plus	Plus
15+	Plus	Plus

Zeichenerklärung: Minus= gegenüber Erwartungswerten geringeres Aufkommen an mutmaßlich in klinischer Ausprägung depressiv belastete Kinder, Plus= ein gegenüber Erwartungswerten höheres Aufkommen; mit = wird ein den Erwartungswerten folgendes Aufkommen angezeigt.

Aus den mit obiger Tabelle 10.4 dargestellten Angaben zur Richtung der nach CDI und DTK auftretenden Abweichungen von erwarteten Zuweisungen an Kindern zu den Risikogruppen dep3 (nach CDI) und DepBel3 (nach DTK) folgern die nachstehenden Aussagen.

10.14 AS: Die Positionierung der von Erwartungswerten abweichenden Zuweisungen an Kinder und Jugendlichen zu den Risikogruppen dep3 und DepBel3 stimmen in 6 von 9 Abweichungen überein und in 3 Positionen nicht.

10.15 AS: Die Richtung der zwischen CDI und DTK übereinstimmenden 6 Abweichungen ist identisch.

10.16 AS: Alle 9 über die Messungen in CDI und DTK auftretenden signifikanten altersspezifischen Abweichungen von den Erwartungswerten für die Zuweisung an Kindern zu den Risikogruppen dep3 und DepBel3 stehen in voller Übereinstimmung zu den weiter oben graphisch veranschaulichten hypothetischen Erwartungen zur altersspezifischen Verteilung der am Grundwert einer jeden Stichprobe orientierten Prozentsätzen des Anteils von Kindern und Jugendlichen in den mit Verfahren der Selbstzuweisungen ermittelten höchsten Risikogruppen für das Bestehen depressiver Störungen.

10.17 AS: Die in dieser Stichprobe befindlichen 9-, 10-, 11-, 12- und 13jährigen Kinder zeigen mit ihren Frequenzen in den CDI-Belastungsgruppen Dep1 und Dep2 nur zufällige Abweichungen.

10.18 AS: Über den Einsatz des CDI und des DTK in einer Stichprobe von 2034 griechischen Kindern ergibt sich, dass mit wachsendem Alter die Hinweise auf das Bestehen depressiver Störungen mit klinischen Ausprägungen allgemein zunehmen.

10.19 AS: Als kritisches Lebensalter für das gegenüber vorangegangenen alterstypischen Zuwachsraten überproportionale Ansteigen der Prozentsätze an Kindern und Jugendlichen mit hohem Risiken für das Auftreten depressiver Störungen sind nach den dargestellten Ergebnissen die Altersstufen ab dem Alter von 14 Jahren anzusehen.

10.20 AS: Vermutet wird ein deutliches Absinken der prozentualen Anteile an risikivoll belasteten Persönlichkeiten ab dem 18. Lebensjahr.

10.21 AS: Der hier mitgeteilte Arbeitsbefund stimmt tendenziell mit den in der einschlägigen Literatur vertretenen Positionen überein, nach denen eine erhöhte Vulnerabilität zur Entwicklung depressiver Störungen zwischen dem 12 und 14 Lebensjahr erwartet wird.

10.5 Variationen der CDI-GPW aus der Interaktion von Alter und Geschlecht der untersuchten Kinder

Nun soll auch für die über den CDI gewonnenen Daten untersucht werden, ob die Interaktion vom Alter und Geschlecht der beteiligten Kinder zu überzufälligen Variationen der Verteilung der Kinder auf die zwei CDI-Risikogruppen führt. Diese Risikogruppen waren in Korrespondenz mit den CDI Gesamtpunktwerten (CDI-GPW) gebildet worden. Beginnend beim GSPW 0 wurden die Kinder nach ihrem individuellen Punktwert den Risikogruppen dep1, dep2 und dep3 zugewiesen. Im Kapitel 9 finden sich die als Trennwerte zwischen den Gruppen bestimmten GSPW und die Begründungen für ihre Wahl. Die drei Gruppen bilden, beginnend bei dep1, in ansteigender Folge über dep2 und dep3 insgesamt drei Risikobereiche für das Bestehen einer depressiven Störung ab. Die CDI-GSPW dienen als Kennwerte für die Zuweisung von Kindern zu einer der drei Gruppen. Auch die auf die Risikogruppen zu beziehende Validitätsproblematik findet sich in Kapitel 9 eingehend erörtert. Die Wahrscheinlichkeit des Bestehens einer individuell gegebenen depressiven Störung nimmt von dep1 über dep2 zu dep3 hin kontinuierlich zu. Die Wahrscheinlichkeit, mit der Zuweisung eines Schulkindes auch ein Kind mit bestehender depressiver Störung getroffen zu haben, ist relativ hoch. Die Treffsicherheit ist aus mehreren Gründen nur abzuschätzen, ein wesentlicher Grund ist, dass auch ärztliche Diagnosen durch Fehldiagnosen belastet werden. Mit den Zuweisungen zur Risikogruppe dep3 dürfte eine geschätzte Treffsicherheit für die Identifizierung ausgeprägter klinischer Depressionen und deren Ränder von etwa 80 % angenommen werden dürfen. Die Einführung dreier Risikogruppen bringt den Vorteil einer größtmöglichen Trennung der klinisch bedeutsamen Gruppe niedrigster und höchster Belastung. Die Existenz der Mittelgruppe dep2 verhindert Überlappungen des Klientels von dep1 und dep3. So wird für beide Gruppen eine relative Homogenität erreicht. Die Risikogruppe dep1 besteht mit hoher Sicherheit vorwiegend aus Kindern aus Kindern mit keiner

oder mit geringer Belastung durch depressive Symptome. Das Forschungsinteresse wandte sich international der Kinder mit fehlendem oder geringerem Risiko für die Ausbildung depressiver Symptome kaum zu. Das könnte als Vernachlässigung gegebener Möglichkeiten für das Auffinden von Einsichten in die Bedingungen einer relativen Stabilität gegenüber depressiven Symptomen gewertet werden. Wenn das Forschungsinteresse nur auf eine der beiden Randgruppen gerichtet ist, kann dem durch Zusammenlegung der beiden anderen Gruppen gefolgt werden. Das geschieht, wie hier schon zuvor bei der Frage nach den altersspezifischen Variationen der CDI-GPW nun auch nachfolgend bei der Analyse der Variation dieser Punktwerte mit Alter und Geschlecht.

Die nachfolgende Tabelle führt die Zahlen und Prozentsätze der Kinder an, die der Risikogruppe dep3 zugewiesen wurden. Mit diesen Daten wird verdeutlicht, für wie viele Kinder jeweils ein hohes Risiko für das Bestehen einer depressiven Störung erkannt wurde. Die Formulierung, das Risiko sei erkannt worden, ist korrekt. Der Begriff des Risikos schließt die Möglichkeit einer relativen Verknennung mit ein.

Tabelle 10.5: Geschlechtsspezifische Verteilung der Kinder auf die eingeführten Risikogruppen in unterschiedlichen Altersstufen

Alter	Geschlecht		Dep1+2	Dep3	Gesamt
8	Jungen	Anzahl	88 **	5 **	93
		Erwartete Anzahl	79,3	13,7	93,0
		Prozent	94,6%	5,4%	100,0%
	Mädchen	Anzahl	69	9	78
		Erwartete Anzahl	66,5	11,5	78,0
		Prozent	88,5%	11,5%	100,0%
9	Jungen	Anzahl	88	17	105
		Erwartete Anzahl	89,6	15,4	105,0
		Prozent	83,8%	16,2%	100,0%
	Mädchen	Anzahl	104 **	9 **	113
		Erwartete Anzahl	96,4	16,6	113,0
		Prozent	92,0%	8,0%	100,0%
10	Jungen	Anzahl	127	16	143
		Erwartete Anzahl	122,0	21,0	143,0
		Prozent	88,8%	11,2%	100,0%
	Mädchen	Anzahl	102	13	115
		Erwartete Anzahl	98,1	16,9	115,0
		Prozent	88,7%	11,3%	100,0%
11	Jungen	Anzahl	129	16	145
		Erwartete Anzahl	123,7	21,3	145,0
		Prozent	89,0%	11,0%	100,0%
	Mädchen	Anzahl	121	17	138
		Erwartete Anzahl	117,7	20,3	138,0
		Prozent	87,7%	12,3%	100,0%
12	Jungen	Anzahl	139	20	159
		Erwartete Anzahl	135,6	23,4	159,0
		Prozent	87,4%	12,6%	100,0%
	Mädchen	Anzahl	138	32	170
		Erwartete Anzahl	145,0	25,0	170,0
		Prozent	81,2%	18,8%	100,0%

(Fortführung Tabelle 10.5)

Alter	Geschlecht		Dep1+2	Dep3	Gesamt
13	Jungen	Anzahl	146	24	170
		Erwartete Anzahl	145,0	25,0	170,0
		Prozent	85,9%	14,1%	100,0%
	Mädchen	Anzahl	135	32	167
		Erwartete Anzahl	142,4	24,6	167,0
		Prozent	80,8%	19,2%	100,0%
14	Jungen	Anzahl	145	31	176
		Erwartete Anzahl	150,1	25,9	176,0
		Prozent	82,4%	17,6%	100,0%
	Mädchen	Anzahl	114 **	32 **	146
		Erwartete Anzahl	124,5	21,5	146,0
		Prozent	78,1%	21,9%	100,0%
15 +	Jungen	Anzahl	37 **	13 **	50
		Erwartete Anzahl	42,6	7,4	50,0
		Prozent	74,0%	26,0%	100,0%
	Mädchen	Anzahl	35	10	45
		Erwartete Anzahl	38,4	6,6	45,0
		Prozent	77,8%	22,2%	100,0%
Gesamt	Anzahl	1717,0	296,0	2013,0	
	Prozent	85,3%	14,7%	100,0%	

Die in Tabelle 10.5 dargestellte nach Alter geschlechtsspezifische Verteilung der Kinder auf die CDI-Risikogruppen Dep1+2 und Dep3 weicht hochsignifikant von den Erwartungswerten ab [$\chi^2(15) = 35,929$; $p = 0,002$].

10.22 AS: Die Verteilung der Schulkinder auf zwei nach CDI-GPW (Cutoff = 19) gebildeten Risikogruppen variiert signifikant mit geschlechtsspezifischen Altersgruppen.

10.23 AS: In der Altersgruppe der 8jährigen Kinder treten bei den Jungen signifikante und bei den Mädchen nur zufällige Abweichungen im Hinblick auf ihre Verteilung auf die zwei CDI-Risikogruppen auf. Die Jungen sind gegenüber den über Randsummen bestimmten Erwartungswerten in der Gruppe niedrigerer Belastung Dep1+2 mit +10,97% überrepräsentiert und in der Gruppe Dep2 mit -63,50% unterrepräsentiert. Die Jungen der Stichprobe bleiben auf dieser Altersstufe hochsignifikant unter den über Randsummen bestimmten Erwartungswerten, während die Mädchen denen entsprechen. Die Gruppe der 8jährigen Jungen schöpft das über die Erwartungswerte bestimmte Potential der Entwicklung depressiver Belastungen nicht aus.

10.24. AS: In der Altersgruppe der 9jährigen Kinder treten bei den Jungen zufällige und bei den Mädchen signifikante Abweichungen im Hinblick auf ihre Verteilung auf die zwei CDI-Risikogruppen auf. Die Mädchen dieses Alters sind gegenüber den über Randsummen bestimmten Erwartungswerten in der Gruppe Dep1+2 mit +7,88% überrepräsentiert und in der Gruppe Dep3 mit -45,78% unterrepräsentiert. Die Jungen dieser Altersstufe entsprechen den Erwartungswerten, während die Mädchen unter den Erwartungswerten bleiben. Die Gruppe der 9jährigen Mädchen zeigt geringere Frequenzen an depressiver Belastung als statistisch erwartet.

10.25 AS: Die 10-, 11-, 12- und 13jährigen Jungen und Mädchen der Stichprobe zeigen mit ihren Frequenzen in den Gruppen Dep1+2 und Dep3 nur zufällige Abweichungen.

10.26 AS: In der Altersgruppe der 14jährigen Kinder treten bei den Jungen zufällige und bei den Mädchen signifikante Abweichungen im Hinblick auf ihre Verteilung auf die zwei gebildeten Belastungsgruppen auf. Die Mädchen dieses Alters sind gegenüber den über Randsummen bestimmten Erwartungswerten in der Gruppe niedrigerer depressiver Belastung Dep1+2 -8,43% unterrepräsentiert und in der Gruppe hoher Belastung Dep3 mit +48,84% überrepräsentiert.

10.27 AS: In der Altersgruppe der 15jährigen Kinder treten bei den Jungen signifikante und bei den Mädchen zufällige Abweichungen im Hinblick auf ihre Verteilung auf die zwei gebildeten Belastungsgruppen auf. Die Jungen dieses Alters sind gegenüber den über Randsummen bestimmten Erwartungswerten in der Gruppe niedriger depressiver Belastung Dep1+2 -13,15% unterrepräsentiert und in der Gruppe hoher Belastung Dep3 mit +48,84% überrepräsentiert. Der Prozentsatz der Mädchen dieses Alters, die aufgrund ihrer CDI-Gesamtpunktwerte in der Gruppe hoher depressiver Belastung Dep3 zugeordnet sind, obschon viel höher als der Gesamtstichprobe (22,2% zu 14,7% entsprechend) weicht statistisch nicht von den Erwartungswerten ab.

Die nachstehende Tabelle gibt einen Überblick über die Richtung der von den Erwartungswerten signifikant abweichenden Zahlen der Verteilung von Jungen und Mädchen auf die unterschiedenen Risikogruppen an. Die Tabelle vermerkt Unterschreitungen mit Minus, Gleichstand mit Gleichheitszeichen und Überschreiten mit Plus. In der Spalte „DTK-Ergebnisse“ sind die Befunde aus den entsprechenden Analysen von DTK-Daten (siehe Kap. 8.5 und dort Tab. 8.6 und 8.7) dargestellt.

Tabelle 10.6: Richtung der bei Jungen und Mädchen auf den Altersstufen von statistischen Erwartungen abweichenden Zuweisungen mutmaßlich depressiv belasteter Kinder nach DTK (Gruppe DepBel3) und nach CDI (Gruppe Dep3)

Alter	Richtung der Abweichung			
	DTK-Ergebnisse		CDI-Ergebnisse	
	Jungen	Mädchen	Jungen	Mädchen
8	Minus	=	Minus	=
9	=	Minus	=	Minus
10	=	=	=	=
11	=	Minus	=	=
12	=	=	=	=
13	=	Plus	=	=
14	=	Plus	=	Plus
15+	Plus	Plus	Plus	=

Zeichenerklärung: Minus= gegenüber Erwartungswerten geringeres Aufkommen an mutmaßlich in klinischer Ausprägung depressiv belastete Jungen oder Mädchen, Plus= ein gegenüber Erwartungswerten höheres Aufkommen; mit = wird ein den Erwartungswerten folgendes Aufkommen angezeigt.

Wie es aus der oben dargestellten zusammenfassenden Tabelle zu ersehen ist, führen die Analysen über die nach Altersgruppen auftretenden geschlechtsspezifischen Unterschiede im Hinblick auf die Zuordnung der Kinder zu den Gruppen hoher depressiver Belastung des CDI und des DTK zu relativ übereinstimmenden Ergebnissen. Im DTK treten in 7 von 16 speziellen Teilgruppen der Kinder (gebildet nach Alter und Geschlecht) signifikante Variationen auf. Von 7 signifikanten Ergebnissen im DTK treten auch 4 im CDI auf. Die Frage, warum die übrigen 3 Teilgruppen von Kindern, die im DTK signifikante Variationen in Vergleich zu der Gesamtheit aufweisen (11-, 13- und 15-jährigen Mädchen), nicht auch im CDI als signifikant unterrepräsentiert oder überrepräsentiert in der CDI Risikogruppe Dep3 erscheinen, kann mit unterschiedlichen Hypothesen zu beantworten versucht werden. Eine von denen wäre, dass das DTK empfindlicher zu entwicklungspezifischen Besonderheiten bei dem selbstberichteten Auftreten depressiver Störungen sei. Diese Hypothese könnte nur unter der Annahme formuliert werden, dass die zwei Verfahren (CDI und DTK) eine und dieselbe „depressive Störung“ messen. Letztere Annahme scheint aber eher unwahrscheinlich zu sein, unter anderem aufgrund der Tatsache, dass die zwei

Verfahren aus Items unterschiedlicher Zahl und Formulierung bestehen. Das Gewicht der über die DTK Skala „Agitiertes Verhalten“ aufgenommenen Informationen dürfte im Übrigen größer sein als im CDIU. Dafür sprechen einige Hinweise, so auch die Validitätsstudie von Zimmermann (in Rossmann, 1993) nach der mit dem DTK auch Hinweise auf ADHS erfasst werden können.

Eine verfahrenüberschreitende Betrachtung der Zahl und der Richtung aller signifikanten Abweichungen von den nach Randsummen berechneten Anteile von Kindern einer Altersgruppe an den höchsten Risikogruppen kommt zu dem Resultat, das die nachstehende Aussage wiedergibt.

10.28 AS: Wenn die untersuchten Altersgruppen in zwei Altersbereiche zusammengefasst werden, in den Bereich A der 8-, 9-, 10- und 11 jährigen Jungen einerseits und den Bereich B der 12-, 13-, 14- und 15-jährigen Jungen und Mädchen, treten im Bereich A insgesamt 5 signifikante Abweichungen von Erwartungswerten der Zugehörigkeit zur höchsten Risikogruppe und im Bereich B 6 auf.

10.29 AS: Die Zahl der im Bereich A und die Zahl der im Bereich auftreten Abweichungen von der erwarteten Gruppengröße für die höchste Risikogruppe ist statistisch gleich.

10.30 AS: Alle nach der vorangegangenen Aussage und nach beiden Messverfahren auftretenden signifikanten Abweichungen im Altersbereich A sind Unterschreitungen der Erwartungswerte.

10.31 AS: Alle nach der vorangegangenen Aussage und nach beiden Messverfahren auftretenden signifikanten Abweichungen im Altersbereich B sind Überschreitungen der Erwartungswerte.

10.32 AS: Nach den mit CDI und DTK vorgenommenen Messungen tritt für die unteren Altersgruppen 8, 9, 10 und 11 Jahre eine gleiche Anzahl an Abweichungen von erarteten Frequenzen für die höchsten Risikogruppen des Bestehens depressiver Störungen auf. Unterschiedlich ist die Richtung der Abweichung. Alle Abweichungen im unteren Altersbereich sind Unterschreitungen der Erwartungswerte. Alle Abweichungen im höheren Altersbereich sind Überschreitungen der Erwartungswerte.

10.33 AS: Die festgestellte Tendenz der altersspezifischen Abweichungen von den erwarteten Frequenzen an Kindern in der höchsten Risikogruppe spricht für eine systematisch auftretende Tendenz in allen Messungen mit CDI und DTK in den angegebenen Altersgruppen. Die Berechnung der Erwartungswerte über die Randsummen der über 8 Altersstufen reichenden Tabelle ergibt eine lineare Beziehung zwischen Größe der Altersgruppe und erwarteter Frequenz an Kinder der höchsten Belastungsgruppe. Dieser linearen Beziehung unterliegt eine empirische Auftretenskurve, die, von Stichprobenschwankungen abgesehen, in den unteren Altersgruppen unter den linearen Erwartungswerten bleibt und in den oberen Altersgruppen die lineare Beziehung anwachsend übersteigt.

10.34 AS: Die vorangehende Aussage kann in die nachstehende Feststellung umformuliert werden. Das mit Alter und Geschlecht über die Randsummen errechnete und linear über die Altersstufen gelegte Gitternetzen von Erwartungswerten für die Frequenzen der höchsten depressionsspezifischen Risikogruppe ist gegenüber dem empirisch nachzuweisenden Kurvenverlauf fehlerhaft. Es liegt in den unteren vier Altersstufen über und in den oberen vier Altersstufen unter dem empirisch beobachteten Kurvenverlauf. Die nachstehende Graphik veranschaulicht die Aussage.

Tabelle 10.7: Nach Schulstufen unterschiedene geschlechtsspezifische Verteilungen der CDI-GPW auf zwei CDI-Risikogruppen (dep1+2 und dep3)

Schulstufe	Geschlecht		Dep1+2	Dep3	Gesamt
Grundschule	Jungen	Anzahl	461 **	56 **	517
		Erwartete Anzahl	441,0	76,0	517,0
		Prozent	89,2%	10,8%	100,0%
	Mädchen	Anzahl	418 **	51 **	469
		Erwartete Anzahl	400,0	69,0	469,0
		Prozent	89,1%	10,9%	100,0%
Gymnasium	Jungen	Anzahl	438	86	524
		Erwartete Anzahl	446,9	77,1	524,0
		Prozent	83,6%	16,4%	100,0%
	Mädchen	Anzahl	400 **	103 **	503
		Erwartete Anzahl	429,0	74,0	503,0
		Prozent	79,5%	20,5%	100,0%
	Gesamt	Anzahl	1717	296	2013
		Prozent	85,3%	14,7%	100%

Die in dieser Tabelle wiedergegebene Verteilung von Jungen und Mädchen auf die unterschiedenen Risikogruppen dep1+2 und dep3 weicht hochsignifikant von Zufallswerten ab. [$\chi^2(3) = 26,251$; $p = 0,000$]. Daraufhin dürfen die folgenden Aussagen formuliert werden:

10.35 AS: Die Verteilung der Kinder auf die zwei CDI-Risikogruppen variiert signifikant mit geschlechtsspezifischen Schulstichproben.

10.36 AS: Sowohl die jüngeren Jungen als auch die jüngeren Mädchen der Stichprobe (beide aus Grundschulen) sind in der CDI Risikogruppe 1+2 überrepräsentiert und in der Gruppe dep3 unterrepräsentiert.

10.37 AS: Die Überrepräsentation der Jungen und Mädchen aus Grundschulen in der Gruppe Dep1+2 beträgt gegenüber den über die jeweiligen Randsummen bestimmten Erwartungswerten +4,54% bei den Jungen und +4,5% bei den Mädchen.

10.38 AS: Die Unterrepräsentation der Jungen und Mädchen aus Grundschulen in der Gruppe Dep3 beträgt gegenüber den über die jeweiligen Randsummen bestimmten Erwartungswerten -26,32% bei den Jungen und -26,09% bei den Mädchen.

10.39 AS: Die in dieser Stichprobe befindlichen älteren Jungen (aus den Gymnasien) zeigen mit ihren Frequenzen in den zwei eingeführten CDI-Risikogruppen nur zufällige Abweichungen.

10.40 AS: Die in dieser Stichprobe befindlichen älteren Mädchen (aus Gymnasien) sind in der Gruppe niedriger Belastung Dep1+2 unterrepräsentiert und in der Gruppe hoher Belastung Dep3 überrepräsentiert.

10.41 AS: Die Unterrepräsentation der Mädchen aus Gymnasien in der Gruppe Dep1+2 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten -6,76%.

10.42 AS: Die Überrepräsentation der Mädchen aus Gymnasien in der Gruppe Dep3 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten +39,19%. Die Überrepräsentation der älteren Mädchen in der CDI Gruppe hoher Belastung von depressiven Symptomen liegt also fast 40% über den Erwartungswerten.

10.43 AS: Die letztgenannte Aussage stimmt in hohem Maße mit den entsprechenden Feststellungen aus den Analysen der DTK Daten überein. Dort wurde gefunden, dass die gymnasialen Mädchen fast 50% über den für die höchste Risikogruppe DepBel3 berechneten Erwartungswerten.

10.44 AS: Die über die Zusammenfassung der beteiligten Jungen und Mädchen sowohl für Grundschulen als auch für Gymnasien gewählte Stichprobenorganisation bestätigt die Grundtendenz der bereits mehrfach wiederholten Feststellung. Die über die Randsummen der Gesamttabelle berechneten Erwartungswerte, werden von Jungen und Mädchen in der Grundschule im Sinne eines allgemein niedrigeren Aufkommens von Kindern mit vielen depressiven Symptomen unterboten und von den Mädchen an Gymnasien im Sinne eines höheren Aufkommens an Mädchen mit vielen depressiven Symptomen überboten.

10.6 Das Wiederholen einer Klasse als potentieller Risikofaktor einer erhöhten depressionspezifischen Vulnerabilität

An mehreren vorangegangenen Stellen dieser Arbeit ist darauf hingewiesen worden, dass ungeplant aber auch realitätsnah in der Untersuchung eine Teilstichprobe von Schulkindern entstanden ist, die ein besonderes Merkmal im Vergleich zu ihren Mitschülern aufweist, nämlich das der Überschreitung ihres jeweiligen klassentypischen Alters. Diese klassenspezifische Überalterung geht in der Regel auf die Wiederholung einer Klassenstufe zurück. Im entsprechenden Teil der Analysen von DTK-Daten wurde in diesem Zusammenhang die Hypothese geäußert, dass das Wiederholen einer Klasse aufgrund der damit eventuell verbundenen Hindergründe als einer der möglichen Risikofaktor zu einer erhöhten Vulnerabilität für die Entwicklung depressiver Störungen verstanden werden könnte. Die durchgeführten Analysen mit den DTK-Daten haben diese Hypothese bestätigt (zu Beschreibung der Teilstichprobe und zu den zuvor erwähnten Ergebnissen siehe Kapitel 8.6).

An dieser Stelle wird nun geprüft, ob auch aus den Analysen der mit dem CDI gewonnen Daten das Wiederholen einer Klasse mit einer signifikanten Überrepräsentation dieser Kindergruppe in der Risikogruppe hoher Belastung verbunden ist und demzufolge als Risikofaktor zur Entwicklung depressiver Störungen verstanden werden sollte.

Die Umschreibung des Sachverhaltes als Risikofaktor sollte erläutert werden. Die Wiederholung einer Klassenstufe kann für das betroffene Kind zum einen die Ursache für das vermehrte Auftreten depressiver Symptome in dieser Persönlichkeit gewesen sein. Aber die entgegengesetzte kausale Beziehung wäre ebenso denkbar, nach der bereits bestehenden depressiven Symptome die Ursache sowohl für schulische Leistungsdefizite als auch für nachfolgendes Wiederholen einer Klassenstufe gewesen wären. Schließlich könnten sich in jeder Schule einige Kinder befinden, die von der einen kausalen Beziehung betroffen wurden und andere, die unter dem Auftreten der anderen gelitten hatten. Kausale Aussagen lassen sich mit den gewonnenen Feststellungen nicht begründen. Für die nicht aufschließbaren Zusammenhänge steht der Begriff von der Wiederholung einer Klasse als Risikofaktor.

Da die Mehrheit der überalterten Kinder (87,80%) eine Klasse des Gymnasiums besucht, wird - ähnlich wie bei der Analyse der DTK-Daten - die durchzuführende χ^2 -Prüfung aus statistischen Gründen einschränkend auf die Stichprobe der Gymnasiasten ($n=1027$) bezogen.

Tabelle 10.8: Verteilung der in gymnasialen Schulklassen altersgerecht oder nicht altersgerecht befindlichen Schülerinnen und Schüler auf definierte depressionsspezifische Risikogruppen dep1+2 und dep3 (nach CDI-GPW)

Klassenbesuch		Dep1+2	Dep3	Gesamt
dem Alter entsprechend	Anzahl	788 **	169 **	957
	Erwartete Anzahl	780,9	176,1	957,0
	Prozent	82,3%	17,7%	100,0%
überalterte Kinder	Anzahl	50 **	20 **	70
	Erwartete Anzahl	57,1	12,9	70,0
	Prozent	71,4%	28,6%	100,0%
Gesamt	Anzahl	838	189	1027
	Prozent	81,6%	18,4%	100,0%

Zeichenerklärung: Dep1+2= Gruppe niedriger oder mittlerer depressiver Belastung nach CDI (bis zu 18 GPW), Dep3 = Gruppe hoher depressiver Belastung nach CDI (ab 19 GPW), **= Felder, die nach korrigierten standardisierten Residuen von den erwarteten Werten signifikant abweichen

Die in Tabelle 10.8 dargestellte Verteilung der Gymnasiasten auf die zwei CDI-Risikogruppen Dep1+2 und Dep3 ist statistisch signifikant [$\chi^2(1) = 5,172$; $p = 0,023$].

Daraus ließen sich die folgende Aussagen ableiten:

10.45 AS: Die Verteilung der Gymnasiasten auf die zwei gebildeten Risikogruppen Dep1+2 und Dep3 variieren signifikant mit dem altersgerechten oder dem davon abweichenden Schulbesuch.

10.46 AS: Die Schüler, die eine ihrem Alter entsprechende Schulklasse besuchen, sind in der Gruppe niedriger Belastung Dep1+2 überrepräsentiert und in der Gruppe hoher Belastung Dep3 unterrepräsentiert.

10.47 AS: Die Überrepräsentation dieser Schüler in der Gruppe Dep1+2 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten +0,91%.

10.48 AS: Die Unterrepräsentation dieser Schüler in der Gruppe Dep3 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten -4,03%.

10.49 AS: Die Schüler, die im Vergleich zu ihren Mitschülern in der Klasse überaltert sind, sind in der Gruppe dep1+2 unterrepräsentiert und in der Gruppe Dep3 überrepräsentiert.

10.50 AS: Die Unterrepräsentation der überalterten Kinder in der Gruppe Dep1+2 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten -12,43%.

10.51 AS: Die Überrepräsentation der überalterten Kinder in der Gruppe Dep3 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten +55,04%.

10.52 AS: Das in 7. Aussage formuliertes Ergebnis stimmt in hohem Maße mit dem entsprechenden aus der Analyse der DTK-Daten. Dort wurde gefunden, dass die Überrepräsentation der überalterten Kinder in der DTK-Gruppe hoher Belastung um 33,05% vermehrt war.

In Übereinstimmung mit der Vorgehensweise im DTK wird auch hier als ein zweiter Schritt die Überprüfung von geschlechtsspezifischen Unterschieden im Hinblick auf die Verteilung der überalterten Gymnasiasten in den zwei CDI-Risikogruppen Dep1+2 und Dep3 vorgenommen.

Tabelle 10.9: Verteilung der überalterten Mädchen und Jungen aus Gymnasien auf die nach CDI Gesamtpunktwerten gebildeten depressionsspezifischen Risikogruppen

		Dep1+2	Dep3	Gesamt
überalterte Jungen	Anzahl	25	12	37
	Erwartete Anzahl	26,4	10,6	37,0
	Prozent	67,6%	32,4%	100,0%
überalterte Mädchen	Anzahl	25	8	33
	Erwartete Anzahl	23,6	9,4	33,0
	Prozent	75,8%	24,2%	100,0%
Gesamt	Anzahl	50	20	70
	Prozent	71,4%	28,6%	100,0%

Zeichenerklärung: Dep1= Gruppe niedriger oder mittlerer depressiver Belastung nach CDI (bis zu 18 GPW), Dep2= Gruppe hoher depressiver Belastung nach CDI (ab 19 GPW), **= Felder, die nach korrigierten standardisierten Residuen von den erwarteten Werten signifikant abweichen

Die in Tabelle 10.9 dargestellte Verteilung ist statistisch nicht signifikant [$\chi^2(1) = ,573$; $p = 0,449$].

Auf diesem Ergebnis basiert dann folgende Aussage:

10.53 AS: Die Verteilung der überalterten Mädchen und Jungen auf die CDI-Gruppen niedriger und hoher depressiver Belastung (Dep1+2 und Dep3) ist in Übereinstimmung mit dem entsprechenden Befund aus den DTK-Daten geschlechtsspezifisch nicht signifikant unterschiedlich.

Das zuvor mitgeteilte Ergebnis verstärkt einen durch die entsprechenden DTK-Analysen gefundenen Hinweis auf eine interessante Beziehung: auch mit dem Verfahren CDI ergibt sich, dass die für die Gesamtstichprobe gefundenen geschlechtsspezifischen Unterschiede im Hinblick auf die Verteilung der Kinder in den Gruppen Dep1+2 und Dep3 zu Lasten der Mädchen aus Gymnasien für die speziellen Teilgruppe der überalterten Kinder nicht gilt. Die daraus abzuleitende Nachfrage, warum dieser Unterschied auftritt, kann hier nun nachfolgend zwar gestellt, aber nicht beantwortet werden.

10.7 Der Wohnort des Kinder als möglicher Faktor für Variationen von CDI-Gesamtwerten

Die an Persönlichkeiten interessierte empirische Forschung sammelt zum einen, ihrem Forschungsinteresse folgend, persönliche Daten. Wenn bestimmt werden sollte, welchem Raum diese Daten zu entnehmen wären, würde der primär über die Persönlichkeit selbst beschrieben werden können. Jeder Mensch, jede Persönlichkeit lebt aber in unauflösbarer Interaktion mit seiner Umwelt. Die Atemluft durchdringt ihn, und ist die Luft seiner Wohnumgebung verschmutzt, wird die mit hoher Wahrscheinlichkeit auch seinen Organismus verschmutzen, oder verstrahlen, oder ganz im Gegenteil ihn biologisch und physiologisch förderlich werden. Mit seiner Lebensumwelt interagiert jeder Mensch über die ihm dazu nutzbaren Sinne. Seine psychischen Inhalte werden außer durch die intrapersonalen Lebensbedingungen auch durch die spezifische und persönlichkeitspezifische Interaktion mit der Umwelt determiniert. Der Wohnort eines Schulkindes ist einerseits eine sprachlich symbolisierte Repräsentation einer geographisch

bestimmten Siedlungsfläche. Andererseits ist im Zusammenwirken mit dem Namen des jeweiligen Kindes ein Symbol für die persönlichkeitspezifischen Lebensumstände eines Kindes an seinem Wohnort. Eine empirisch überaus schwierig zu bewältigende Aufgabe ist die Entschlüsselung der durch den Namen eines Wohnortes symbolisierten persönlichkeitspezifischen Lebensumstände. Sie stünden, bezogen auf das Auftreten depressiver Symptome, wie alle Lebensumstände einer Persönlichkeit, im Verdacht, mit denen in Verbindung zu stehen. Der Versuch, den Wohnort der Schulkinder in der untersuchten Stichprobe 2034GR, in die Datenanalyse einzubeziehen, folgt diesem allgemein interaktionalen Ansatz, der hier ausdrücklich auf den von Magnusson und Edler (1977) bezogen wird. David Magnusson ging von der These aus, dass Persönlichkeit immer als Interaktion von Person und Umwelt zu verstehen sei. Damit nimmt er die von Kurt Lewin vertretene Position auf. In der hier geführten Untersuchung, die dem Auftreten depressiver Symptome bei griechischen Schulkindern folgt, ist in der genannten Perspektive der Wohnort der untersuchten Schulkinder eine spezielle Symbolisierung der allgemeinen Interaktion von Person und Umwelt. Die Analyse der aufgenommenen Daten folgt dem, und versucht, depressionsspezifische Informationen, die mit dem Wohnort verbunden sind, aufzuschließen.

Im Folgenden wird dementsprechend der Frage nachgegangen, ob die von den beteiligten Kindern aufgenommenen CDI-Werte in signifikanter Weise mit den unterschiedlichen Wohnorten variieren. Die nachstehende Tabelle stellt für die Gesamtstichprobe die auf Wohnorte bezogene Verteilung der CDI-Gesamtpunkte dar, wobei die über die Zugehörigkeit zu den nach solchen Werten gebildeten Risikogruppen dep1+2 und dep3 vertreten werden. Die CDI-Gesamtpunktwerte werden somit in zwei Datenklassen dargestellt, der von 0 bis 18, jeweils inklusive, als Risikogruppe dep1+2 bezeichnet und der von 19 bis zum Maximalwert, die Risikogruppe dep3 genannt wird. Diese beiden Gruppen können auch als unter Zuhilfenahme des Cutoff-Wertes von GPW = 19 gebildete Risikogruppen niedriger und hoher Belastung mit depressiven Symptomen bezeichnet werden.

Um alters- und geschlechtsspezifische Einflüsse zu berücksichtigen, werden dann nachfolgend die nach Wohnorten gruppierte geschlechtsspezifische Verteilung der Grundschul Kinder und die der Gymnasiasten in ihrer Verteilung auf die genannten zwei CDI-Risikogruppen dargestellt (Tabellen 10.11 und 10.12).

Wie schwierig der Umgang mit der Einbeziehung des Wohnortes wird, demonstriert in der nachstehenden Tabellenüberschrift die ersatzweise für Wohnort eingesetzte Bezeichnung Schulort. Von jedem Kind wird gewusst, dass es die Schule an dem Ort besucht, an dem die Untersuchung durchgeführt worden ist, eben am Schulort. Nicht auszuschließen ist aber, dass ein oder mehrere Kinder an dem jeweiligen Untersuchungsort als Fahrschüler zu führen wären. Die würden außerhalb des Ortes wohnen und zum Schulbesuch in den jeweiligen Schulort gebracht werden.

Tabelle 10.10: Getrennt nach Schulort (Untersuchungsort) unterschiedene Verteilung der Gesamtstichprobe auf die CDI-Risikogruppen Dep1+2 und Dep3

		Dep1+2	Dep3	Gesamt
Veria	Anzahl	170	22	192
	Erwartete Anzahl	163,8	28,2	192,0
	Prozent	88,5%	11,5%	100,0%
Thessaloniki	Anzahl	218	39	257
	Erwartete Anzahl	219,2	37,8	257,0
	Prozent	84,8%	15,2%	100,0%
Galatista	Anzahl	114	18	132
	Erwartete Anzahl	112,6	19,4	132,0
	Prozent	86,4%	13,6%	100,0%
N. Agioneri Kilkis	Anzahl	78	15	93
	Erwartete Anzahl	79,3	13,7	93,0
	Prozent	83,9%	16,1%	100,0%
Athen	Anzahl	356	49	405
	Erwartete Anzahl	345,4	59,6	405,0
	Prozent	87,9%	12,1%	100,0%
Ksanthi	Anzahl	159	24	183
	Erwartete Anzahl	156,1	26,9	183,0
	Prozent	86,9%	13,1%	100,0%
Stavroupoli	Anzahl	64 **	22 **	86
	Erwartete Anzahl	73,4	12,6	86,0
	Prozent	74,4%	25,6%	100,0%
Rethymnon	Anzahl	115	22	137
	Erwartete Anzahl	116,9	20,1	137,0
	Prozent	83,9%	16,1%	100,0%
Episkopi	Anzahl	126	23	149
	Erwartete Anzahl	127,1	21,9	149,0
	Prozent	84,6%	15,4%	100,0%
Heraklion	Anzahl	105 **	32 **	137
	Erwartete Anzahl	116,9	20,1	137,0
	Prozent	76,6%	23,4%	100,0%
Karditsa	Anzahl	116	14	130
	Erwartete Anzahl	110,9	19,1	130,0
	Prozent	89,2%	10,8%	100,0%
Itea	Anzahl	96	16	112
	Erwartete Anzahl	95,5	16,5	112,0
	Prozent	85,7%	14,3%	100,0%
Gesamt	Anzahl	1717	296	2013
	Prozent	85,3%	14,7%	100,0%

Zeichenerklärung: dep1+2 = Gruppierung der Schulkinder mit einem CDI Gesamtpunktwert zwischen 0 und 18 einschließlich; dep3 = Gruppierung der Schulkinder mit einem CDI Gesamtpunktwert zwischen 19 und Maximalwert, jeweils einschließlich

Die oben stehende Tabelle erweist sich als statistisch signifikant [$\chi^2(11) = 22,666$; $p = 0,020$]. Dieses Ergebnis führt dann zu den nachfolgenden Aussagen. Für die ist ausdrücklich hervorzuheben, dass sie die gesamten Schulkinder der untersuchten Stichprobe betreffen und dies,

nachdem mehrfach altersspezifische und geschlechtsspezifische Unterschiede aus der Untersuchung 2034GR bekannt geworden sind. Die Aussagen stehen damit gewissermaßen unter Verdacht, sie könnten aufgetretene altersspezifische oder geschlechtsspezifische Unterschiede durch die Zusammenfassung aller Altersgruppen und beider Geschlechtsgruppen aufgehoben haben. Festzustellen bleibt aber dennoch:

10.54 AS: Die Verteilung der untersuchten Schulkinder auf die zwei gebildeten CDI-Gruppen unterschiedlichen Grades von depressiver Belastung variiert signifikant mit dem Wohnort der Kinder.

10.55 AS: In 10 von insgesamt 12 besuchten Orten treten dennoch keine signifikante Abweichungen im Hinblick auf die Verteilung der dort lebenden Kinder auf die CDI-Gruppen Dep1+2 und Dep3 auf. Diese Orte sind, wie in DTK, Veria, Thessaloniki, Galatista, Athen, Ksanthi, Rethymno, Episkopi, Itea und auch N. Agioneri Kilkis und Karditsa.

10.56 AS: Im Gegensatz dazu werden in 2 von allen 12 Untersuchungsorten signifikante Abweichungen im Hinblick auf die Zugehörigkeit der dort lebenden Kinder zu einer von beiden Belastungsgruppen bekannt. Diese Orte sind Stavroupoli und Heraklion.

10.57 AS: In den oben genannten zwei Orten ist die Wahrscheinlichkeit der dort lebenden Kinder, der Gruppe hoher Belastung zugewiesen zu werden, höher als in den übrigen Schulorten der Gesamtstichprobe. Diese Wahrscheinlichkeit beträgt um 74,6% in Stavroupoli und 59,20% in Heraklion.

Mit den genannten Ergebnissen ist sicher, dass in den genannten zwei Orten alle für das Zustandekommen der Verteilung von CDI Gesamtpunktwerten verantwortlichen Bedingungen so sind, dass sich schließlich das Endresultat ergibt. Diese Besonderheit an den genannten zwei Orten könnten alle, bekannten oder unbekannt, Determinanten des Gesamturteils einzeln oder in Kombination bewirkt haben. Da aus den in dieser Untersuchung bereits bekannt gewordenen Untersuchungen ebenso wie aus der internationalen Fachliteratur hinreichend bekannt ist, dass geschlechtsspezifische und altersspezifische Unterschiede in den CDI Werten auftreten, kämen die als erste in den Verdacht, die Besonderheiten der Gesamtergebnisse in den beiden Orten bewirkt zu haben. Hinreichend für das Zustandekommen der signifikanten Abweichungen wäre beispielsweise die zahlenmäßige Dominanz älterer Mädchen an diesen beiden Schulorten. Die würden das Auftreten signifikanter Überbesetzungen der Risikogruppe dep3 erwarten lassen. Nachstehend wird geprüft, ob solche geschlechtsspezifischen Effekte im Zusammenwirken mit altersspezifischen Effekten an den einzelnen Orten vorliegen. Die Altersvariable wird in der nachstehenden Tabelle durch den Zusammenschluss von Kindern der Grundschule repräsentiert. Die Tabelle zeigt also Kinder im Grundschulalter.

Tabelle 10.11: Geschlechtsspezifische Verteilung der jüngeren Kinder der Stichprobe (Grundschulkinder) auf die Risikogruppen Dep1+2 und Dep3

Stadt	Geschlecht		Dep1+2	Dep3	Gesamt
Veria	Jungen	Anzahl	55	7	62
		Erwartete Anzahl	55,3	6,7	62,0
		Prozent	88,7%	11,3%	100,0%
	Mädchen	Anzahl	46	7	53
		Erwartete Anzahl	47,2	5,8	53,0
		Prozent	86,8%	13,2%	100,0%

(Fortführung Tabelle 10.11)

Stadt	Geschlecht		Dep1+2	Dep3	Gesamt
Thessaloniki	Jungen	Anzahl	60	8	68
		Erwartete Anzahl	60,6	7,4	68,0
		Prozent	88,2%	11,8%	100,0%
	Mädchen	Anzahl	63	5	68
		Erwartete Anzahl	60,6	7,4	68,0
		Prozent	92,6%	7,4%	100,0%
Galatista	Jungen	Anzahl	28	5	33
		Erwartete Anzahl	29,4	3,6	33,0
		Prozent	84,8%	15,2%	100,0%
Galatista	Mädchen	Anzahl	25	4	29
		Erwartete Anzahl	25,9	3,1	29,0
		Prozent	86,2%	13,8%	100,0%
N. Agioneri Kilkis	Jungen	Anzahl	19	4	23
		Erwartete Anzahl	20,5	2,5	23,0
		Prozent	82,6%	17,4%	100,0%
	Mädchen	Anzahl	7	3	10
		Erwartete Anzahl	8,9	1,1	10,0
		Prozent	70,0%	30,0%	100,0%
Athen	Jungen	Anzahl	108	9	117
		Erwartete Anzahl	104,3	12,7	117,0
		Prozent	92,3%	7,7%	100,0%
	Mädchen	Anzahl	99	11	110
		Erwartete Anzahl	98,1	11,9	110,0
		Prozent	90,0%	10,0%	100,0%
Ksanthi	Jungen	Anzahl	43	6	49
		Erwartete Anzahl	43,7	5,3	49,0
		Prozent	87,8%	12,2%	100,0%
	Mädchen	Anzahl	46	4	50
		Erwartete Anzahl	44,6	5,4	50,0
		Prozent	92,0%	8,0%	100,0%
Stavroupoli	Jungen	Anzahl	19	5	24
		Erwartete Anzahl	21,4	2,6	24,0
		Prozent	79,2%	20,8%	100,0%
	Mädchen	Anzahl	13	3	16
		Erwartete Anzahl	14,3	1,7	16,0
		Prozent	81,3%	18,8%	100,0%
Rethymnon	Jungen	Anzahl	28	6	34
		Erwartete Anzahl	30,3	3,7	34,0
		Prozent	82,4%	17,6%	100,0%
	Mädchen	Anzahl	24	0	24
		Erwartete Anzahl	21,4	2,6	24,0
		Prozent	100,0%	,0%	100,0%
Episkopi	Jungen	Anzahl	27	4	31
		Erwartete Anzahl	27,6	3,4	31,0
		Prozent	87,1%	12,9%	100,0%
	Mädchen	Anzahl	29	2	31
		Erwartete Anzahl	27,6	3,4	31,0
		Prozent	93,5%	6,5%	100,0%

(Fortführung Tabelle 10.11)

Stadt	Geschlecht		Dep1+2	Dep3	Gesamt
Heraklion	Jungen	Anzahl	29	1	30
		Erwartete Anzahl	26,7	3,3	30,0
		Prozent	96,7%	3,3%	100,0%
	Mädchen	Anzahl	19	7	26
		Erwartete Anzahl	23,2	2,8	26,0
		Prozent	73,1%	26,9%	100,0%
Karditsa	Jungen	Anzahl	30	0	30
		Erwartete Anzahl	26,7	3,3	30,0
		Prozent	100,0%	,0%	100,0%
	Mädchen	Anzahl	33	3	36
		Erwartete Anzahl	32,1	3,9	36,0
		Prozent	91,7%	8,3%	100,0%
Itea	Jungen	Anzahl	15	1	16
		Erwartete Anzahl	14,3	1,7	16,0
		Prozent	93,8%	6,3%	100,0%
	Mädchen	Anzahl	14	2	16
		Erwartete Anzahl	14,3	1,7	16,0
		Prozent	87,5%	12,5%	100,0%
Gesamt		Anzahl	879	107	986
		Prozent	89,1%	10,9%	100,0%

Zeichenerklärung: dep1+2 = Gruppierung der Schulkinder mit einem CDI Gesamtpunktwert zwischen 0 und 18 einschließlich; dep3 = Gruppierung der Schulkinder mit einem CDI Gesamtpunktwert zwischen 19 und Maximalwert, jeweils einschließlich

Die in Tabelle 10.11 dargestellte Verteilung weicht nicht signifikant von Erwartungswerten ab [$\chi^2(23) = 30,521$; $p = 0,135$]. Daraus folgern die nachstehenden Aussagen:

10.58 AS: Die geschlechtsspezifische Verteilung der Schülerinnen und Schüler aus den von uns untersuchten Grundschulen variiert nicht signifikant mit dem Wohnort der Kinder.

10.59 AS: Dieses Ergebnis steht in Widerspruch mit den entsprechenden aus der relevanten Analyse der DTK-Daten.

Dem folgt nun die entsprechende Verteilung der Gymnasiasten.

Tabelle 10.12: Geschlechtsspezifische Verteilung der älteren Kinder der Stichprobe (Gymnasiasten) auf die Risikogruppen Dep1+2 und Dep3

Stadt	Geschlecht		Dep1+2	Dep3	Gesamt
Veria	Jungen	Anzahl	27	3	30
		Erwartete Anzahl	24,5	5,5	30,0
		Prozent	90,0%	10,0%	100,0%
	Mädchen	Anzahl	42	5	47
		Erwartete Anzahl	38,4	8,6	47,0
		Prozent	89,4%	10,6%	100,0%
Thessaloniki	Jungen	Anzahl	41	11	52
		Erwartete Anzahl	42,4	9,6	52,0
		Prozent	78,8%	21,2%	100,0%
	Mädchen	Anzahl	54	15	69
		Erwartete Anzahl	56,3	12,7	69,0
		Prozent	78,3%	21,7%	100,0%

(Fortführung Tabelle 10.12)

Stadt	Geschlecht		Dep1+2	Dep3	Gesamt
Galatista	Jungen	Anzahl	34	7	41
		Erwartete Anzahl	33,5	7,5	41,0
		Prozent	82,9%	17,1%	100,0%
	Mädchen	Anzahl	27	2	29
		Erwartete Anzahl	23,7	5,3	29,0
		Prozent	93,1%	6,9%	100,0%
N. Agioneri Kilkis	Jungen	Anzahl	25	7	32
		Erwartete Anzahl	26,1	5,9	32,0
		Prozent	78,1%	21,9%	100,0%
	Mädchen	Anzahl	27 **	1 **	28
		Erwartete Anzahl	22,8	5,2	28,0
		Prozent	96,4%	3,6%	100,0%
Athen	Jungen	Anzahl	91	14	105
		Erwartete Anzahl	85,7	19,3	105,0
		Prozent	86,7%	13,3%	100,0%
	Mädchen	Anzahl	58	15	73
		Erwartete Anzahl	59,6	13,4	73,0
		Prozent	79,5%	20,5%	100,0%
Ksanthi	Jungen	Anzahl	38	10	48
		Erwartete Anzahl	39,2	8,8	48,0
		Prozent	79,2%	20,8%	100,0%
	Mädchen	Anzahl	32	4	36
		Erwartete Anzahl	29,4	6,6	36,0
		Prozent	88,9%	11,1%	100,0%
Stavroupoli	Jungen	Anzahl	20	3	23
		Erwartete Anzahl	18,8	4,2	23,0
		Prozent	87,0%	13,0%	100,0%
	Mädchen	Anzahl	12 **	11 **	23
		Erwartete Anzahl	18,8	4,2	23,0
		Prozent	52,2%	47,8%	100,0%
Rethymnon	Jungen	Anzahl	27	3	30
		Erwartete Anzahl	24,5	5,5	30,0
		Prozent	90,0%	10,0%	100,0%
	Mädchen	Anzahl	36	13	49
		Erwartete Anzahl	40,0	9,0	49,0
		Prozent	73,5%	26,5%	100,0%
Episkopi	Jungen	Anzahl	39	9	48
		Erwartete Anzahl	39,2	8,8	48,0
		Prozent	81,3%	18,8%	100,0%
	Mädchen	Anzahl	31	8	39
		Erwartete Anzahl	31,8	7,2	39,0
		Prozent	79,5%	20,5%	100,0%
Heraklion	Jungen	Anzahl	31	8	39
		Erwartete Anzahl	31,8	7,2	39,0
		Prozent	79,5%	20,5%	100,0%
	Mädchen	Anzahl	26 **	16 **	42
		Erwartete Anzahl	34,3	7,7	42,0
		Prozent	61,9%	38,1%	100,0%

(Fortführung Tabelle 10.12)

Stadt	Geschlecht		Dep1+2	Dep3	Gesamt
Karditsa	Jungen	Anzahl	29	6	35
		Erwartete Anzahl	28,6	6,4	35,0
		Prozent	82,9%	17,1%	100,0%
	Mädchen	Anzahl	24	5	29
		Erwartete Anzahl	23,7	5,3	29,0
		Prozent	82,8%	17,2%	100,0%
Itea	Jungen	Anzahl	36	5	41
		Erwartete Anzahl	33,5	7,5	41,0
		Prozent	87,8%	12,2%	100,0%
	Mädchen	Anzahl	31	8	39
		Erwartete Anzahl	31,8	7,2	39,0
		Prozent	79,5%	20,5%	100,0%
Gesamt		Anzahl	838	189	1027
		Prozent	81,6%	18,4%	100,0%

Zeichenerklärung: dep1+2 = Gruppierung der Schulkinder mit einem CDI Gesamtpunktwert zwischen 0 und 18 einschließlich; dep3 = Gruppierung der Schulkinder mit einem CDI Gesamtpunktwert zwischen 19 und Maximalwert, jeweils einschließlich

Die in Tabelle 10.12 dargestellte Verteilung weicht statistisch sehr signifikant von Zufallswerten ab. [$\chi^2(23) = 44,097$; $p = 0,005$].

Dieses Ergebnis führt dann zu nachstehenden Aussagen:

10.60 AS: Die geschlechtsspezifische Verteilung der Schülerinnen und Schüler aus Gymnasien auf die zwei nach CDI-Gesamtpunkte gebildeten Risikogruppen variiert signifikant mit dem Wohnort der Kinder.

10.61 AS: Signifikante Abweichungen von den Erwartungswerten treten in drei Orten (N. Agioneri Kilkis, Stavroupoli und Heraklion) auf.

10.62 AS: In N. Agioneri Kilkis sind die Mädchen in der Gruppe hoher Belastung Dep3 um 80,77% unterrepräsentiert.

10.63 AS: In Stavroupoli und in Heraklion sind die Mädchen in der Gruppe Dep3 um 161,90% respektive 107,79% überrepräsentiert.

Den Arbeitsschritten in der Analyse der vergleichbaren Daten folgend, wird nachfolgend auch für die CDI-Daten geprüft, ob die zuvor mitgeteilten signifikanten Abweichungen der Verteilung von Schulkindern auf die CDI Risikogruppen in den Schulorten möglicherweise von der Größe der Siedlungsflächen (Groß-, Mittel- und Kleinstadt) beeinflusst werden. Die drei nachstehenden Tabellen beschreiben die Verteilung der Gesamtstichprobe wie auch die geschlechtsspezifische Verteilung der Grundschul Kinder und Gymnasiasten auf die CDI-Gruppen Dep1+2 und Dep3 und prüfen sie auf Signifikanz.

Tabelle 10.13: Nach drei Stadtgrößen geordnete Verteilung der Gesamtstichprobe auf die nach CDI-GPW gebildeten Gruppen Dep1+2 und Dep3

		Dep1+2	Dep3	Gesamt
Großstadt	Anzahl	679	120	799
	Prozent	85,0%	15,0%	100,0%
Mittelstadt	Anzahl	560	82	642
	Prozent	87,2%	12,8%	100,0%
Kleinstadt	Anzahl	478	94	572
	Prozent	83,6%	16,4%	100,0%
Gesamt	Anzahl	1717	296	2013
	Prozent	85,3%	14,7%	100,0%

Zeichenerklärung: dep1+2 = Gruppierung der Schulkinder mit einem CDI Gesamtpunktwert zwischen 0 und 18 einschließlich; dep3 = Gruppierung der Schulkinder mit einem CDI Gesamtpunktwert zwischen 19 und Maximalwert, jeweils einschließlich

Die in Tabelle 10.13 dargestellte Verteilung ist nach einer χ^2 -Prüfung nicht signifikant [$\chi^2(2) = 3,337$; $p = 0,189$].

10.64 AS: Die Verteilung der an der Untersuchung 2034GR beteiligten Schulkinder auf den zwei CDI-Risikogruppen für das Bestehen depressiver Belastung variiert nur zufällig mit den nach den Einwohnerzahlen als Groß-, Mittel- und Kleinstädten unterschiedenen Schulstandorten.

10.65 AS: Das zuvor mitgeteilte Ergebnis stimmt mit den entsprechenden Befunden aus der Durchführung des DTK überein.

Nachfolgend findet sich die Tabelle mit der Verteilung der Grundschüler aus Schulorten unterschiedlicher Größe auf die Risikogruppe dep1+2 und dep3.

Tabelle 10.14: Nach drei Stadtgrößen geordnete Verteilung der Kinder aus den Grundschulen auf die nach CDI-GPW gebildeten Gruppen Dep1+2 und Dep3

			Dep1+2	Dep3	Gesamt
Großstadt	Jungen	Anzahl	197	18	215
		Prozent	91,6%	8,4%	100,0%
	Mädchen	Anzahl	181	23	204
		Prozent	88,7%	11,3%	100,0%
Mittelstadt	Jungen	Anzahl	156	19	175
		Prozent	89,1%	10,9%	100,0%
	Mädchen	Anzahl	149	14	163
		Prozent	91,4%	8,6%	100,0%
Kleinstadt	Jungen	Anzahl	108	19	127
		Prozent	85,0%	15,0%	100,0%
	Mädchen	Anzahl	88	14	102
		Prozent	86,3%	13,7%	100,0%
Gesamt	Anzahl	879	107	986	
	Prozent	89,1%	10,9%	100,0%	

Zeichenerklärung: dep1+2 = Gruppierung der Schulkinder mit einem CDI Gesamtpunktwert zwischen 0 und 18 einschließlich; dep3 = Gruppierung der Schulkinder mit einem CDI Gesamtpunktwert zwischen 19 und Maximalwert, jeweils einschließlich

Die in Tabelle 10.14 dargestellte Verteilung weicht nicht signifikant von Erwartungswerten ab [$\chi^2(23) = 5,354$; $p = 0,374$].

10.66 AS: Die geschlechtsspezifische Verteilung von Grundschulkindern auf die zwei CDI-Risikogruppen variiert nur zufällig mit der Unterscheidung der Schulstandorte Orte als Groß-, Mittel- und Kleinstädten.

10.67 AS: Das zuvor mitgeteilte Ergebnis stimmt mit dem entsprechenden Befunden aus den Analysen der DTK-Daten überein.

Die nachstehende Tabelle enthält die Daten über die geschlechtsspezifische Verteilung aller CDI-GPW in Schulstandorten unterschiedlicher Größe.

Tabelle 10.15: Nach drei Stadtgrößen geordnete Verteilung der Kinder aus den Gymnasien auf die nach CDI-GPW gebildeten Gruppen Dep1+2 und Dep3

			Dep1+2	Dep3	Gesamt
Großstadt	Jungen	Anzahl	163	33	196
		Prozent	83,2%	16,8%	100,0%
	Mädchen	Anzahl	138	46	184
		Prozent	75,0%	25,0%	100,0%
Mittelstadt	Jungen	Anzahl	121	22	143
		Prozent	84,6%	15,4%	100,0%
	Mädchen	Anzahl	134	27	161
		Prozent	83,2%	16,8%	100,0%
Kleinstadt	Jungen	Anzahl	154	31	185
		Prozent	83,2%	16,8%	100,0%
	Mädchen	Anzahl	128	30	158
		Prozent	81,0%	19,0%	100,0%
Gesamt	Anzahl	838	189	1027	
	Prozent	81,6%	18,4%	100,0%	

Zeichenerklärung: dep1+2 = Gruppierung der Schulkinder mit einem CDI Gesamtpunktwert zwischen 0 und 18 einschließlich; dep3 = Gruppierung der Schulkinder mit einem CDI Gesamtpunktwert zwischen 19 und Maximalwert, jeweils einschließlich

Die in Tabelle 10.15 dargestellte Verteilung ist nicht signifikant [$\chi^2(2) = 7,176$; $p = 0,208$].

10.68 AS: Die geschlechtsspezifische Verteilung von Gymnasiasten auf die zwei CDI-Risikogruppen variiert nur zufällig mit den nach Einwohnerzahlern als Groß-, Mittel- und Kleinstädte unterschiedenen Schulstandorten.

10.69 AS: Das über die Zuweisung der Gymnasiasten in Schulstandorten unterschiedlicher Größe an die beiden CDI-Gruppen zuvor mitgeteilten Ergebnis weicht von dem entsprechenden Befunden aus der Analyse der DTK-Daten ab. Dort wurde gefunden, dass in der DTK-Gruppe hoher Belastung, DepBel3, die älteren Jungen aus Großstädten und Mittelstädten signifikant unterrepräsentiert und die Mädchen aus Großstädten überrepräsentiert sind.

Die über die Verteilung CDI Gesamtpunktwerte ermittelte Zuweisung von Jungen und Mädchen an die CDI Risikogruppen dep1+2 und dep3 sollte, wie schon bei entsprechenden Gelegenheiten zuvor, mit den korrespondierenden Ergebnissen aus der Durchführung des DTK verglichen werden. Die DTK Risikogruppen sind als DepBel1+2 und DepBel3 bezeichnet worden. Beide Formen der Gruppierung, die für den CDI und die für den DTK praktizierte, verfahren nach

identischen Grundprinzipien. Die Gesamtverteilung der Punktwerte aus der Durchführung des jeweiligen Verfahrens wird in drei Gruppen unterteilt, von denen hier die mit den niedrigeren Punktwerten, vom Punktwert 0 bis zum jeweiligen Trennwert, als dep1+2 und DepBel1+2 bezeichnet worden sind. Da die Punktwerte über die Durchführung unterschiedlichen Verfahren bekannt werden, sind die ähnliche Bezeichnung und die Übereinstimmungen in der Bestimmung der Gruppen keine hinreichenden Belege für Übereinstimmungen zwischen den CDI Gruppen und den DTK Gruppen. Der über die Gruppenbildung vertretene Anspruch ist demgegenüber aber identisch. Dep3 und DepBel3 sollen die Schulkinder identifizieren, die mit hoher Wahrscheinlichkeit depressive Störungen entwickelt haben. Idealerweise müssten dann, falls der Einspruch eingelöst werden sollte, über die unterschiedlichen Verfahren identische Kinder benannt werden. Ob dieser Einspruche eingelöst wird, lässt sich prüfen. An dieser Stelle wird unterhalb dieses Anspruchsniveaus überprüft, ob die Zuweisungen an die höchsten Risikogruppen in ihren Mengenangaben und in der Richtung der Zuweisung, in erwartbarer Annäherung übereinstimmen.

10.8 Die Befundlage nach der Analyse der Beziehungen zwischen CDI-Gesamtpunktwerten und verschiedenen Teilstichproben in dieser Studie

Das nahezu allen empirischen Untersuchungen zum Auftreten depressiver Störungen bei Kindern und Jugendlichen unterliegende gemeinsame Forschungsinteresse ist auf die Erschließung von Einsichten in die Auftretensbedingungen dieser Störung und auf Fortschritte in der gegenüber ihr wirksamen Präventionen und Therapien gerichtet. Die Analyse der Beziehungen zwischen CDI Gesamtpunktwerten und unterschiedlichen Teilstichproben in dieser Untersuchung vermag dazu das beizutragen, was die analysierten Befunde an Informationen bieten.

Die Möglichkeit der Erarbeitung von Informationen ist durch den verfügbaren Datensatz und die einsetzbaren Analysetechniken determiniert.

Die CDI-Gesamtpunktwerte sind, je nach ihrer Ausprägung unterschiedlich gewichtige Indizien für die Wahrscheinlichkeit des Bestehens depressiver Störungen bei den untersuchten Schulkindern.

Die über die Suche nach signifikanten Unterschieden in variierenden Verteilungen dieser Werte auf die unterschiedenen Risikogruppen dep1.dep2 und dep3 gefundenen Ergebnisse sind in den vorangegangenen Abschnitten dieses Kapitels eingehend gewürdigt worden. Insgesamt ist zu vermerken, dass sie auf Bedingungen für das Auftreten depressiver Störungen bei Kindern und Jugendlichen hinweisen, die, insbesondere unter Bezugnahme auf die internationale Fachliteratur, als spezifisch für depressive Störungen bei Kindern und Jugendlichen angesehen werden dürfen. In dieser Untersuchung wird nachgewiesen, dass sie auch in Griechenland auftreten. In diesem Kapitel gilt dies speziell für die Befunde zur Variation der Hinweise auf depressive Störungen mit durch Teilstichproben in dieser Untersuchung vertretenen Merkmalen.

Als in diesem Sinne spezifisch sollten insbesondere die nachstehenden empirisch begründeten Hinweise angesehen werden:

Die mit dem CDI erfassten Indizien für das Bestehen depressiver Störungen variieren mit dem Alter der Schulkinder. Der Anteil an Kindern die selbst relevante Hinweise auf ihre Betroffenheit von depressiven Symptomen geben steigt in der Altersspanne zwischen 8 Jahren und dem Alter von 15 Jahren (und mehr) zunehmend an.

Der immer wieder zu gebende Hinweis darauf, dass in den unteren Altersgruppen von Grundschulern häufiger Unterschreitungen von Erwartungswerten für das Auftreten von Hinweisen auf depressive Störungen zu beobachten sind, darf nicht dazu verleiten, das Auftreten von Hinweisen auf depressive Störungen in unteren Altersklassen zu übersehen und in seiner

Bedeutung zu unterschätzen. Alle Hinweise auf das Bestehen depressiver Störungen sind hoch relevant, auch wenn sie signifikant unter Erwartungswerten bleiben sollten.

Das mit zunehmendem Alter beobachtete Ansteigen der prozentualen Anteile von Mädchen an den Zuweisungen zur jeweils höchsten Risikogruppe gilt auch für griechische Kinder und Jugendliche.

Die gegenüber den Jungen deutlichen geschlechtsspezifischen Unterschiede in dem mit zunehmendem Alter zu beobachtenden Anwachsen der Hinweise auf bestehende depressive Störungen bei Mädchen sind zum einen epidemiologisch bedeutsam. Zum anderen sprechen sie dafür vermehrt zu versuchen, die verantwortlichen Bedingungen zu erfassen.

In dieser griechischen Untersuchung finden sich an mehreren Stellen, so auch in der Analyse der CDI Werte Hinweise darauf, dass sowohl depressive Störungen das Risiko für schulisches Versagen erhöhen, als auch schulisches Versagen die Wahrscheinlichkeit für das Auftreten depressiver Symptome ansteigen lässt.

Obschon die Hinweise darauf nicht in demonstrativen Ausmaß auftreten, besteht empirisch begründeter Anlass, auf mutmaßliche Beziehungen zwischen der Ausprägung des Auftretens depressiver Symptome und dem jeweils wirksamen lokalen Lebensumständen bestehen. Welche Qualitäten unter der Symbolisierung als lokale Lebensumstände wirksam sind, kann bislang vermutet, aber nicht nachgewiesen werden. Da dieser Bedingungsaspekt für das Auftreten depressiver Störungen bislang wenig beobachtet worden ist, wird an dieser Stelle ausdrücklich auf ihn verwiesen.

Einige, auch leicht zu übersehende Unterschiede in den über DTK und CDI gewonnenen Befunden könnten Hinweise auf Unterschiede in der Sensibilität der Verfahren gegenüber spezifischen Aspekten des depressiven Störungsbildes sein. Als solche stellen sie Anlässe für nachforschende Untersuchungen dar.

Die in diesem Kapitel dargestellten Ergebnisse und Befunde sind als Einzelergebnisse beachtenswert. Beachtet werden muss aber auch, dass sie Teilaspekte in der Gesamtheit der mit dieser Untersuchung erarbeiteten Ergebnisse sind. Insbesondere sind sie, wie alle CDI Ergebnisse mit denen aus der Anwendung des DTK zu vergleichen. Im Idealfall bilden beide Verfahren die Quelle für verfahrensübergreifende Befunde.

Kapitel 11

Analyse der mit dem SDQ gewonnenen Daten

„In the context of clinical diagnostics, behavioural assessment should not be restricted to a narrow focus (e.g. anxiety) but should rather attempt to cover a broader area of different behavioural aspects, thus establishing a complete psychopathological profile” (Rothenberger und Woerner, 2004, S. II/1). Die von Rothenberger und Woerner zitierte Äußerung weist auf den Hintergrund der innerhalb dieser Untersuchung getroffenen Entscheidung über den Einsatz eines Verfahrens zur allgemeinen Erfassung psychischer Beeinträchtigungen der befragten griechischen Schüler und Schülerinnen hin. Speziell innerhalb der Depressionsforschung bei Kindern stellen die wiederholt beobachteten komorbiden Störungsbilder einen Gemeinplatz dar, so dass man selten über eine „reine“ depressive Störungen sprechen kann (Reicher, 1998). Das gleichzeitige Auftreten von depressiven Symptomen und solchen, die nach den aktuellen Klassifikationssystemen eher anderen Störungsbildern zuzuordnen wären, sind insofern für die betroffenen Kinder von besonderer Bedeutung, weil solche Symptome sowohl die Stabilität als auch den Verlauf der depressiven Befindlichkeit beeinflussen (DuBois et al., 1995). Sowohl die aktuellen als auch die künftigen Folgen des Bestehens solcher Komorbiditätsmuster als individuelles Störungsbild sind, belegt durch empirische Ergebnisse, als potentiell folgenreich einzuschätzen. Selbst wenn weder die depressiven noch die komorbiden Symptome den Kriterien der entsprechenden klinischen Störungsbilder voll entsprechen sollten, ist von deren Einwirkung auf das individuelle Gesamtbefinden und von Wechselwirkungen auszugehen. “Even if scores in no single area satisfy clinical diagnostic criteria, such pattern of comorbidity may contribute to the persistence of existing adaptive difficulties and the emergence of more serious impairments in functioning in one or more areas.” (DuBois et al., 1995, S. 393)

Für den Ansatz der Untersuchung 2034GR, die hier erörtert wird, kann die zitierte Überlegung als Begründung dafür verwendet werden, neben depressionsspezifischen Messinstrumenten auch ein weiteres, aber weitaus unspezifischeres Instrument einzusetzen, eben eines, das auf klinisch relevante individuelle psychische Belastungen außerhalb des Störungsbildes der Depression aufmerksam macht.

Als Verfahren zur Erfassung einer solchen noch unspezifischen und daher zunächst als allgemein bezeichneten psychischen Beeinträchtigung wurde innerhalb dieser Untersuchung das Verfahren “Strengths and Difficulties Questionnaire” (SDQ) von Goodman (1997) ausgewählt. Nach der speziellen Theorie dieses Tests und nach seiner Struktur, sollte der Einsatz des SDQ 1. im Hinblick auf die als depressiv eingeschätzten Mädchen und Jungen das Aufnehmen von Indizien über das potentielle Auftreten komorbider Störungen ermöglichen. Zugleich würde 2. für alle als sehr wahrscheinlich depressiv unbelastet erkannten Mädchen und Jungen durch SDQ-Daten die Frage zu beantworten sein, ob Verdachtsmomente für das Vorliegen anderer psychischer Störungen zu beachten wären. Schließlich würde 3. bekannt werden, ob etwa für die wegen entsprechend niedriger Gesamtpunktwerte im DTK oder CDI als mutmaßlich nicht depressiv belastet eingestuften Mädchen und Jungen durch deren auffallende Werte in der Subskala „Emotional symptoms“ des SDQ die Möglichkeit beachtet werden sollte, dass erst in einem dritten Verfahren, eben dem SDQ, Selbstberichte über das Auftreten depressiver Symptome bekannt werden.

Wie schon in Kapitel 6.5 zur Prüfung der Messqualitäten des SDQ innerhalb dieser Studie angegeben worden war, ist das Verfahren SDQ das einzige unter den bei der hier beschriebenen

Untersuchung verwendeten Verfahren, für das eine griechische Übersetzung, die von Bibou, Stogiannidou, Kioseoglou und Papagoergiou schon vorab verfügbar war und dass auch schon in Untersuchungen mit griechischen Kindern in dieser Form Anwendung gefunden hat (Bibou et al., 2001; Kypraiou, 2002). Der Einsatz des SDQ innerhalb der Untersuchung 2034GR diente somit dann auch einer Überprüfung der Messqualitäten dieses Verfahrens in einer größeren und repräsentativeren Stichprobe, nachdem sich über die Studien von Bibou und Mit. (2001) und Kypraiou (2002) nur relativ niedrige Reliabilitätswerte für die Subskalen des Verfahrens ergeben hatten. Unter Kenntnisnahme dieser Werte war der Einsatz des SDQ zur Übernahme der weiter oben angeführten drei Funktionen für die Untersuchung 2034GR entsprechend kritisch zu bedenken.

Das ist der Grund dafür, dass an dieser Stelle der Erörterungen erneut auf die Messqualitäten des SDQ eingegangen wird.

11.1 Das SDQ als Screening Verfahren zur Identifizierung psychisch belasteter Kinder

Das Verfahren „Strengths and Difficulties Questionnaire“ (SDQ; Goodman, 1994, 1997; Goodman, Meltzer & Bailey, 1998) bezieht sich durch seine Items auf verhaltensnahe Emotionen und zwischenmenschliche Beziehungen. Seine Entwicklung basiert auf dem Rutter Questionnaire (Rutter, 1967), das sich in verschiedenen Zusammenhängen als ein valides und reliables psychometrisches Verfahren erwiesen hat (Elander & Rutter, 1996). Die aktuell gültige Form des SDQ umfasst 25 Fragen, die sich in 5 Skalen unterteilen. Die werden von Goodman wie folgt bezeichnet: 1. „Prosocial behaviour“, 2. „Hyperactivity-Inattention“, 3. „Emotional symptoms“, 4. „Conduct problems“ und 5. „Peer problems“. Die 25 Fragen des SDQ stimmen nach Angaben des Autors sowohl mit den nosologischen Konzepten des DSM-IV als auch mit den im ICD-10 vertretenen überein. Die Konstruktion basierte, wie üblich, auf einem Itempool, aus dem die sich bewährenden Items nach eine Reihe von Faktorenanalysen im Zusammenhang mit Probetests bei Kindern durchschnittlicher Intelligenz (Goodman, 1997) ausgewählt und auf den jetzt verwendeten Itemssatz reduziert worden sind. Nach Goodman und Scott (1999) trifft die nachstehende Aussage zu: „...the fact that the SDQ was designed on the basis of theory as well as previous factor analysis does not seem to have undermined its factor structure...“ (S. 18).

Die Entwicklung eines gegenüber dem Rutter Questionnaire aktualisierten und kürzeren Verfahren, in dem auch nach positiven Eigenschaften der Kinder gefragt wird, sollte nach Goodman einerseits zu einem in der Anwendung ökonomischeren Instrument und dann auch zu einem mit höherer Akzeptanz seitens der befragten Personen führen. Das Verfahren ist zur Anwendung durch Eltern und Lehrer von Kindern und Jugendlichen zwischen 4 und 16 Jahren geeignet und kann auch bei Kindern und Jugendliche zwischen 11 und 16 Jahren eingesetzt werden.

Seit etwa dem Jahr 2000 ist eine größere Zahl von Publikationen bemüht gewesen, die Nützlichkeit des SDQ als Messinstrument zur Erfassung psychischer Störungen bei Kindern durch Prüfung seiner Messqualitäten empirisch zu belegen. Für seine Nutzbarkeit sprechen inzwischen die etwa in 40 Sprachen vorhandenen Versionen des Verfahrens. Sie sind alle unter der Internetadresse www.sdqinfo.com verfügbar.

In den Veröffentlichungen von Marzocchi et al. (2004) und von Woerner et al. (2002) sind vor allem die Untersuchungspläne und auch einige erste Ergebnisse aus der Anwendung des SDQ in Kinderpopulationen aus Italien, Spanien, Portugal, Kroatien und Frankreich (Marzocchi et al., 2004) und aus Brasilien, Kanada, Australien, Gaza, Yemen, und Pakistan (Woerner et al., 2004)

dargestellt. Des Weiteren sind Ergebnisse aus der Anwendung des SDQ in europäischen Ländern außerhalb Englands in den nachfolgend genannten Berichten, Erörterungen und Analysen zu finden. Ausgewählt sind hier aber ausschließlich Studien, die Angaben über die Durchführung der Self-report Version des Verfahrens bieten.

Die bisher in unterschiedlichen Ländern veröffentlichten Daten zu Reliabilität und Validität des SDQ sind in hohem Maße übereinstimmend. Das betrifft interessanterweise sowohl erfreuliche als auch problematische Befunde über die psychometrische Verlässlichkeit des Verfahrens. So berichten Goodman (1999) und Muris et al. (2003) von einer befriedigenden bis guten Test-Retest Reliabilität. Darüber hinaus finden sich für die hypothetische Struktur des Verfahrens immer wieder Bestätigungen durch Ergebnisse aus Faktorenanalysen mit Daten aus unterschiedlichen Stichproben (Goodman, 2001; Muris et al., 2003; Rønning et al., 2004; Smedje et al., 1999). Problematisiert werden in der einschlägigen Literatur die relativ niedrigen Werte für die interne Konsistenz der SDQ-Subskalen. Die nachstehende Tabelle gibt einen Überblick über die einschlägigen Daten aus unterschiedlichen Stichproben. Die Tabelle stellt Reliabilitätswert für die Subskalen und die Gesamtskala des SDQ dar. Als kritisch werden in der Literatur vor allem die Konstrukte „Conduct problems“ und „Peer problems“ diskutiert, wobei von den meisten Autoren ein Koeffizient von $\alpha = 0,60$ als Wertlimit angesetzt ist. Wenn aber in Abweichung davon, so wie bei Bortz & Döring, (2003), für Forschungsinstrumente ein Alpha-Wert von 0,70 verlangt wird, dann müssen, wie aus Tabelle 11.1 zu ersehen ist, alle 5 Subskalen des Verfahrens als für Forschungszwecke ungeeignet eingestuft werden.

Tabelle 11.1: Reliabilitätswerte der SDQ-Subskalen aus unterschiedlichen Studien

Autor	Untersuchung			Alpha-Werte der SDQ-Subskalen					
	Ort	N	Alter	Td	Es	Cp	Hi	Pp	Pb
Goodman et al., 1998 ¹	England	199	11-16	.82	.75	.72	.69	.61	.65
Goodman, 2001	England	3983	11-15	.80	.66	.60	.67	.41	.66
Koskelainen et al., 2000	Finnland	601	9-15	.71	.69	.57	.66	.63	.69
Muris et al., 2003	Niederlande	562	9-15	.78	.71	.45	.72	.54	.62
van Widenfelt et al., 2003	Niederlande	1288	11-16	.70	.63	.47	.66	.39	.60
Rønning et al., 2004	Norwegen	2159 J 2017 M	11-16		.64 .67	.54 .44	.66 .67	.58 .61	.62 .66
Bibou et al., 2001	Griechenland	238	11-15		.70	.45	.57	.54	.69
Kypraiou, 2002	Griechenland	861	10 und 13	.69	.69	.42	.61	.49	.52
Kampisiou, 2003	Griechenland	1034	11-16	.78	.70	.43	.60	.60	.62

Zeichenerklärung: ¹= Stichprobe von 83 Schulkindern und 116 psychiatrisch hospitalisierten Kindern; Td= Total difficulties, Es = Emotional symptoms, Cp= Conduct problems, Hi= Hyperactivity-inattention, Pp= Peer problems, Pb= Prosocial behaviour; J= Jungen, M= Mädchen

Die Kürze der Subskalen, die positive Formulierung einiger Items, wie auch die eventuell fehlende Bereitschaft der Jugendlichen, von ihren Problemen zu berichten, werden als mögliche Faktoren genannt, die zu den niedrigen Reliabilitätskoeffizienten geführt haben könnten. So erklären beispielsweise Muris et al. (2003) die in ihrer eigenen Studie auftretenden testtheoretisch inakzeptablen Alpha-Werte für die Subskalen „Conduct problems“ und „Peer problems“ (0,45 und 0,54) wie folgt: „Given that SDQ subscales only consist of 5 items, this result was hardly

surprising. In addition, inspection of the item-total correlations of the pertinent scales showed that the low reliability coefficients could be attributed to a number of reversely scored and low frequent items” (Muris et al., 2003; S. 7). Bei van Widenfelt et al. (2003) wird wie folgt argumentiert: “...the weak internal consistencies of the self-report SDQ conduct problems and peer problems subscales may indicate that one or two items do not fit with the rest of the scale and/or the conduct and peer problems are partly reported (van Widenfelt und Mit., 2003, S. 285). Eine weitere Erklärung, die zu den bereits geschilderten gewisse Ähnlichkeiten aufweist, ist die von Rønning et al. (2004) abgegebene: “Low alphas indicate that the items included were measuring somewhat different properties.... Common to these domains in this study were very small intra-scale correlations between oppositely posed questions. This suggests a potential problem with the transformation of some of the positively worded items to negative scales. It could be that the respondents were not interpreting these questions in the way intended by the designer of the questionnaire.” (S. 80). Dementsprechend empfehlen diese Autoren: “...avoiding semantic similarities, positive and negative wording within the same construct, and two sentences within the same item. It might also be worth linking some of the items to more than one latent construct“ (S. 81).

Zur Kenntnis zu nehmen ist, dass, der Literaturlage folgend, die Messqualitäten der einzelnen Skalen des SDQ in Frage zu stellen sind.

Wenn schon die Messqualitäten der Einzelskalen im SDQ kritisiert werden, so wäre nun die Betrachtung der Messqualitäten der Gesamtskala des SDQ um so eingehender zur Kenntnis zu nehmen.

Weiterhin ist dann aber auch dem Problem der Validität des SDQ nachzugehen. Rønning et al. (2004) äußern sich dazu wie folgt: “Moreover, the SDQ discriminates well between children with and without psychopathological symptoms (Goodman, 1999; Goodman, Meltzer & Bailey, 1998; Klasen et al., 2000; Mullick & Goodman, 2001) and there is evidence to indicate that it can be employed as an effective screen for child psychiatric disorders in community samples (Goodman, 2001; Goodman, Ford, Simmons et al., 2000)

Studien, die das SDQ unter dem Aspekt des Konzepts der konvergenten Validität untersucht haben, verweisen auf eine hohe und in der richtigen Richtung verlaufende Korrelation des Verfahrens mit dem YSR, das ist die Self-report Form des etablierten CBCL, (Goodman, 1997; Goodman und Scott, 1999; Klasen et al., 2000; Koskelainen et al., 2000; van Widenfelt et al., 2003). Solche Aussagen liegen auch für weitere Verfahren vor, so für die Beziehungen zwischen SDQ und dem depressionsspezifischen Verfahren CDI, dem RCMAS zur Erfassung von Angststörungen und dem ADHDQ, der typische Symptome einer ADHD-Störung zur erfassen vermag. (Muris et al., 2003; van Widenfelt et al., 2003). Alle diese Ergebnisse würden für den SDQ eine breite Spezifität für die Erfassung von Informationen über das Auftreten unterschiedlicher Störungsbilder annehmen lassen. Dafür müsste das Verfahren sensibel gegenüber Indikatoren sein, die in allen erwähnten Störungsbildern auftreten.

Eine gegenüber der konvergenten Validität abzusetzende spezielle Form ist die diskriminante Validität.

Goodman and Scott (1999), Klasen et al. (2000) and Becker et al. (2004) berichten, dass die Self-report Form des SDQ und das YSR sich als vergleichbar effizient bei der Unterscheidung zwischen klinisch unauffälligen Kindern einerseits und in klinischen Abteilungen hospitalisierten Kindern andererseits bewiesen hätten.

Weitere Angaben über die diskriminante Leistung des SDQ sind unter anderen aus einer Studie von Goodman, Renfrew und Mullick (2000) bekannt geworden. Diese Studie untersuchte in einer

klinischen Stichprobe von 101 in London lebenden Kindern mit dem Durchschnittsalter von 10,3 Jahren die prädiktive Fähigkeit der SDQ-Scores zur Identifizierung klinisch diagnostizierter Kinder. Die Auswahl solcher Kinder wurde nach einem von Goodman entwickelten Algorithmus vollzogen (s. ebd., S. 131). In diesem Artikel werden auch entsprechende Daten aus einer klinischen Stichprobe von 89 Kindern in Dhaka, Bangladesh, dargestellt. Die gefundenen Korrelationen zwischen den über den verwendeten Algorithmus errechneten SDQ-Scores und den klinischen Befunden (Diagnosen als „conduct“, „emotional“ oder „hyperactivity“) werden für beide Stichproben befriedigend bis gut genannt (Kendall's tau b zwischen 0,49 und 0,73; $p < 0,001$). Tabelle 11.2 stellt in Werten von Sensitivität, Spezifität und positiver und negativer Prädiktion die diagnostische Leistungsfähigkeit des SDQ nach seiner Anwendung in den oben erwähnten Stichproben dar. Die Autoren bewerten die im Allgemeinen übereinstimmenden Ergebnisse aus den zwei kulturell unterschiedlichen Stichproben als erfreulich und dies sowohl im Hinblick auf die psychometrische Qualität des SDQ, als auch in Bezug auf den in beiden Stichproben übereinstimmend verwendeten Algorithmus. Für den wird dann aber wegen entsprechender empirischer Daten aus einer Stichprobe von klinisch unbelasteten Schulkindern empfohlen, ihn nur in klinischen Stichproben zu verwenden (Goodman, Ford, Simmons, Gatward & Meltzer, 2000). Das Befolgen einer solchen Empfehlung führt unabweisbar zu der Frage nach dem dann noch verbleibenden klinischen Nutzen des SDQ. Messtheoretisch wird der Verdacht genährt, dass die in klinischen Stichproben genutzten Indikatoren auch in außerklinischen Gruppen gehäuft auftreten könnten.

Tabelle 11.2: Effizienz des SDQ in zwei klinischen Stichprobe von Kindern aus London (N=101) und Kindern aus Dhaka (N=89) (Angaben aus Goodman, Renfrew & Mullick, 2000, S. 133)

Diagnose	Sensitivität		Spezifität		Positiver prädiktiver Wert		Negativer prädiktiver Wert	
	London	Dhaka	London	Dhaka	London	Dhaka	London	Dhaka
Conduct	90%	86%	47%	82%	68%	60%	83%	95%
Emotional	81%	86%	80%	84%	52%	86%	94%	90%
Hyperactivity	89%	89%	78%	81%	70%	35%	93%	98%

Hinweise über die diskriminante Validität des deutschen SDQ bieten Klasen et al. (2000) mit Werten aus ROC-Kurven („area under curve“: total score: 0,91, emotional problems scale: 0,85, conduct problems scale: 0,97, hyperactivity: 0,94, prosocial scale: 0,78). Die diskriminante Leistung des SDQ in dieser Studie war innerhalb der klinischen Stichprobe, wie erwartet, leicht reduziert („area under curve“: emotional disorder: 0,72, conduct disorder: 0,81, hyperactivity disorder: 0,77). Die Autoren dieser Studie betonen die Tatsache, dass die kurze Form des SDQ Validitätsleistungen aufweist, die der Vergleichsform entsprechen, obschon doch kürzere Verfahren in der Regel wenig reliabel sind.

Zum Anwendungsbereich des SDQ äußerten sich Goodman et al. wie folgt: „Goodman originally suggested that the cut-off points of the SDQ could act as useful „warning signals“ to the clinician that an adolescent might have a disorder...” (Goodman, Meltzer & Bailey, 1998, unter Hinweis auf Goodman, 1997). Diesen Hinweis ergänzt das nachfolgende Zitat: „...researchers could appropriately use the self-report SDQ to examine group differences, though they could not expect to make accurate diagnoses on individual subjects. Similarly, clinicians could sensibly use the self-report SDQ to assess young people's degree of awareness of their own problems...”

(Goodman, Meltzer & Bailey, 1998, S. 128-129). Andere Autoren äußern sich wie folgt dazu: "...efficient and economical screening instrument for preventive research on large community samples ..."(Rønning, Handegaard and Mørch, 2004).

11.2 Die aktuelle Datenbasis der nachfolgenden Analysearbeit aus der Anwendung des SDQ in der Studie 2034GR

Die für die anstehende Analysen der mit dem SDQ gewonnenen Daten verfügbare Datenbasis besteht aus den Antworten, die die Mädchen und Jungen aus den griechischen Gymnasien unserer Stichprobe auf die 25 Aussagen des Verfahrens gegeben haben. Wie schon dargestellt, bilden diese 25 Aussagen fünf Subskalen mit jeweils gleicher Itemszahl. Diese Subskalen wurden von Goodman wie folgt benannt: 1. „Prosocial behaviour“, 2. „Hyperactivity-Inattention“, 3. „Emotional symptoms“, 4. „Conduct problems“ und 5. „Peer problems“. Die Skalen 2 bis 5 bilden zusammen eine mit "Total difficulties" bezeichnete Gesamtskala, die Indizien für den Grad der psychischen Belastung der befragten Kinder liefern sollte.

Mit dem SDQ werden die Kinder gebeten, jede einzelne vorgegebene Aussage, die das Auftreten spezieller Symptome betreffen, als "nicht zutreffend", "teilweise zutreffend" oder "eindeutig zutreffend" durch das Ankreuzung der jeweils auf sie selbst zureffenden Bewertung auszuwählen.. Die letztgenannte Aussage(„eindeutig zutreffend“) verweist auf die hochwahrscheinliche Anwesenheit des jeweils nachgefragten Symptoms. Das Gegenteil gilt für positiv formulierte Aussagen (Items 7, 11, 14, 21, 25 und für die Items der Subskala „Prosocial behaviour“, nämlich 1, 4, 9, 17 und 20).

Der höchste Gesamtpunktwert, den ein befragtes Kind in jeder Skala erreichen kann, ist 10 (5 Items * 2 Punkte). Demzufolge kann ein individueller Gesamtpunktwert in der Gesamtskala zwischen 0 und 40 Punkte liegen (4 Subskalen * 10 Punkte).

Die fünfte Subskala des Verfahrens („Prosocial behaviour“) soll nicht Schwierigkeiten, sondern Stärken der Kinder erfassen und wird aus diesem Grund nicht in den Gesamtscore eingerechnet: „The prosocial score is not incorporated in the reverse direction into the total difficulties score since the absence of prosocial behaviour is conceptually different from the presence of psychological difficulties“ (Goodman, Meltzer & Bailey, 1998, S. 126).

Da die Self-Report Form des SDQ für Kinder und Jugendlichen zwischen 11 und 16 Jahren konzipiert worden ist, wurde das Verfahren in der hier beschriebenen Untersuchung auch nur bei den älteren Kindern der Stichprobe angewendet. Geantwortet auf die Fragen des SDQ hat damit auch nur ein Teil der Gesamtstichprobe, nämlich nur die Kinder aus den von uns besuchten Gymnasien.

11.3 Verteilung der Werte der SDQ-Skala „Total difficulties“ in der Untersuchung 2034GR

Die in den bis hier geführten Ausführungen zur eigenen Untersuchung als 2034GR bezeichnete Stichprobe umfasst Kinder und Jugendliche, die nach den griechischen Gegebenheiten als Kinder und Jugendliche aus Grundschulen oder aus Gymnasien bezeichnet werden. Die beiden Schultypen korrespondieren mit den ihnen entsprechenden Altersklassen. Die für die Anwendung des SDQ geeigneten Altersklassen finden sich vornehmlich in Gymnasien. Das bestimmte die Anwendung des SDQ in der Untersuchung 2034GR. Damit aber reduziert sich die mit dem SDQ untersuchte Anzahl von Kindern gegenüber der Gesamtstichprobe auf eine SDQ-Teilstichprobe mit $n = 1029$ Jungen und Mädchen.

Die Tabelle 11.3 zeigt die Verteilung der von den beteiligten Jungen und Mädchen erreichten Gesamtpunktwerte in der SDQ-Gesamtskala „Total Difficulties“. Die in dieser Tabelle dargestellten Werte beziehen sich auf die Antworten derjenigen Gymnasiasten ($n = 1029$), die auf alle 20 Fragen der 4 SDQ-Subskalen, mit denen die Gesamtskala „Total difficulties“ gebildet wird, vollständig geantwortet haben.

Tabelle 11.3: Verteilung der Gesamtwerte aus der SDQ-Skala „Total Difficulties“ („Td“-GPW)

„Td“-GPW	Frequenz	Prozent	Kum. Prozent	„Td“-GPW	Frequenz	Prozent	Kum. Prozent
0	7	,7	,7	16	44	4,3	82,6
1	11	1,1	1,7	17	34	3,3	85,9
2	30	2,9	4,7	18	22	2,1	88,0
3	40	3,9	8,6	19	29	2,8	90,9
4	52	5,1	13,6	20	22	2,1	93,0
5	45	4,4	18,0	21	20	1,9	94,9
6	54	5,2	23,2	22	14	1,4	96,3
7	62	6,0	29,3	23	16	1,6	97,9
8	61	5,9	35,2	24	6	,6	98,4
9	71	6,9	42,1	25	5	,5	98,9
10	63	6,1	48,2	26	3	,3	99,2
11	64	6,2	54,4	27	1	,1	99,3
12	59	5,7	60,2	28	2	,2	99,5
13	72	7,0	67,2	29	3	,3	99,8
14	59	5,7	72,9	30	1	,1	99,9
15	56	5,4	78,3	34	1	,1	100,0
				Gesamt	1029	100,0	

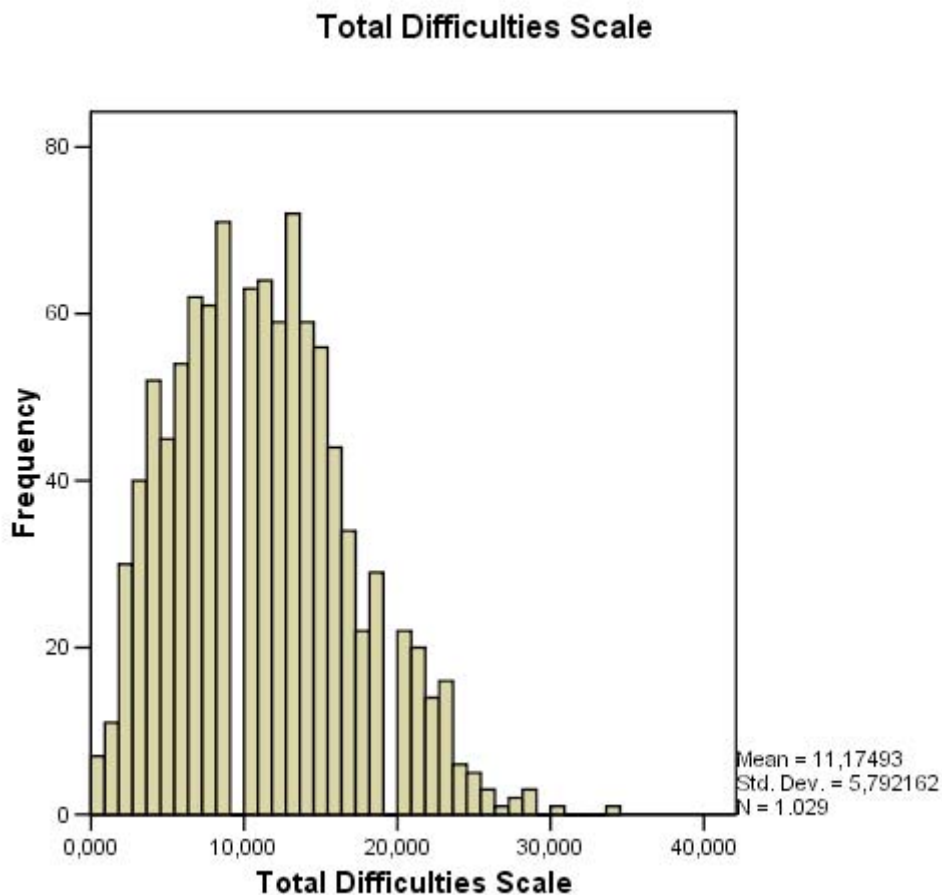
Zeichenerklärung: „Td“-GPW= „Total difficulties“- Gesamtpunktwerte, Kum. Prozent = Kumulierter Prozentsatz

Zur Häufigkeitsverteilung der SDQ-Gesamtwerte gehören die nachstehend aufgeführten Kennwerte.

Mean	11,17493	Variance	33,549	Skewness	,449
Std. Error of Mean	,180565	Range	34,000	Std. Error of Skewness	,076
Median	11,00000	Minimum	,000	Kurtosis	-,087
Mode	13,000	Maximum	34,000	Std. Error of Kurtosis	,152
Std. Deviation	5,792162	Sum	11499,000		

Die nachfolgende Abbildung veranschaulicht die in Tabelle 11.3 dargestellte Verteilung der „Total difficulties“-Gesamtpunktwerte.

Abbildung 11.1: Verteilung der SDQ Gesamtpunktwerte



Beachtenswert in der oben stehenden Abbildung sind die zwei auffälligen Spitzen in der dargestellten Kurve der Gesamtpunktwerte aus der Gesamtskala des SDQ „Total difficulties“. Die Ergebnisse aus der Durchführung des Kolmogorov-Smirnov Tests bestätigen die Hypothese einer signifikanten Abweichung von der Normalverteilung (Kolmogorov-Smirnov $Z = 2,154$; $p = 0,000$). Die nach Pearson berechnete positive Schiefe (Skewness = ,449) spricht für eine linkssteile Asymmetrie der dargestellten Verteilung. Problematisch ist in diesem Zusammenhang die Tatsache, dass die für die Bezeichnung der Verteilung geeignete statistische Kennwerte (arithmetisches Mittel, Medianwert und Modalwert) nicht genau so positioniert sind, wie es für den Fall der linksteilen Verteilung nach Bortz (1999, 5. Aufl., S. 46) angegeben ist.

Mit der nach der Durchführung des Kolmogorov-Smirnov Tests zu verteidigenden signifikanten Abweichung der Verteilung von der Normalverteilung ergibt sich für die griechische Anwendung eine andere Verteilungsform als von anderen Autoren beschrieben: „...that total difficulties score was approximately normally distributed in both community and clinical samples, with the community sample being shifted 1.4 standard deviations to the right. The mean was 11.4 (5.2) for the community sample and 18.6 (6.1) for the clinic sample. (Goodman, Meltzer & Bailey, 1998, S. 127).

Der Sachverhalt war festzustellen und ist entsprechend zu vermerken.

11.1 AS: Die Verteilung der SDQ-Gesamtwerte in der Untersuchung 2034GR weicht signifikant von einer Normalverteilung ab.

11.2 AS: Die theoretisch verfügbare Variationsbreite der zu erreichenden Gesamtpunktwerte in der SDQ-Gesamtskala liegt zwischen 0 und 40 Gesamtpunkten. Die von den beteiligten Kindern erbrachten individuellen Gesamtpunktwerte nehmen eine Breite von 0 bis zu 34 Gesamtpunkten in Anspruch, wobei die Mehrheit der Kinder (99,9%) mit ihren individuellen Werten nur bis zu 30 Punkten variieren. Der höchste Wert (Score 34) ist nur von einem Kind erreicht worden.

11.3 AS: Über 75% der Angehörigen dieser Stichprobe verteilen sich bis und mit dem Punktwert 15 auf der unteren Hälfte des in Anspruch genommenen Skalenabschnittes.

11.4 AS: Annähernd 95% der Angehörigen dieser Stichprobe sind mit ihrem individuellen Gesamtpunktwert auf etwa 70% des insgesamt in Anspruch genommenen Skalenteils platziert.

11.5 AS: Auf die von und mit 0 gezählten ersten 14 Werte der von den Jungen und Mädchen in Anspruch genommenen Gesamtpunktwerte entfallen 67,2% der SDQ-Stichprobe, auf die letzten 14 Punktwerte dem gegenüber nur 5,1%. Die Angaben veranschaulichen die zu den höheren Werten hin unter den Kurvenverlauf der Normalverteilung abflachenden SDQ Gesamtpunktwerte.

11.6 AS: Die Verteilung der SDQ-Gesamtpunktwerte in dieser Stichprobe lässt vermuten, dass höheren Punktwerten und den Abständen zwischen ihnen auch ein höheres Gewicht für die mit der Abweichung vom Mittelwert begründeten klinischen Aussagen zuzuschreiben ist.

Aus der Durchführung des SDQ sind nun nicht nur die individuellen Gesamtpunktwerte auf der Skala „Total Difficulties“ („Td“-GPW) bekannt geworden, sondern mit ihnen auch die der im SDQ verwendeten Subskalen. Auf die Ergebnisse aus diesen Skalen und auf einige Aspekte der Beziehungen zwischen ihnen, so wie sie sich nach der Anwendung des SDQ in der Untersuchung 2034 darstellen, wird im Folgenden eingegangen.

11.4 Die Interkorrelationen zwischen den SDQ Subskalen in den Daten der Stichprobe 2034GR

In den nachstehenden Ausführungen werden die aus der Durchführung des SDQ in der Untersuchung 2034GR bekannt gewordenen Ergebnisse und Teilergebnisse unter dem Aspekt der Zusammenhänge zwischen den mit dem SDQ eingeführten und benutzten Konstrukte betrachtet. Als Konstrukte werden die Begriffe verstanden, die von Goodman mit der von ihm entwickelten Theorie dieses Tests eingeführt und untereinander in Beziehung gesetzt werden. Die Operationalisierungen dieser Konstrukte sind die Skalen. Sie definieren über die zu den Skalen gehörenden Fragen die von den Skalen repräsentierten Konstrukte. Indem sie in diesem Test, so etwa mit dem Namen der Subskalen, auch verbale Repräsentanten haben, entsteht eine definitorische Spannung zwischen den vom Namen der Skalen gesetzten Konstrukt und den mit den Fragen der entsprechenden Skala repräsentierten Inhalten. Letztlich ist diese definitorische Spannung als Validitätsproblem zu begreifen. In Frage steht beispielsweise ob die Skala

„Prosocial behaviour“ mit den über sie erhobenen Werten einen Indikator für das liefert, was unter „Prosocial behaviour“ verstanden wird, oder ob der entsprechende Wert die Antwort auf fünf Testitems repräsentiert. Indem der erstgenannte Fall bedacht wird, bleibt zunächst offen, Was alles „Prosocial behaviour“ umfasst. Eine denkbare Interpretation wäre auch die, nach der „Prosocial behaviour“ im Sinne dieses Tests nichts weiter ist als eine Sammelbezeichnung für eine auf fünf Fragen und Antworten beruhende Ministichprobe an Selbstauskünften über „Prosocial behaviour“ als winzige Teilmenge eines unübersehbaren Pools an Selbstauskünften über prosoziales Verhalten. Für die Anwender des Tests ist die Beantwortung der Frage wesentlich, welches Verständnis sie benutzen sollen, um eine optimale Nützlichkeit des Verfahrens erreichen zu können. Dieser Problematik wird hier nicht nachgegangen. Wohl aber soll geprüft werden, welche Relationen nach den Ergebnissen aus der erfolgten Anwendung zwischen den testinternen Konstrukten erkennbar sind.

Die dabei gewonnenen Informationen könnten Hinweise auf die psychometrischen Eigenschaften des Verfahrens und dessen innere Struktur geben, zumindest so wie die sich nach dessen Einsatz in der vorliegenden Stichprobe von griechischen Kindern aus Gymnasien zeigen. Andererseits besteht dadurch auch die Möglichkeit, Im übrigen lassen sich so Vergleiche mit entsprechenden Ergebnissen aus Studien mit Stichproben außerhalb Griechenlands fundieren, was zur begonnenen Diskussion über die interkulturelle Stabilität des Verfahrens beitragen könnte.

Empirisch begründete Angaben über Zusammenhänge zwischen den SDQ-Subskalen sind unter anderem in den Veröffentlichungen von Muris et al., 2003, sowie 2004 zu finden.

In der Studie von Muris et al. (2003), die mit 562 Schulkindern zwischen 9 und 15 Jahren in den Niederland durchgeführt wurde, ergaben sich zwischen den SDQ-Subskalen Korrelationen, die zwischen 0,19 und 0,40 lagen. Nach Meinung der Autoren dieser Studie erweist sich damit: „... that SDQ subscales tap relatively independent domains of difficulties...“ (S. 4). Die höchste Korrelationen wurden zwischen den Subskalen „Emotional problems“ und „Peer problems“ ($r = 0,40$) und zwischen den Skalen „Conduct problems“ und „Hyperactivity“ ($r = 0,38$) gefunden. Die niedrigsten Korrelationen bestanden zwischen den Subskalen „Peer problems“ und „Hyperactivity“ ($r = 0,19$) wie auch zwischen „Peer problems“ und „Conduct problems“ ($r = 0,20$). Die Skala „Prosocial behaviour“ hat mit allen Subskalen eine negative Korrelation aufgewiesen, die stärkste unter denen mit der Subskala „Conduct problems“ ($r = -0,27$). Alle 5 Subskalen korrelierten in dieser Studie höher mit der Gesamtskala „Total difficulties“ als untereinander.

Muris et al. (2003) beschreiben über die von ihnen errechneten Korrelationen eine Struktur des SDQ, der ihn als ein Instrument erscheinen lässt, das Informationen aus recht unterschiedlichen Bereichen der Selbstwahrnehmung aufnimmt und die zu einer Gesamtskala vereint, die aus gleichsinnig verrechneten Antworten spezifische Punktwerte bildet, deren Validität aus dieser Strukturbeschreibung nicht verlässlich bestimmbar ist. Allerdings ergibt sich so das Denkmodell, über bekannt werdende Gesamtscores könnten, möglicherweise faktorenanalytisch, die aus den Subskalen aufsummierten Eigenheiten (Faktoren) erfasst werden.

In hoher Maße übereinstimmende Ergebnisse berichten auch van Widenfelt et al. (2003) und Rønning et al. (2004). Die höchste Korrelation in der Studien von van Widenfelt wurde zwischen den Subskalen „Conduct problems“ und „Hyperactivity“ ($r = 0,40$) gefunden und die niedrigste zwischen „Emotional problems“ und „Prosocial behaviour“ ($r = 0,07$). In der Studie von Rønning et al. (2004) lagen die Korrelationen ähnlich wie bei der Studie von Muris (s.o.) zwischen den 4 „pathologischen“ SDQ-Subskalen zwischen 0,17 und 0,49 (r conduct problems and hyperactivity = 0,49; r emotional-peer = 0,40; r conduct-peer = 0,29; r hyperactivity-emotional = 0,29; r conduct-emotional = 0,24; r peer-hyperactivity = 0,17).

Einige dieser Korrelationen, so beispielsweise die zwischen den Subskalen „Conduct problems“ und „Hyperactivity“ ($r = 0,40$) lassen sich widerspruchsfrei als theoriekonform interpretieren. Im Beispielfall wären vermehrt auftretende Probleme für hyperaktiv belastete Jungen und Mädchen erwartbar. Auch in der Studie 2034GR wurden niedrige negative Korrelationen zwischen der Skala „Prosocial behaviour“ und den übrigen 4 Subskalen gefunden, die zwischen $-0,15$ und $-0,32$ lagen.

Aus Tabelle 11.4 wird die Höhe der Interkorrelationen der SDQ-Subskalen innerhalb der griechischen Stichprobe, über die hier berichtet wird, ersichtlich. Die dargestellten Korrelationen aus der Untersuchung 2034GR beruhen auf individuellen Daten von Jungen und Mädchen aus griechischen Gymnasialklassen.

Tabelle 11.4: Interkorrelationen zwischen den 5 Subskalen des SDQ und der SDQ-Gesamtskala

	Emotional symptoms	Conduct problems	Hyperactivity	Peer problems	Prosocial behaviour
Conduct problems	,269 **				
Hyperactivity	,331 **	,444 **			
Peer problems	,409 **	,323 **	,238 **		
Prosocial behaviour	,061 *	-,281 **	-,223 **	-,219 **	
Total Difficulties	,749 **	,681 **	,713 **	,683 **	-,212 **

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

Die in dieser Tabelle dargestellten Ergebnisse ergeben sich aus der Durchführung einer speziellen griechischen Untersuchung. Inwieweit solche Ergebnisse zu verallgemeinern wären, bedürfte der intensiven Prüfung. Immerhin könnten aber die beobachteten strukturellen Besonderheiten des SDQ zur Begründung von Hypothesen benutzt werden. Die explizite Formulierung von Mutmaßungen eröffnete immerhin die Möglichkeit der empirischen Überprüfung. Das gilt dann so auch für das nachfolgende Beispiel.

In der inneren Struktur des SDQ fällt auf, dass ein in diesem Test als „Prosocial behaviour“ bezeichnetes Antwortmuster in den Selbstberichten nur mit $r = 0,061^*$ zu „Emotional symptoms“ in Beziehung steht und mit $r = -0,281^{**}$ zu „Conduct problems“ und weiterhin mit $r = -0,223^{**}$ zu „Hyperactivity“ und $r = -0,219^{**}$ zu „Peer problems“. Dann würde die nachstehende Vermutung die beobachteten Korrelationen in einen relativ widerspruchsfreien Zusammenhang bringen. „Das vom SDQ gemessene selbstberichtete „Prosocial behaviour“ ist mit höherer Wahrscheinlichkeit ein Indikator für die individuelle Abwesenheit von „Emotional symptoms“, von „Conduct problems“, „Hyperactivity“ und „Peer problems“.

Diese Vermutung geht über das hinaus, was die Korrelationen anzeigen. Das Ausbleiben der Korrelationen ist lediglich eine Auffälligkeit, die Interpretationen anregt. Der dargestellte Zusammenhang ließe sich aber unter Akzentuierung der aufgenommenen Selbstzuschreibungen begründbar konstruieren. Obschon die von den einzelnen Teilnehmern gegebenen vielen Antworten von den Jungen und Mädchen nicht wirklich auf innere Konsistenz, auf Übereinstimmung von Frage zu Frage geprüft werden, ist doch eine relative Widerspruchsfreiheit zu erwarten. Kognitive Dissonanzen entstünden, wenn einerseits prosoziales Verhalten behauptet wird und andererseits Verhaltensprobleme oder Streitigkeiten mit den Gleichaltrigen zugegeben werden müssten. Da auch Hyperaktive permanent der Gefahr unterliegen, in soziale Konflikte zu kommen, ließen sich auch Aussagen über eigene Hyperaktivität potentiell in die geäußerte

Vermutung einbeziehen. Der selbstgegebene Hinweis auf eigene emotionale Probleme könnte demgegenüber relativ widerspruchsfrei zusammen mit Angaben über eigenes prosoziales Verhalten auftreten.

Verkürzt ließe sich das zu der folgenden Hypothese zusammenfassen: „Selbstberichtetes prosoziales Verhalten im Sinne des SDQ ist relativ inkompatibel mit Hinweisen auf selbstberichtete „Conduct problems“, „Hyperactivity“ und „Peer problems“ im Sinne des SDQ.

Mit einer solchen Hypothese würde dem Aspekt der Selbstangaben und deren relativer Widerspruchsfreiheit eine sehr spezielle Funktion bei der subjektiven Beantwortung der SDQ Items zugeordnet werden. An dieser Stelle dienen diese Spekulationen zunächst nur der Verdeutlichung von Gründen, die auch für die analytische Befassung mit der inneren Struktur des SDQ sprechen könnten. Solche, in eine Theorie des SDQ eingreifende Überlegung könnte auf Faktorenanalysen beruhen, die als Hypothesen generierende statistische Verfahren verstanden werden. Dem ist hier nicht so, denn das aus der Tabelle der Interkorrelationen entnommene Beispiel zeigt lediglich eine denkbare analytische Perspektive und verdeutlicht diese eine Sichtweise. Nun wird nachfolgend dennoch, den unterliegenden empirischen Daten exemplarisch nachgegangen. Das folgt, ebenso exemplarisch, der Frage, ob solche Details wirklich beachtenswert sein könnten.

Die Datenbasis besteht aus den SDQ Daten, die mit dessen Anwendung in der Untersuchung 2034GR aufgenommen worden sind. Vermutet wird die nachstehend formulierte Beziehung zwischen zwei Subskalen des SDQ.

Hypothese (pros. Hy.): „Die Anwesenheit selbst berichteten prosozialen Verhaltens vermehrt in Gruppen die Wahrscheinlichkeit der Abwesenheit von selbst berichteter Hyperaktivität.“

Die Überprüfung der Hypothese erfolgt nonparametrisch. Die jeweiligen Skalendaten sind gleichsinnig auf drei aufsteigende Wertebereiche, auf die statistischen Klassen 1, 2 und 3 verteilt worden.

Tabelle 11.5: Kreuztabelle der in 2034GR beobachteten Verteilung der Werte aus den SDQ Subskalen Prosoziales Verhalten und Hyperaktivität (jeweils in 3 Klassen zusammengefasst)

		ProSoz				
		1,00	2,00	3,00	Gesamt	
Hyper	1,00	Anzahl	14	167	389	570
		Erwartete Anzahl	16,6	197,2	356,3	570,0
		% von Hyper	2,5%	29,3%	68,2%	100,0%
		% von ProSoz	46,7%	46,8%	60,3%	55,2%
		% der Gesamtzahl	1,4%	16,2%	37,7%	55,2%
	2,00	Anzahl	14	177	242	433
		Erwartete Anzahl	12,6	149,8	270,6	433,0
		% von Hyper	3,2%	40,9%	55,9%	100,0%
		% von ProSoz	46,7%	49,6%	37,5%	42,0%
		% der Gesamtzahl	1,4%	17,2%	23,4%	42,0%
	3,00	Anzahl	2	13	14	29
		Erwartete Anzahl	,8	10,0	18,1	29,0
		% von Hyper	6,9%	44,8%	48,3%	100,0%
		% von ProSoz	60,7%	3,6%	2,2%	2,8%
		% der Gesamtzahl	,2%	1,3%	1,4%	2,8%
		Korrigierte Residuen	1,3	1,2	-1,6	

(Fortführung Tabelle 11.5)

		ProSoz			1032
		1,00	2,00	3,00	Gesamt
Gesamt	Anzahl	30	357	645	1032
	Erwartete Anzahl	30,0	357,0	645,0	1032,0
	% von Hyper	2,9%	34,6%	62,5%	100,0%
	% von ProSoz	100,0%	100,0%	100,0%	100,0%
	% der Gesamtzahl	2,9%	34,6%	62,5%	100,0%

Zeichenerklärung: ProSoz = SDQ Subskala prosoziales Verhalten; Hyper = SDQ Subskala Hyperaktivität; 1, 2, 3: gleichsinnige Werteklassen, aufsteigend geordnet.

Chi-Quadrat-Tests

	Wert	Df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	19,563(a)	4	,001
Likelihood-Quotient	19,144	4	,001
Zusammenhang linear-mit-linear	17,468	1	,000
Anzahl der gültigen Fälle	1032		

Die Gesamttabelle ist signifikant. Demnach dürfen die einzelnen Felder zur Bildung von feldspezifischen Aussagen verwendet werden:

11.7 AS: Der Anteil von Kindern mit hohem selbstberichteten prosozialem Verhalten (SDQ) nimmt mit ansteigendem selbst berichteten hyperaktiven Verhalten (SDQ) ab.

11.8 AS: Der Anteil sinkt in drei aufsteigenden Klassen selbst berichteter Hyperaktivität von 68,2% der Teilgruppe über 55,9% auf 48,3%.

11.9 AS: In den unterschiedenen drei Teilgruppen fällt die Tendenz, über prosoziales Verhalten zu berichten, mit der Zunahme von auf drei Werteklassen verteilten Angaben zur eigenen Hyperaktivität um etwa 20%.

Die auf die SDQ Subskala „Hyperactivity“ bezogene Hypothese wird im Folgenden sinngemäß auf zwei weitere SDQ Subskalen übertragen. Danach ist zu erwarten, dass Selbstberichte über prosoziales Verhalten im Sinne des SDQ sich als relativ inkompatibel mit selbst abgegebenen Hinweisen auf eigene Verhaltensprobleme, „Conduct problems“ und Problemen mit Gleichaltrigen, „Peer problems“ im Sinne des SDQ erweisen werden. Die Prüfung dieser beiden Annahmen erfolgt mit den gleichen Arbeitsschritten wie sie bei der eben zuvor überprüften Subskala angewendet worden sind.

Tabelle 11.6: Kreuztabelle der in 2034GR beobachteten Verteilung der Werte aus den SDQ Subskalen Prosoziales Verhalten und „Conduct problems“ (in drei statistischen Klassen zusammengefasst)

			ProSoz			Gesamt	
			1,00	2,00	3,00		
Con	1,00	Anzahl	12	209	488	709	
		Erwartete Anzahl	20,0	245,7	443,3	709,0	
		% von Con	1,7%	29,5%	68,8%	100,0%	
		% von ProSoz	41,4%	58,5%	75,8%	68,8%	
		% der Gesamtzahl	1,2%	20,3%	47,4%	68,8%	
			Korrigierte Residuen	-3,2	-5,2	6,2	
	2,00	Anzahl	14	142	154	310	
		Erwartete Anzahl	8,7	107,4	193,8	310,0	
		% von Con	4,5%	45,8%	49,7%	100,0%	
		% von ProSoz	48,3%	39,8%	23,9%	30,1%	
		% der Gesamtzahl	1,4%	13,8%	15,0%	30,1%	
			Korrigierte Residuen	2,2	4,9	-5,6	
	3,00	Anzahl	3	6	2	11	
		Erwartete Anzahl	,3	3,8	6,9	11,0	
		% von Con	27,3%	54,5%	18,2%	100,0%	
% von ProSoz		10,3%	1,7%	,3%	1,1%		
% der Gesamtzahl		,3%	,6%	,2%	1,1%		
		Korrigierte Residuen	4,9	1,4	-3,1		
Gesamt	Anzahl	29	357	644	1030		
	Erwartete Anzahl	29,0	357,0	644,0	1030,0		
	% von Con	2,8%	34,7%	62,5%	100,0%		
	% von ProSoz	100,0%	100,0%	100,0%	100,0%		
	% der Gesamtzahl	2,8%	34,7%	62,5%	100,0%		

Zeichenerklärung: ProSoz = SDQ Subskala prosoziales Verhalten; Con = SDQ Subskala Conduct Problems; 1, 2, 3: gleichsinnige Werteklassen, aufsteigend geordnet.

Chi-Quadrat-Test

	Wert	Df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	63,739(a)	4	,000
Likelihood-Quotient	49,565	4	,000
Zusammenhang linear-mit-linear	49,430	1	,000
Anzahl der gültigen Fälle	1030		

Auch diese Gesamttabelle ist signifikant. Danach sind die folgenden Feststellungen zu vermerken:

11.10 AS: Die Anwesenheit selbst berichteten prosozialen Verhaltens vermehrt in Gruppen die Wahrscheinlichkeit der Abwesenheit von selbst berichteten Verhaltensproblemen.

11.11 AS: Der Anteil von Kindern mit hohem selbstberichteten prosozialen Verhalten (SDQ) nimmt mit ansteigendem selbst berichteten Verhaltensproblemen (SDQ) ab.

11.12 AS: Der Anteil sinkt in drei aufsteigenden statistischen Klassen selbst berichteter Verhaltensprobleme von 68,8% der Teilgruppe über 49,7% auf 18,2%.

11.13 AS: In Teilgruppen fällt die Tendenz über prosoziales Verhalten zu berichten, mit der Zunahme von Angaben auf einer dreistufigen Skala zu eigenen Verhaltensproblemen um etwa 40%.

Tabelle 11.7: Kreuztabelle der in 2034GR beobachteten Verteilung der Werte aus den SDQ Subskalen Prosoziales Verhalten und „Peer problems“ (in Klassen zusammengefasst)

			ProSoz			Gesamt
			1,00	2,00	3,00	
Peer	1,00	Anzahl	15	267	547	829
		Erwartete Anzahl	24,1	287,1	517,8	829,0
		% von Peer	1,8%	32,2%	66,0%	100,0%
		% von ProSoz	50,0%	74,8%	84,9%	80,4%
		% der Gesamtzahl	1,5%	25,9%	53,1%	80,4%
		Korrigierte Residuen	-4,3	-3,3	4,7	
	2,00	Anzahl	13	85	89	187
		Erwartete Anzahl	5,4	64,8	116,8	187,0
		% von Peer	7,0%	45,5%	47,6%	100,0%
		% von ProSoz	43,3%	23,8%	13,8%	18,1%
		% der Gesamtzahl	1,3%	8,2%	8,6%	18,1%
		Korrigierte Residuen	3,6	3,4	-4,6	
	3,00	Anzahl	2	5	8	15
		Erwartete Anzahl	,4	5,2	9,4	15,0
		% von Peer	13,3%	33,3%	53,3%	100,0%
% von ProSoz		6,7%	1,4%	1,2%	1,5%	
% der Gesamtzahl		,2%	,5%	,8%	1,5%	
	Korrigierte Residuen	2,4	-,1	-,7		
Gesamt	Anzahl	30	357	644	1031	
	Erwartete Anzahl	30,0	357,0	644,0	1031,0	
	% von Peer	2,9%	34,6%	62,5%	100,0%	
	% von ProSoz	100,0%	100,0%	100,0%	100,0%	
	% der Gesamtzahl	2,9%	34,6%	62,5%	100,0%	

Zeichenerklärung: ProSoz = SDQ Subskala prosoziales Verhalten; Peer = SDQ Subskala Peer Problems; 1, 2, 3: gleichsinnige Werteklassen, aufsteigend geordnet

Chi-Quadrat-Tests

	Wert	Df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	35,754(a)	4	,000
Likelihood-Quotient	30,724	4	,000
Zusammenhang linear-mit-linear	28,208	1	,000
Anzahl der gültigen Fälle	1031		

Da die Gesamttabelle sich als signifikant erwiesen hat, sind die folgenden Feststellungen entsprechend begründet.

11.14 AS: Mit der Zunahme von selbst berichteten prosozialem Verhalten vermehrt sich in Gruppen die Wahrscheinlichkeit der Abwesenheit von selbst berichteten Problemen mit Gleichaltrigen.

11.15 AS: Der Anteil von Kindern mit hohem selbst berichteten prosozialem Verhalten (SDQ) nimmt mit ansteigendem selbst berichteten Problemen mit Gleichaltrigen (SDQ) ab.

11.16 AS: Der Anteil variiert in drei aufsteigenden Klassen selbst berichteter Probleme mit Gleichaltrigen von 66% der Teilgruppe über 49,4% auf 53,3%.

Hinzugesetzt werden darf an dieser Stelle auch der Hinweis auf das erwartete Ausbleiben von Signifikanzen im Vergleich der Subskalen Prosoziales Verhalten (SDQ) und Emotionale Probleme.

Die exemplarisch analysierten Zusammenhänge in der Verteilung von Subskalen des SDQ öffnen die Perspektive auf die Beziehungen zwischen selbstberichteten Verhaltensweisen und der Selbstwahrnehmung des eigenen Verhaltens. Wenn auf mögliche Zusammenhänge zwischen den Selbstangaben und dem Auftreten kognitiver Dissonanzen verwiesen wurde, war weniger an Bemühungen der befragten Jungen und Mädchen um widerspruchsfreie Angaben gedacht worden, als viel mehr an die angestrebte Homogenität der Persönlichen Theorien der Jungen und Mädchen. Wenn das zutrifft, was S. Eppstein (1981) über das Interesse der jeweiligen Persönlichkeit an der Homogenität ihrer Persönlichen Theorien postulierte, dann ist die nach der jeweiligen Selbsttheorie zu vertretende Inkompatibilität mancher Verhaltensweisen sehr wahrscheinlich ein im Verhalten nachweisbares Phänomen. Die Selbstwahrnehmung der eigenen Persönlichkeit als prosozial ist dann beispielsweise relativ inkompatibel mit Hinweisen auf vermehrte Probleme mit Gleichaltrigen.

Obschon die bis hier dargestellten Analysen zum SDQ in gewissermaßen beiläufiger Wahrnehmung von Zusammenhängen zwischen Korrelationen einzelner Subskalen veranlasst wurden, eröffnen sie Einsichten in die innere Struktur der SDQ Messungen. Bei allen ansonsten feststellbaren Schwächen in den Beziehungen zwischen den Subskalen, verdeutlichen die mitgeteilten Analyseergebnisse inhaltsvalide Zusammenhänge in den selbst berichteten Einschätzungen der jeweils eigenen Situation. Dem wäre mit Aussicht auf Gewinn an gemessenen Informationen nachzugehen. Dazu werden die nachstehenden Feststellungen formuliert: 1. Selbstangaben zum prosozialem Verhalten erweisen sich im SDQ mit den empirischen Daten subjektiv als relativ inkompatibel mit Selbstangaben zum Vorliegen von Phänomenen auf anderen, zum prosozialem Verhalten dissonanten Skalen (s.o.). 2. Diese Inkompatibilität tritt als Gruppenphänomen auf und ist demnach tendenziell wirksam. Zugleich sind individuelle Abweichungen zu beobachten. Die kann als relative Widersprüchlichkeit der Selbstangaben bezeichnet werden. 3. Auf die bezieht sich die nachstehende Hypothese: Die exemplarisch nachgewiesene relative Widersprüchlichkeit von Selbstaussagen im SDQ ist sehr wahrscheinlich eine klinisch nutzbare Quelle relevanter Informationen über die jeweils betroffenen Kinder und Jugendlichen.

Für die den psychischen Störungsbildern nahe stehenden Subskalen des SDQ ist aus Tabelle 11.4 zu ersehen, dass deren Interkorrelationen zwischen r -Werten von 0,30 und 0,44 liegen. In Abschätzung der Bedeutung solcher Werte ist hervorzuheben, dass bereits weniger als 1/5 an Übereinstimmungen zwischen zwei Skalen ausreicht, um diese Werte hervorzubringen. Das nun wiederum dürfte theoriegeleiteten Erwartungen entsprechen. Die Subskalen sollen sensibel gegenüber unterschiedlichen Symptomen sein.

Die engsten Zusammenhänge bestehen zwischen den Subskalen „Conduct problems“ und „Hyperactivity“ ($r = 0,44$) sowie zwischen den Skalen „Emotional problems“ und „Peer problems“ ($r = 0,41$). Diese Ergebnisse stimmen mit den hier zuvor wiedergegebenen Daten aus anderen Studien überein. Weiterhin korrelieren diese 4 Subskalen alle jeweils stärker mit der Gesamtskala des Verfahrens „Total difficulties“ als untereinander, was so auch von Muris et al. (2003) berichtet wurde.

Insgesamt stimmen die aus der untersuchten griechischen Stichprobe gefundenen Korrelationen zwischen den SDQ-Testskalen untereinander mit bisher bekannt gewordenen Ergebnissen aus

nicht griechischen Stichproben überein. Diese Übereinstimmung spricht für eine von Anwendung zu Anwendung auch über kulturelle Grenzen hinweg bestehende Stabilität des Verfahrens in der Informationsaufnahme über die mit den Antworten auf die SDQ-Items erfassten konstruktvaliden Selbstzuschreibungen von Phänomenen.

Die Bedeutung der individuell erreichten SDQ –Gesamtpunktwerte für die Untersuchung 2034GR ist ohne Zuhilfenahme externer Kriterien nicht hinreichend bestimmbar.

Wenn den von Goodman (1997) gegebenen Hinweisen gefolgt wird, sind höhere SDQ-Werte in der definierten Gesamtskala ein klinisch relevantes Indiz für das potentielle Vorliegen einer klinisch beachtenswerten psychischen Störung. Die daraufhin zu lösende Problematik der Bestimmung von validen Trennwerten (cut offs) führt, entsprechend dem Forschungsstand, in ähnliche Probleme, wie sie bereits im Zusammenhang mit der Bestimmung von Trennwerten für den CDI und den DTK zu erörtern waren. Zu bedenken bleibt, ob die Anwendung des von Goodman et al. (1998) vorgeschlagenen Algorithmus die Nutzbarkeit der SDQ Gesamtwerte und deren Nützlichkeit fördert. Insgesamt ließe sich so erhoffen, dass über die SDQ Werte Jungen und Mädchen bekannt werden, die mit beachtenswerter Wahrscheinlichkeit als nach klinischen Maßstäben potentiell psychisch belastet gelten würden. Die doppelte Bedingung in diesem Satz – mit beachtenswerter Wahrscheinlichkeit ...potentiell klinisch belastet – geht nicht auf sprachliche Ungeschicklichkeit zurück, sie gibt vielmehr korrekt die Relativität dieser Aussagen wieder, die so lange Bestand haben wird, wie keine höheren Validitätswerte als die bislang bekannten genutzt werden können.

Wenn die bislang empirisch zu vertretende Validität als hinreichend genommen werden sollte, wäre der Nutzwert der SDQ Daten in der Untersuchung 2034GR der hier zuvor genannte. Ein gewisser Prozentsatz der Jungen und Mädchen mit höheren SDQ-Daten dürfte als potentiell psychisch belastet gelten, was in der Folge, falls solche Ressourcen vorhanden wären, mit entsprechender Zuwendung zu deren individuelle Problematik beantwortet werden könnte. Die Realisierung dieser Vorgehensweise ist in der Untersuchung 2034GR, wie auch in vielen anderen, wegen des Fehlens aller dafür erforderlichen Ressourcen nicht möglich. Wohl aber kann geprüft werden, ob die bekannt gewordenen SDQ Daten nutzbringend zu den Daten aus den anderen in dieser Untersuchung benutzten Messverfahren in Beziehung gesetzt werden könnten.

Mit der Durchführung der Untersuchung 2034GR sind schon zahlreiche andere klinisch relevante Informationen über die Jungen und Mädchen der untersuchten Stichprobe bekannt geworden. Indem solche Daten mit denen aus der Durchführung des SDQ in Verbindung gebracht werden, ergeben sich wechselseitig zusätzliche Aufschlüsse über den Informationswert der Daten. Das ist so, weil alle Daten eine bestimmte Gruppe an Persönlichkeiten betreffen, eben die Jungen und Mädchen der Stichprobe. Mit einer wachsenden Zahl von Daten wächst das sie betreffende Informationsnetz.

Da als ein Hauptziel der Untersuchung 2034GR die Aufnahme von Informationen über die mutmaßliche Verbreitung depressiver Störungen in der untersuchten Stichprobe griechischer Mädchen und Jungen bestimmt wurde, könnte zunächst auch geprüft werden, welche Aufschlüsse die Anwendung des SDQ dazu beitragen kann. Spiegelgleich dazu wird zu bestimmen versucht, welche Informationen über den SDQ mit der Anwendung des CDI und des DTK bekannt werden.

Eine theoriegeleitete Abschätzung der Relationen zwischen den Messinstrumenten, SDQ einerseits und CDI wie auch DTK andererseits, führt zur Annahme der nachstehenden Verhältnisse.

Da einerseits die höheren SDQ Gesamtpunktwerte ein Indiz für das potentielle Bestehen einer klinisch bedeutsamen Störung sein sollen (Goodman, 1997), andererseits aber der DTK nach

Rossmann (2005) und der CDI nach Kovacs (2004) mit ihren höheren Scores hoch relevante Indizien für das Bestehen depressiver Störungen sein sollen, müssten Jungen und Mädchen mit höheren SDQ Werten auch solche sein, die höhere CDI und DTK Werte auf sich vereinen.

Die abgegebene Aussage erkennt den Messinstrumenten nur eine oberbegriffliche Übereinstimmung zu, indem auf das allen drei Instrumenten zugeschriebene Erkennen klinisch bedeutsamer Störungen verwiesen wird. Zu bestimmen bleibt danach jeweils, welches Instrument über das wahrscheinliche Auftreten welcher Störung informiert.

Aus der zu beschreibenden Spezifität der Verfahren sind die Möglichkeiten für Informationsgewinne abzuleiten.

Der die Instrumente betreffende Forschungsstand zwingt wegen widersprüchlicher oder fehlender Erkenntnisse zur Relativierung der Angaben über die Messleistungen. Für jedes Verfahren können unterschiedliche Angaben heran gezogen werden.

Mit dem SDQ sollen nach Goodman über entsprechende Ausprägungen des SDQ Gesamtpunktwertes klinisch beachtenswerte Hinweise auf das Vorliegen von klinischen Störungen gewonnen werden. Die Skala der Gesamtpunkte vereint im Vergleich zu SDQ Subskalen die höchsten Messqualitäten auf sich. Die auftretenden Variationen der SDQ Gesamtpunkte bieten demnach voneinander unterscheidbare Hinweise auf das Auftreten psychischer Störungen. Sie haben generellen Hinweischarakter.

Demgegenüber wird entsprechenden Gesamtpunkten der Verfahren CDI und DTK eine depressionsspezifische Aussagekraft zugeschrieben. Die auftretenden unterschiedlichen Gesamtpunkte aus DTK und CDI geben inhaltlich unterscheidbare Hinweise auf die Wahrscheinlichkeit des Bestehens depressiver Störungen.

Wenn den genannten Verhältnissen gefolgt wird, sollten für alle Jungen und Mädchen, die nach DTK oder CDI hochwahrscheinlich von klinischen Ausprägungen depressiver Störungen betroffen sind, auch entsprechend hohe individuelle SDQ Gesamtpunkte vorliegen.

Jungen und Mädchen, die nach DTK oder CDI hochwahrscheinlich nicht durch klinische Ausprägungen depressiver Störungen belastet sind, dürfen, da ja durchaus anderen psychischen Störungen vorliegen können, als Gruppe auch einen Anteil höherer SDQ Gesamtpunkte auf sich ziehen. Eine Schätzung des Prozentsatzes dieses Anteils wird durch hohe Ungewissheit belastet. Der Prozentsatz sollte nicht über 40% liegen. Grundlage dieser Schätzung sind Aufsummierungen der Prävalenzraten unter Kindern und Jugendlichen verbreiteten psychischen Störungen.

Ob die empirisch nachweisbaren Beziehungen zwischen den Daten aus der Anwendung des CDI, des DTK und des SDQ den mehr oder minder zweifelsfrei begründeten Erwartungen entsprechen, ist den Datenanalysen zu entnehmen.

11.5 Die verfahrensinternen korrelativen Beziehung der SDQ Items

Bei den nachstehenden Analysen wird auf die Frage eingegangen, wie niedrig oder hoch die zu den einzelnen SDQ-Subskala gehörenden Items miteinander korrelieren. Anlass für diese Frage geben die von Rønning et al. (2004) gefundenen niedrigen Korrelationen zwischen einigen Items, ein Ergebnis, das die Autoren als mögliche Erklärung für die niedrige Reliabilitätswerte heran gezogen hatten. Die aus ihren Daten unter dem gesetzten Limit von $r = 0,20$ gefundenen Korrelationskoeffizienten diskutieren die Autoren als ein Indizien dafür, dass „...the items included were measuring somewhat different properties...“ (S. 80). Die problematisierten Werte fanden sich, für beide Geschlechter, in den Korrelationen zwischen den Items 11-23 und 11-14 in

der Skala „Peer problems“, zwischen 7-22, 7-18 und 7-12 in der Skala „Conduct problems“ und zwischen den Items 21-15 und 21-10 in der Skala „Hyperactivity“.

Wie Rønning et al. (ebd.) haben van Widenfelt et al. (2003) argumentiert. Die niedrigen Alpha-Werte würden auftreten weil „...one or two items do not fit with the rest of the scale...“ (S. 285). Solche Äußerungen basieren auf Annahmen der klassischen Testtheorie, nach denen die Hohe des Zusammenhanges zwischen zwei Testsfragen ein Index dafür ist, „...wieweit beide Aufgaben den gleichen oder einen verschiedenen Aspekt des untersuchten Persönlichkeitsmerkmal erfassen...“ (Lienert und Raatz, 1998; 6. Aufl., S. 99).

Tabelle 11.8: Korrelationen zwischen den Items der SDQ-Skala „Emotional symptoms“

„Emotional symptoms“	3	8	13	16	24
3. Ich habe oft Kopfschmerzen oder Bauchschmerzen, mir wird oft schlecht					
8. Ich mache mir häufig Sorgen	,302**				
13. Ich bin oft unglücklich oder niedergeschlagen; ich muss häufig weinen	,319**	,424**			
16. Neue Situationen machen mich nervös; ich verliere leicht das Selbstvertrauen	,219**	,288**	,362**		
24. Ich habe viele Ängste; ich fürchte mich leicht	,319**	,355**	,358**	,263**	
Gesamtscore „Emotional symptoms“	,618**	,714**	,729**	,644**	,672**

** Correlation is significant at the 0.01 level (2-tailed).

Die höchste Korrelation innerhalb der Subskala „Emotional symptoms“ (Alpha-Wert = 0,70) besteht zwischen den Selbstaussagen „Ich mache mir häufig Sorgen“ und „Ich bin oft unglücklich oder niedergeschlagen; ich muss häufig weinen“ (Items 8 und 13 entsprechend; $r = 0,42$). In dieser Skala fällt keine Korrelationen zwischen den dazugehörigen Fragen unter das Limit von $r = 0,20$. Die Korrelationen aller Items mit dem Gesamtscore dieser Skala sind hoch und liegen zwischen 0,62 und 0,73.

Tabelle 11.9: Korrelationen zwischen den Items der SDQ-Skala „Conduct problems“

„Conduct problems“	5	7	12	18	22
5. Ich werde leicht wütend; ich verliere oft meine Beherrschung					
7. Normalerweise tue ich, was man mir sagt ¹	,068*				
12. Ich schlage mich häufig; ich kann Andere zwingen zu tun, was ich will	,288**	,061*			
18. Andere behaupten oft, dass ich lüge oder mogele	,217**	,031	,170**		
22. Ich nehme Dinge, die mir nicht gehören (von zu Hause, in der Schule oder anderswo)	,113**	,011	,157**	,189**	
Gesamtscore „Conduct problems“	,673**	,444**	,619**	,571**	,428**

* Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed). ¹= umgepolte Frage

Zwischen den Fragen der Skala „Conduct problems“, die in der vorliegenden Untersuchung 2034GR einen Alpha-Wert von 0,43 aufweist, ergeben sich sehr niedrige Itemkorrelationen.

Tendenziell überschreiten das gesetzte Limit von $r = 0,20$ nur die Korrelationen zwischen dem Item 5 („ich werde leicht wütend; ich verliere oft meine Beherrschung“) und den Fragen 12 und 18 („ich schlage mich häufig; ich kann Andere zwingen zu tun, was ich will“ und „Andere behaupten oft, dass ich lüge oder mogele“ entsprechend). Alle anderen Zusammenhänge erweisen sich als sehr niedrig, vor allem diejenige, die zwischen der 7. Aussage („Normalerweise tue ich, was man mir sagt“) und den übrigen 4 Items dieser Testskala bestehen. Dieses Ergebnis stimmt mit den von Rønning und Mit. (2004, s.o.) mitgeteilten überein. Für das als problematisch erscheinende Item 7 ist zusätzlich anzumerken, dass es das einzige Item dieser Skala ist, das positiv formuliert wurde.

Tabelle 11.10: Korrelationen zwischen den Items der SDQ-Skala „Hyperactivity“

„Hyperactivity“	2	10	15	21	25
2. Ich bin oft unruhig, ich kann nicht lange stillsitzen					
10. Ich bin dauernd in Bewegung und zappelig	,366**				
15. Ich lasse mich leicht ablenken; ich finde es schwer, mich zu konzentrieren	,314**	,244**			
21. Ich denke nach, bevor ich handele ¹	,149**	,158**	,223**		
25. Was ich angefangen habe, mache ich zu Ende; ich kann mich lange genug konzentrieren ¹	,159**	,099**	,195**	,344**	
Gesamtscore „Hyperactivity“	,678**	,632**	,656**	,562**	,548**

** Correlation is significant at the 0.01 level (2-tailed). ¹= umgepolte Frage

Ebenfalls problematisch sind in hoher Masse die Korrelationen zwischen den Aussagen der Skala „Hyperactivity“. Aus der Tabelle 11.10 ist zu ersehen, dass sehr niedrige Korrelationen zwischen positiv (Items 21 und 25) und negativ (Items 2, 10, 15) abgefassten Aussagen bestehen. Zugleich finden sich zwischen den positiven Items untereinander und ebenso zwischen den negativen Items nur Zusammenhänge, die das Wertlimit von $r = 0,20$ überschreiten. Über ähnliche Beobachtungen, nämlich geringe Korrelationen zwischen unterschiedlich gepolten Formulierungen (positiv vs. negativ), berichteten auch Rønning et al. (2004).

Zum Nachdenken führt vor allem der niedrige Zusammenhang zwischen der Aussage 15 „Ich lasse mich leicht ablenken; ich finde es schwer, mich zu konzentrieren“ und der Aussage 25 „Was ich angefangen habe, mache ich zu Ende; ich kann mich lange genug konzentrieren“. Beide Äußerungen teilen anscheinend einen ähnlichen Sachverhalt mit, der jeweils die Konzentrationsfähigkeit thematisiert, einmal positiv und einmal negativ formuliert. Unterschiedlich ist jedoch der erste Teil der zwei Äußerungen. Dass jemand sich leicht ablenken lässt (Item 15), schließt aber nicht unbedingt aus, dass jemand - auch wenn mit vielen Schwierigkeiten - eine Aufgabe nicht zu Ende bringen kann (Item 25). Das ergibt sich zumindest aus den Antworten der beteiligten griechischen Kinder. Wenn es wirklich so wäre, würde es dann eventuell darauf hinweisen, dass die befragten Kinder eher auf den ersten Teil der hier diskutierten Äußerungen aufmerksam geworden sind. Die zitierten Ergebnisse stellen die für den SDQ angewendeten Prinzipien der Itemkonstruktion in Frage. Vermutet werden sollte, dass die Kombination zweier thematisch unterschiedlicher Itemhälften auf die Beantwortung dieser Items als Störvariable einwirkt. Der Sachverhalt korrespondiert mit den psychischen Abläufen, die zu der Interviewregel veranlassten, keinesfalls zwei Fragen gleichzeitig zu stellen. Die Annahme, durch eine Überarbeitung der Itemkonstruktion seien die messtechnischen Kennwerte des SDQ zu verbessern, erscheint als durchaus begründet. Insoweit ist dem Vorschlag von Rønning et al.

(2004) zur messtechnischen Verbesserung des Verfahrens SDQ zu folgen. Der steht in Übereinstimmung mit den in der weiteren Diskussion zur Reliabilität des SDQ zu findenden kritischen Anmerkungen (s.o. Abschnitt 11.1).

Tabelle 11.11: Korrelationen zwischen den Items der SDQ-Skala „Peer problems“

„Peer problems scale“	6	11	14	19	23
6. Ich bin meistens für mich alleine; ich beschäftige mich lieber mit mir selbst					
11. Ich habe einen oder mehrere gute Freunde oder Freundinnen ¹	,208**				
14. Im allgemeinen bin ich bei Gleichaltrigen beliebt ¹	,160**	,349**			
19. Ich werde oft anderen gehänselt oder schikaniert	,308**	,209**	,306**		
23. Ich komme besser mit Erwachsenen aus als mit Gleichaltrigen	,232**	,131**	,209**	,181**	
Gesamtscore „Peer problems“	,621**	,578**	,644**	,644**	,604**

** Correlation is significant at the 0.01 level (2-tailed). ¹= umgepolte Frage

Die bestehende Zusammenhänge zwischen den 5 Items der SDQ-Skala „Peer problems“ sind, wie in allen anderen bisher dargestellten Subskalen des Verfahrens, im Allgemeinen niedrig. Wieder sind Korrelationskoeffizienten zu finden, die unter dem gesetzten kritischen Wert von $r = 0,20$ fallen (Items 6-14, 23-11 und 23-19). Die Werte für diese drei Itempaare scheinen keiner gemeinsamen Variationsregel zu folgen.

Tabelle 11.12: Korrelationen zwischen den Items der SDQ-Skala „Prosocial behaviour“

„Prosocial behaviour“	1	4	9	17	20
1. Ich versuche, nett zu anderen Menschen zu sein, ihre Gefühle sind mir wichtig					
4. Ich teile normalerweise mit Anderen (z.B. Süßigkeiten, Spielzeug, Buntstifte)	,222**				
9. Ich bin hilfsbereit, wenn andere verletzt, krank oder traurig sind	,226**	,279**			
17 Ich bin nett zu jüngeren Kindern	,118**	,202**	,224**		
20. Ich helfe anderen oft freiwillig (Eltern, Lehrern oder Gleichaltrigen)	,253**	,318**	,404**	,271**	
Gesamtscore „Prosocial behaviour“	,581**	,668**	,648**	,580**	,687**

** Correlation is significant at the 0.01 level (2-tailed).

Schließlich ergeben sich auch für die Testskala „Prosocial behaviour“ zu problematisierende Korrelationen. Hier ist die zwischen den Items 1 und 17 niedriger als der Limitwert (Tabelle 11.12). Die Korrelationen einer jeder Aussage mit der Gesamtscore der Skala sind im Übrigen aber - wie auch schon bei allen anderen Testskalen - befriedigend bis gut.

11.6 Die empirisch gesicherten Beziehungen zwischen SDQ, DTK und CDI in der Untersuchung 2034GR: Aspekte der konvergenten Validität

Die Nachprüfung des Grades der Übereinstimmung zwischen zwei oder auch mehreren Messverfahren ist verbreitete Praxis. Wenn für unterschiedliche Messverfahren eine übereinstimmende Spezifität ihrer Messleistungen behauptet wird, ist der Vergleich der mit ihnen an übereinstimmenden Stichproben gewonnenen Daten ein entsprechend begründbare Analyseschritt. Weitgehende Übereinstimmung von an identischen Stichproben gewonnenen Messergebnissen bietet Argumente für die den jeweiligen Test betreffende Theorie. In diesem Zusammenhang wird der Begriff der konvergenten Validität benutzt. Zwar ist die Validierung des einen Tests an einem vergleichbaren anderen Test nur ein Nachweis für die Übereinstimmung der verglichenen Verfahren, entbehrt aber nicht der Nützlichkeit und hilft zumindest, das Fehlen von Möglichkeiten zur eindeutigeren Validierung des Verfahrens zu relativieren. Konvergente Validierung ist insoweit pragmatisch, und dies in Absetzung gegenüber der Konstruktvalidität. „Konstruktvalidierung ist weniger pragmatisch als theoretisch orientiert. Ihre Bedeutung liegt daher nicht in einer unmittelbaren praktisch-diagnostischen Verwendbarkeit, sondern in der theoretischen Klärung dessen, was der betreffende Test misst“ (Lienert und Raatz, 1998; 6. Aufl., S. 266).

Indem nachfolgend betrachtet wird, ob die an einer identischen Stichprobe über CDI, DTK und SDQ gewonnene Daten übereinstimmen, ergibt das zunächst auch nur Antworten auf diese Frage. Welche Nutzbarkeit dem zukommt, bliebe danach zu prüfen. Zu beachten ist die Datenbasis der in der nachstehenden Tabelle mitgeteilten Korrelationen. Die beziehen sich auf eine Teilstichprobe innerhalb der untersuchten Stichprobe, nämlich der Jungen und Mädchen aus Klassen an den in die Untersuchung einbezogenen griechischen Gymnasien. Diese Teilstichprobe ist damit zugleich eine Teilstichprobe der älteren Jungen und Mädchen.

Tabelle 11.13: Korrelationen der SDQ-Skalen mit denen des DTK und des CDI innerhalb der Untersuchung 2034GR

SDQ	DTK				CDI	
	Skala 1: Dysphorie	Skala 2: Agitiertes Verhalten	Skala 3: Müdigkeit	Skala 1 + 2	CDI - „Schulbesuch“	CDI – 26 Items
“Total Difficulties”	,631 **	,513 **	,526 **	,710 **	,715 **	,714 **
“Emotion symptoms”	,653 **	,186 **	,515 **	,570 **	,555 **	,563 **
“Conduct problems”	,323 **	,518 **	,282 **	,487 **	,458 **	,445 **
“Hyperactivity”	,323 **	,542 **	,331 **	,499 **	,475 **	,466 **
“Peer problems”	,437 **	,247 **	,324 **	,439 **	,529 **	,535 **
“prosocial behaviour”	,046	-,311 **	-,082 **	-,115 **	-,219 **	-,203 **

** Correlation is significant at the 0.01 level (2-tailed).

Weil die in Tabelle 11.13 dargestellten Korrelationen zwischen dem SDQ und DTK und CDI auf den Antworten der Gymnasiasten in dieser Untersuchung beruhen, könnte eine Tabelle mit Werten von Kindern aus der Grundschule andere Werte enthalten. Das ergibt sich allein schon aus dem mit dem Geschlecht und dem Alter variierenden Anteilen an Kindern mit höheren DTK- und CDI-Gesamtpunktwerten.

Die Korrelationen zwischen den SDQ-Skalen und den DTK-Skalen bieten für beide Verfahren insgesamt Hinweise auf eine erwartbare Übereinstimmung der Verfahren. Die höchste Korrelation ($r = 0,71$) besteht zwischen den Gesamtskalen der zwei Verfahren, ein Ergebnis, das theoretisch zu erwarten war. Dies weil zum einen die Gesamtskalen beider Verfahren sowohl emotionale als auch verhaltensbezogene Symptome zu messen vermögen. Zum anderen würde die bekannt werdende Korrelation auf eine zwischen den Verfahren als bestehend geschätzte Übereinstimmung von etwa 50% verweisen. Das aber stünde nicht in Widerspruch zu den weiter oben begründeten Erwartungen. Denen würde eher eine höhere Korrelation zwischen den aufgeführten Skalen der verschiedenen Messinstrumente entgegen stehen.

Die Gesamtskala des SDQ „Total Difficulties“ korreliert darüber hinaus relativ hoch mit der DTK-Skala „Dysphorie-Selbstwert“. Im Sinne der nach Goodman zu interpretierenden speziellen Theorie zum SDQ ist diese Ausprägung der Korrelation als den Erwartungen entsprechend zu bezeichnen. Sie sollte keineswegs höher sein. Der SDQ gilt in dieser Betrachtung als ein Störungsbilder übergreifendes Messinstrument. Eine höhere oder sehr hohe Korrelation mit der entsprechenden Skala des DTK bedeutete eine im Sinne der Theorie über den Test unerwünschte Übereinstimmung mit der Erfassung von vor allem depressionsspezifischen Merkmalen.

Die SDQ-Skala „Emotion symptoms“ korreliert angemessen hoch mit den DTK-Konstrukten „Dysphorie-Selbstwert“ und „Müdigkeit/autonome Reaktionen“ (0,65 und 0,52). Das ist beachtenswert und bedenkenswert zugleich, weil von den insgesamt 5 SDQ-Fragen der letztgenannten Skala nur eine explizit nach somatischen Beschwerden fragt (Item 3). Dieser Befund verweist auf die für SDQ und DTK gleichermaßen erkennbaren und speziell auf die Subskalen zu beziehenden Validitätsprobleme. Das Auftreten somatischer Symptome innerhalb einer depressiven Störung ist vom Autor des DTK (Rossmann, 2005) mit einer eigens darauf bezogenen Skala in die Konstruktion seines Verfahrens einbezogen worden.

Die zwei SDQ-Skalen, die definitionsgemäß eher nach externalisierten Äußerungen der Befindlichkeit fragen, nämlich die Skalen „Conduct problems“ und „Hyperaktivität“ hängen innerhalb dieser Untersuchung stärker mit den in der zweiten DTK-Skala „Agitiertes Verhalten“ erreichten Gesamtwerten zusammen, was wiederum den in der jeweiligen Konstruktion beider Verfahren angelegten Merkmale entspricht.

Die SDQ-Skala „Peer problems“ weist ihre höchsten Korrelationen mit der DTK-Skala „Dysphorie/Selbstwert“ und der DTK-Gesamtskala auf ($r = 0,44$ in beiden Fällen). Die diese Skala bildenden Items sind inhaltlich, wie auch der Name dieser Skala ankündigt, mit dem individuellen sozialen Verhalten oder dessen subjektiver Wahrnehmung und der Bewertung sozialer Beziehungen befasst (6 *Ich bin meistens für mich alleine; ich beschäftige mich lieber mit mir selbst*; 11. *Ich habe einen oder mehrere gute Freunde oder Freundinnen*; 14. *Im Allgemeinen bin ich bei Gleichaltrigen beliebt*; 19. *Ich werde oft von anderen gehänselt oder schikaniert*; 23. *Ich komme besser mit Erwachsenen aus als mit Gleichaltrigen*). Eine auf Äußerungen zur Theorie über das jeweilige Messverfahren zurückgreifende Interpretation der empirisch gefundenen korrelativen Beziehungen vermag nicht mehr als eine bestenfalls plausible Übereinstimmung darzustellen. Wenn der gefolgt werden soll, dürfen die zwischen SDQ, DTK und CDI gefundenen Korrelationen als zumindest den theoriegeleiteten Erwartungen über die Beziehungen zwischen den Verfahren nicht entgegen stehend bewertet werden.

Im weitesten Sinne quantifizierbare Aufschlüsse über die Beziehungen zwischen SDQ und DTK sind zu gewinnen, wenn den weiter oben begründeten Erwartungen über die Beziehungen gefolgt wird. Danach sollten sich zunächst die nachstehend formulierten Erwartungen bestätigen lassen. Den schon gegebenen Begründungen folgend wird erwartet, dass

1. die Jungen und Mädchen, die wegen ihrer hohen DTK Gesamtpunkte aus Skala 1 + 2 bestimmungsgemäß der DTK Risikogruppe deprisk2 zuzurechnen waren, höhere SDQ Gesamtpunkte aufweisen als alle anderen Teilnehmer. Weiterhin wird angenommen,
2. die Jungen und Mädchen, die wegen ihrer hohen CDI Gesamtpunkte bestimmungsgemäß der CDI Risikogruppe deprisk2 zugerechnet wurden, werden höhere SDQ Gesamtpunkte aufweisen als alle anderen Teilnehmer.

Die zum Entscheid über das Eintreffen dieser Erwartungen einzusetzende Prüfung wird nachfolgend nonparametrisch über Kreuztabellen durchgeführt. Dazu wurden zunächst die SDQ Gesamtpunkte dichotomisiert. Dazu wurde der formulierten Erwartung gefolgt. Die in der Gesamtverteilung der SDQ Werte höheren Punktwerte sollten anderen Vorhersagen folgen als die vergleichsweise dann niedrigeren Werte. Dazu waren die erfassten Punktwerte an einem vorab zu definierenden Trennpunkt einerseits in eine auf 0 zu gehende und andererseits dann in eine auf den Höchstwert zu gehende Teilgruppe zu trennen. Der Trennpunkt wurde ohne Einbeziehung irgendwelcher Aufschlüsse über diagnostische Leistungen des SDQ und dann ausschließlich über eine Orientierung an der verbalisierten Relation festgelegt. Bestimmt wurde, dass die oberen 25% der Werte in der vorhandenen Werteliste die gegenüber dem größeren und auf 1 hin sinkenden Bereich höheren Werte umfassen. Die SDQ Gesamtpunkte von 1 bis 33 einschließlich wurden als 1 kodiert und die von 23 ab aufwärts vorhandenen Werte mit 2. Diese Kodierung war auch bei den DTK Gesamtpunkten aus DTK-Skala 1+2 angewendet worden. Der kodierte Wert 1 gibt für jede der beiden Skalen den unteren und der Wert 2 den oberen Bereich an. Die nachfolgende Tabelle verhilft mit den in ihr enthaltenen Daten zur Entscheidung darüber, ob die Erwartungen über die Relationen zwischen SDQ Gesamtpunktskala und DTK Gesamtpunkte aus Skala 1+ 2 erfüllen.

Tabelle 11.14: Kreuztabelle der dichotomisierten Gesamtpunkte aus SDQ und DTK (Skala 1+2)

			DTKdi		Gesamt	
			1,00	2,00		
SDQdi	1,00	Anzahl	605	154	759	
		Erwartete Anzahl	510,9	248,1	759,0	
		% von SDQdi	79,7%	20,3%	100,0%	
		% von DTKdi	87,2%	45,7%	73,6%	
		% der Gesamtzahl	58,7%	14,9%	73,6%	
			Korrigierte Residuen	14,2	-14,2	
	2,00	Anzahl	89	183	272	
		Erwartete Anzahl	183,1	88,9	272,0	
		% von SDQdi	32,7%	67,3%	100,0%	
		% von DTKdi	12,8%	54,3%	26,4%	
% der Gesamtzahl		8,6%	17,7%	26,4%		
		Korrigierte Residuen	-14,2	14,2		
Gesamt	Anzahl	694	337	1031		
	Erwartete Anzahl	694,0	337,0	1031,0		
	% von SDQdi	67,3%	32,7%	100,0%		
	% von DTKdi	100,0%	100,0%	100,0%		
	% der Gesamtzahl	67,3%	32,7%	100,0%		

Zeichenerklärung: SDQdi = SDQ Gesamtpunkte, dichotomisiert, 1 für unter, 2 für über 75%;
DTKdi = DTK Gesamtpunkte, dichotomisiert, 1 für Punktwert niedriger als 17, für gleich oder höher 17

Die Gesamttabelle ist hoch signifikant. Die Variation der Werte ist nicht als zufällig aufzufassen. Für alle Tabellenfelder sind auch die korrigierten standardisierten Residuen hoch signifikant. Daraus ergeben sich die folgenden Feststellungen:

11.17 AS: Die dichotomisierten Gesamtpunktwerte von SDQ und DTK (Skala 1+1) variieren hochsignifikant miteinander.

11.18 AS: Von den nach DTK als hochwahrscheinlich mit depressiven Störungen belasteten Jungen und Mädchen werden nur 54,3% auch nach SDQ Gesamtpunktwerten als wahrscheinlich von psychischen Störungen betroffen klassifiziert. Da das nach DTK anzunehmende Vorliegen depressiver Störungen auch den nach SDQ zu unterscheidenden Störungen zugehört, ergibt sich nur eine partielle Übereinstimmung von SDQ und DTK.

11.19 AS: Die nach DTK wahrscheinlich nicht von depressiven Störungen betroffenen Jungen und Mädchen werden auch nach SDQ Gesamtpunktwerten zu den insgesamt eher nicht mit psychischen Störungen belasteten Jungen und Mädchen zuzuordnen. Der SDQ weicht nur in 12,8% der vom DTK vollzogenen Klassifikation (Jungen und Mädchen nicht mit depressiven Störungen belastet) ab. Diese Abweichung ist anders zu bewerten als die unter 13.18 AS. Nach DTK wäre keine depressive Gefährdung anzunehmen. Der SDQ gibt Hinweise auf das Vorliegen einer, einer von mehreren möglichen, psychischen Störung ab. Beide Klassifikationen widersprechen einander nicht.

Nachfolgend wird in gleicher Weise geprüft, ob die für den DTK formulierten Erwartungen auf die Verteilung der nach CDI bestimmten zwei Risikogruppen übertragen werden können.

Tabelle 11.15: Kreuztabelle der dichotomisierten Gesamtpunkte aus SDQ und CDI

			CDIdi		Gesamt
			1,00	2,00	
SDQdi	1,00	Anzahl	707	52	759
		Erwartete Anzahl	635,3	123,7	759,0
		% von SDQdi	93,1%	6,9%	100,0%
		% von CDIdi	81,9%	31,0%	73,6%
		% der Gesamtzahl	68,6%	5,0%	73,6%
		Korrigierte Residuen	13,7	-13,7	
	2,00	Anzahl	156	116	272
		Erwartete Anzahl	227,7	44,3	272,0
		% von SDQdi	57,4%	42,6%	100,0%
		% von CDIdi	18,1%	69,0%	26,4%
% der Gesamtzahl		15,1%	11,3%	26,4%	
	Korrigierte Residuen	-13,7	13,7		
Gesamt	Anzahl	863	168	1031	
	Erwartete Anzahl	863,0	168,0	1031,0	
	% von SDQdi	83,7%	16,3%	100,0%	
	% von CDIdi	100,0%	100,0%	100,0%	
	% der Gesamtzahl	83,7%	16,3%	100,0%	

Zeichenerklärung: SDQdi = SDQ Gesamtpunktwerte, dichotomisiert, 1 für unter, 2 für über 75%;
 CDIdi = CDI Gesamtpunktwerte, dichotomisiert, 1 für Punktwert niedriger als 19, für gleich oder höher 19

Chi-Quadrat-Tests					
	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	188,112(b)	1	,000		
Kontinuitätskorrektur(a)	185,497	1	,000		
Likelihood-Quotient	166,295	1	,000		
Exakter Test nach Fisher				,000	,000
Zusammenhang linear-mit-linear	187,930	1	,000		
Anzahl der gültigen Fälle	1031				

Die Gesamttabelle ist hoch signifikant. Die Variation der Werte ist nicht als zufällig aufzufassen. Für alle Tabellenfelder sind auch die korrigierten standardisierten Residuen hoch signifikant.

Daraus ergeben sich die folgenden Feststellungen:

11.20 AS: Die dichotomisierten Gesamtpunktwerte von SDQ und CDI variieren hochsignifikant miteinander.

11.21 AS: Von den nach CDI als hochwahrscheinlich mit depressiven Störungen belasteten Jungen und Mädchen werden 69% auch nach SDQ Gesamtpunktwerten als wahrscheinlich von psychischen Störungen betroffen klassifiziert. Da das nach DTK anzunehmende Vorliegen depressiver Störungen auch den nach SDQ zu unterscheiden Störungen zugehört, ergibt sich eine nur wenig unter den Erwartungen liegende Übereinstimmung von SDQ und CDI.

11.22 AS: Die nach DTK wahrscheinlich nicht von depressiven Störungen betroffenen Jungen und Mädchen werden auch nach SDQ Gesamtpunktwerten zu den insgesamt eher nicht mit psychischen Störungen belastet Jungen und Mädchen zuzuordnen. Der SDQ weicht nur in 18,1% der vom DTK vollzogenen Klassifikation (Jungen und Mädchen nicht mit depressiven Störungen belastet) ab. Diese Abweichung ist anders zu bewerten als die unter 11.21 AS. Nach CDI wäre keine depressive Gefährdung anzunehmen. Der SDQ gibt Hinweise auf das Vorliegen einer, einer von mehreren möglichen, psychischen Störung ab. Beide Klassifikationen widersprechen einander nicht.

Nach diesen Überprüfungen der vorab formulierten Erwartungen zu den Relationen zwischen den Messwerten von SDQ, DTK und CDI ergeben sich empirisch begründete Einschätzungen, die von einer näherungsweise Übereinstimmung zwischen SDQ und CDI und von partieller Übereinstimmung zwischen SDQ und DTK sprechen lassen.

Nach den aus der Tabelle mit den Korrelationen zu entnehmenden Informationen waren keine anderen Befunde zu erwarten.

11.7 Die SDQ-Gesamtwerte als Information über das potentielle Vorliegen subklinischer und klinischer psychischer Beeinträchtigungen in der untersuchten Stichprobe

Das SDQ als Screening Verfahren zur Identifizierung von Kindern, die psychische Schwierigkeiten aufweisen, basiert, wie auch der DTK und CDI, auf die Annahme, dass höhere Gesamtpunktwerte Indizien für stärkere Belastungen darstellen. Auf die Suche nach diesem Trennwert, der eine „richtige“ Unterscheidung zwischen klinisch und nicht klinisch belasteten Kinder gewährleistet sollte, treten alle methodische und empirische Probleme auf, die ausführlich in vorangegangenen Abschnitten dieser Arbeit erörtert worden sind. Für den SDQ ergibt sich zusätzlich, vermehrt als beim DTK oder CDI, die Frage, auf welche genau klinisch auffälligen

psychischen Beeinträchtigungen die Werte der Gesamtskala „Total difficulties“ anhand eines Cutoff-Wertes hinweisen. Aus den Daten der vorliegenden Arbeit kann weder die Frage nach dem „idealen“ Grenzwert noch die Frage nach den zu messenden Konstrukten psychischer Störungen beantwortet werden. Zu Beantwortung der Forschungsfrage nach den Prozentsatz der beteiligten Kinder, die laut ihren Selbstangaben beim SDQ ernst zu nehmende psychische Beeinträchtigungen aufweisen, soll zwangsläufig in den relevanten Informationen aus vorherigen Studien mit diesem Verfahren eingegriffen werden. Dementsprechend wird hier angenommen, dass hohe Werte in der Gesamtskala des SDQ auf vermutlichen emotionalen Störungen oder Verhaltensstörungen oder Hyperaktivität hinweisen. Nach Goodman sollte ein SDQ-Wert von 20 Punkten als ein „warning signal“ (Goodman, Meltzer & Bailey, 1998) für das wahrscheinliche Vorliegen einer der oben genannten Diagnosen betrachtet werden. Tabelle 11.16 zeigt die von Goodman vorgeschlagenen Wertbereiche, die für die Identifizierung unterschiedlicher Ausprägungsgrade von „psychischen Beeinträchtigungen“ auch innerhalb dieser Arbeit verwendet werden.

Tabelle 11.16: Kritische SDQ-Gesamtpunktwerte (nach Goodman)

„Td“-GPW	Belastungsgruppe (nach Goodman)
0-15	„normal“
16-19	„borderline“
20-40	„abnormal“

Tabelle 11.17 zeigt die Zahl der an der Untersuchung 2034GR beteiligten Gymnasiasten, die nach ihren Selbstzuschreibungen von gefragten Symptomen im SDQ als alltäglich, subklinisch oder klinisch belastet gelten sollen.

Tabelle 11.17: Verteilung der Kinder aus der Stichprobe 2034GR auf nach Schweregraden unterschiedenen Gruppen mit mutmaßlich unterschiedlicher Belastung durch „psychische Störungen“

Belastungsgruppe	Td 1	Td 2	Td 3
„Td“-GPW	0 - 15	16 – 19	20 – 40
Schülerzahl	806	129	94
und Prozentsatz	78,3%	12,5%	9,1%

Aus der Durchführung eines nonparametrische Tests (χ^2) ergibt sich, dass die Anwendung der nach Goodman vorgeschlagenen Cutoffs zu solchen Gruppenbildungen führt, die signifikant von einer Zufallsverteilung abweichen ($\chi^2 = 939,259$; $p = 0,000$). Demnach dürfen die in Tabelle 11.17 dargestellten Informationen zur Formulierung des nachstehenden Arbeitsergebnisses genutzt werden.

11.23 AS: Über den Einsatz des SDQ in einer Stichprobe von 1034 griechischen Kinder zwischen 12 und 16 Jahren alt hat sich unter Anwendung der von Goodman empfohlenen Cutoff-Werte ergeben, dass 9,1% der gefragten Kinder in einem klinischen Ausprägungsgrade von

psychischen Problemen berichten und weiterhin 12,5% der Kinder von Symptomen subklinischer Ausmaßes berichten.

Der hier gefundene Prozentsatz der vermutlich klinisch psychisch belasteten Kinder ist etwa zweimal höher als der entsprechende aus einer Studie von Goodman, Meltzer und Bailey (1998) bei 83 Kinder zwischen 11 und 16 Jahren aus London (5% der Kinder aus dieser Arbeit als „abnormal“ belastet und 28% als „borderline“ belastet eingestuft).

In der entsprechenden Stelle bei der Datenanalyse der von DTK und CDI gewonnenen Daten (siehe Abschnitte 7.1 und 9.4) wurde für die Identifizierung der innerhalb dieser Stichprobe mutmaßlich klinisch depressiv belasteten Kinder neben den von Pötz und Kovacs entsprechend empfohlenen Grenzwerte auch ein zweites Kriterium verwendet. Dieses Kriterium wurde als „Frequenzunterschreitung von GPW=0“ bezeichnet und basiert auf die Annahme, dass dieser Trennwert eine kritische Abweichung innerhalb des Verlaufes der individuell erbrachten Gesamtwerte der jeweiligen Skala-Verteilung anzeigt.

Die nachstehende Tabelle stellt die nach Verwendung beider Schätzungen ergebnen Prävalenzraten von Kindern mit psychischen Auffälligkeiten eines mutmaßlich klinischen Ausprägungsgrades.

Tabelle 11.18: Nach SDQ-Gesamtpunktwerten geschätzte Prävalenzraten für das Vorliegen psychischer Störungen in der Stichprobe der Untersuchung 2034GR

Nr. der Schätzung	Trennwert	Zahl der Kinder (N= 1029)	Prozentsatz
1. Schätzung	Trennwert 20 nach Goodman	94	9,1%
2. Schätzung	Kriterium der Frequenzunterschreitung nach Kampisiou	22	2,1%

11.24 AS: Nach der Analyse der in der Untersuchung HE 2034 von 1029 Kindern gewonnenen SDQ-Gesamtpunktwerte ist anzunehmen, dass die Prävalenzrate der klinisch psychisch belasteten Kinder der vorliegenden Stichprobe mit hoher Wahrscheinlichkeit zwischen 2,1% und 9,1% liegt.

Kapitel 12

Das Auftreten depressiver Symptome bei Schulkindern in der Perspektive von Lehrkräften: Analyse der mit dem ESDM aufgenommenen Informationen

Untersuchungen zum Auftreten depressiver Symptome bei Kindern und Jugendlichen können unter anderem auch Anlass für die Nachfrage geben, ob die Lehrkräfte der jeweils untersuchten Kinder wissen, welche ihrer Schüler und Schülerinnen nach deren Eigenerleben und in ihrer über CDI oder DTK abgegebenen Selbsteinschätzung depressive Symptome zu tragen und zu ertragen haben. Das Interesse für eine Beantwortung dieser Frage ist keineswegs auf die Erörterung denkbarer Zusammenhänge von Urteilen aus personal unterschiedlichen Perspektiven beschränkt. Wenn Lehrkräfte, was begründbar angenommen werden könnte, über Informationen zum Auftreten depressiver Symptome unter den Kindern ihrer Schulklassen verfügten, wäre das von Bedeutung für die psychosoziale Versorgung solcher Mädchen und Jungen. Soweit das Auftreten depressionsspezifischer Symptome bei jüngeren Menschen als ein entsprechende Interventionen verlangendes Gefahrensignal im Hinblick auf deren künftige Gesundheit und Entwicklung verstanden werden sollte, könnten Lehrkräfte in einem beachtlichen Umfang helfen, notwendig werdende Hilfe zu realisieren.

Mit solchen Überlegungen wurde innerhalb der hier zu erörternden Untersuchung 2034GR der zusätzliche Einsatz eines Messinstrumentes zur Aufnahme von Informationen aus der Perspektive von Lehrkräften neben den bereits dargestellten Instrumenten zur Selbsteinschätzung des Auftretens depressiver Symptome bei Schulkindern beschlossen. Die Wahl fiel, vor allem wegen des Einsatzes dieses Instrumentes bei einer Untersuchung von mehreren Tausend spanischen Schulkindern, auf die Skala ESDM, „Escala de Depresión para Maestros“, von Domènech und Mit. (1985). Entsprechend den dargestellten Mutmaßungen war eine mit dem Einsatz der Skala ESDM verbundene Hauptfrage, ob sich für griechischen Lehrkräfte nachweisen ließe, dass die über ein beachtenswertes Wissen zum Auftreten depressiver Symptome unter ihren Schulkindern verfügten. Für diesen Fall bliebe auch zu erwägen, in welcher Weise dann Lehrkräfte zu Helfern in der Versorgung der durch depressive Störungen bedrohten Schulkinder werden könnten.

Die Analyse der mit dem ESDM in der Untersuchung 2034GR aufgenommenen Daten schafft eine empirische Basis für die Beantwortung der anstehenden Fragen. Der sich dabei ergebende Befund stünde bereits in Beziehungen zu einem vorzufindenden speziellen Forschungsbereich und zu dem dort erreichten Forschungsstand.

12.1 Die Lehrkräfte als Informationsquellen zum Auftreten depressiver Störungen bei Schulkindern

Im Allgemeinen gelten die Berichte der Kinder über ihre eigene emotionale Befindlichkeit als reliabel und valide (Barrett et al., 1991; Edelbrock, Costello et al., 1986; Polaino & Domènech, 1993). Das gilt vor allem für das momentane Erleben depressiver Symptome, allerdings weniger für Angaben zu deren Dauer oder zum Zeitpunkt ihres Erstauftretens (Puura et al., 1997). Eine qualitativ differenzierte Beschreibung der eigenen emotionalen Befindlichkeit hängt darüber hinaus vom Entwicklungsstand der Kinder ab (in Hoffmann, 1991). Verwiesen wird insbesondere auf die Abhängigkeit der Selbsteinschätzungen von den kognitiven und sprachlichen Fähigkeiten, vom Problembewusstsein und von der Fähigkeit zur Selbstreflektion. Zusätzlich scheint Verhalten

von Kindern in seinen Variationen stärker situationsabhängig zu sein als das der Erwachsenen (u.a. Verhulst und Koot, 1992).

Die genannten psychischen Aspekte geben Anlass, für die Diagnose depressiver Störungen bei Kindern und Jugendlichen explizit die Forderung nach Informationen aus unterschiedlichen Quellen als Standard zu setzen. Spätestens seit dem Beginn der 80iger Jahre des vergangenen Jahrhunderts werden diese Überlegungen innerhalb der psychiatrischen und psychologischen Forschung und Praxis diskutiert. Die Bemühungen um Informationen aus unterschiedlichen Quellen ergeben sich methodisch zwangsläufig aus dem Mangel an objektiv erfassbaren eindeutigen Anzeichen für das Bestehen einer depressiven Störung. Die durch DSM und ICD festgelegten Symptome und Störungsbilder umfassen eine Vielzahl von vornehmlich subjektiv aufnehmbaren und entsprechend restriktiv kommunizierbaren Phänomenen. Anders als etwa bei der Fraktur eines Knochens fehlen zweifelsfreie Indikatoren für das Bestehen depressiver Störungen. Alle bisher diagnostisch genutzten Informationen vermitteln Wahrscheinlichkeiten für das Auftreten der Störungen, aber niemals entsprechende Gewissheit. Die neben anderem mit dem Kürzel MMMT belegte diagnostische Verfahrensweise mit der die viel diskutierten Forderungen nach „multi method, multi source“ und „multi trait“ (siehe dazu auch Bortz und Döring, 2002) verbunden sind, versucht, durch methodische Absicherungen, zu denen die Durchführung unterschiedlicher Informationsaufnahmen gehört, die Verlässlichkeit von Daten, so auch von Vorhersagen, nutzbringend zu erhöhen. Die „Idee einer diagnostischen Vorgehensweise“ mit der „ein vollständiges Bild der psychischen Funktionalität von jüngeren Personen“ vermittelt werden könnte, (u.a. Kazdin & Petti, 1982; Verhulst & Koot, 1992) hat, trotz der Realität bildgebender Verfahren in der gegenwärtigen Medizin noch utopischen Charakter, was aber keineswegs ein Argument gegen die Vision selbst sein soll.

Die Hoffnung auf diagnostisch nutzbaren Informationszuwachs durch das in solche Prozesse einzubeziehende Wissen von Lehrkräften hat hohe Plausibilität.

Schulische Lehrkräfte bilden neben den Eltern und Mitschülern eine der Personengruppen, die nahezu täglichen Kontakt mit schulpflichtigen Kindern und Jugendlichen hat. Insoweit sind sie auch eine unter vielen identifizierbaren Informationsquellen. Wenn das als Argument für den Wert der durch sie zu erfahrenen Informationen über psychische Störungen bei Kindern genommen werden sollte, müssten die Lehrkräfte aus verfahrenslogischen Gründen in alle entsprechenden diagnostischen Prozesse einbezogen werden. Ob dies verfahrensökonomisch wäre, bleibt bei solcher Einschätzung unbeantwortet. Ihre Einbeziehung bei dem Versuch, psychisch belastete Schulkinder zu identifizieren, ist nach Sacco und Graves (1985) auch deswegen wichtig, weil die Lehrkräfte die Kinder in einem Kontext erleben, der relativ hohe Anforderungen sowohl an fachbezogene schulische Leistungen als auch an spezifische soziale Fertigkeiten stellt. Nach Meinung der Autoren sollten sich depressive Symptome in einem solchen Kontext offenkundig manifestieren. Die Schule wäre demnach ein diagnostisch relevantes Handlungsfeld und könnte durch seine Einbeziehung in die diagnostische Arbeit sowohl zum frühzeitigen Erkennen depressiver Störungen als auch zur Effizienzsteigerung von Interventionen nutzbar werden. Zuvor müsste der bislang vorherrschende Status der Annahme von Möglichkeiten aber durch die Beschreibung nachgewiesener Nutzbarkeit überwunden werden.

Die Einbeziehung von Schulen wird von manchen Autoren auch in Zusammenhang mit ökonomischen Aspekten erörtert, so beispielsweise von Essau, Karpinski und Mit. (1998): „Die Früherkennung sich entwickelnder depressiver Störungen kann auf effiziente und kostengünstige Weise erfolgen, besonders wenn die Jugendlichen noch zur Schule gehen. Lehrer könnten darin geschult werden, frühe Anzeichen spezifischer Störungen rechtzeitig zu erkennen“ (Essau,

Karpinski und Mit., 1998; S. 27). Das Zitat beschreibt zunächst aber nur eine Vision. Ob die wiedergegebene These realitätsnah ist, darf zumindest in Frage gestellt werden. Sie könnte auch eine zweckoptimistische Einschätzung sein. Dafür lassen sich Indizien anführen. Die Problematik wird aber an anderer Stelle dieser Arbeit eingehend zu diskutieren sein, denn der in dieser Untersuchung eingesetzte ESDM verspricht empirisch fundierte Aufschlüsse über die durch Lehrkräfte erbrachten Leistungen zur Identifizierung psychisch belasteter Kinder, wenn auch nur für eine Stichprobe griechischer Mädchen und Jungen.

Der bis hier beschriebene Diskussionsstand ist inhaltlich durch die Aussage zu kennzeichnen, dass die gleichzeitige Anwesenheit von Lehrkräften und Schulkindern in der Schule für die diagnostischen Kompetenzen von Lehrkräften beim Erkennen bestehender depressiver Störungen sprechen würde. Die mit solcher Aussage unterstellte kausale Beziehung folgt dem Muster der interpersonalen Gleichheit durch gleichzeitige Anwesenheit an einem Ort. Dieses Denkmuster wird hochfrequent im Alltagsleben verwendet. Psychologisch steht der Erklärungswert solcher Annahmen in Frage. Das beträfe dann aber auch die gegenüber Lehrkräften gehegten Kompetenzerwartungen.

In diesem Zusammenhang soll ein weiterer Sachverhalt erwähnt werden, der als Argument für die Untersuchung der möglicherweise depressionsrelevanten Wahrnehmungen von Lehrkräften herangezogen werden sollte. In der fast alltäglichen schulischen Interaktion im Klassenzimmer könnten und müssten, falls die entsprechenden Symptome bei einem oder mehreren der Kinder einer Schulklasse ausgebildet wären, im Verlauf eines Schuljahres fast alle in den Symptomlisten depressiver Störungen genannten Auffälligkeiten auftreten. Fraglich ist, welche davon als beobachtbar oder als zumindest erschließbar gelten sollten. Davon hinge ab, welche Symptome für die Lehrkräfte erfassbar wären. Einige depressive Symptome gelten als nicht beobachtbar und nur kommunizierbar. Lehrkräfte hätten zu denen keinen direkten Zugang.

Weiterhin ist zu beachten, dass die Verhaltensweisen der Kinder in ihren Schulklassen sowohl durch die Personen als auch durch die Situationen determiniert werden. Folglich enthalten dann Aussagen oder Beschreibungen zum Schulverhalten von Kindern oder Jugendlichen auch Inhalte, die vor allem situationsspezifisch die Interaktion von Kind und bestehenden schulischen Bedingungen erfassen. Was als Eigenheit des Kindes erscheint, zeigte sich an diesem Kind in anderen situativen Zusammenhängen möglicherweise nicht wieder. So könnte das Bekunden von Angst gelegentlich auch eine persönlich entwickelte und genutzte schulische Strategie sein. Schulunterricht folgt nicht zwangsläufig den Bedürfnissen der Schulkinder, sondern versucht auch, die möglichst effiziente und sozial reibungslose Durchführung des Unterrichts zu sichern. Das intendiert die Beeinflussung der psychischen Zustände möglichst aller Schulkinder. Zugleich evaluieren Lehrkräfte die Resultate ihrer Interventionsversuche, um deren Erfolg oder Misserfolg festzustellen. Dazu brauchen sie wieder und wieder Fakten über das Verhalten der Kinder in ihrer Klasse. Das spräche erneut für die Annahme der Ansammlung depressionsspezifischen Wissens durch Lehrkräfte. Andererseits werden alle Beobachtungen unmittelbar interpretiert und dabei Bewertungen des Beobachtenden unterzogen. Erst durch die Interpretation wird das mehrdeutige Faktum zur Information. Damit werden aus mancherlei Gründen und unabweisbar auch Fehlinterpretationen provoziert. In einer psychologischen Perspektive lösen sich viele plausible Annahmen in kritische Nachfragen auf. Warum auch sollten Lehrkräfte ihre Schulkinder so erleben, wie die sich selbst. So könnte beispielsweise das gleichzeitige Auftreten von emotionaler Belastung und agitiertem Verhalten bei einem Kind in der Interpretation durch die Lehrkraft, nicht wie bei Rossmann (2005) zu einem depressionsspezifischen Hinweis, sondern stattdessen zu einem provokant frechen Verhalten eines bekannt aufsässigen Schülers umgedeutet werden. Da die Ursachen eines beobachtbaren Schülerverhaltens nicht mit dem zusammen beobachtbar

auftreten, sind Fehlurteile über Verursachungen auffälligen Verhaltens außerordentlich wahrscheinlich. Damit fehlte dann aber auch vielen Verhaltensinterventionen der Lehrkräfte gegenüber ihren Schülern und Schülerinnen die als erwiesen geltende Rechtfertigung. Für unzutreffend begründete pädagogische Maßnahmen erhöht sich dann aber auch die Wahrscheinlichkeit, pädagogisch unangemessen zu sein. Den Begriff eines schulischen Verhaltensunfalls als gedachtem Gegenstück zu einem Verkehrsunfall gibt es nicht. Immerhin veranschaulichte ein solcher Vergleich die durch regelverletzendes pädagogisches Verhalten entstehenden potentiellen Gefahren für Leib und Leben der an den Interaktionen beteiligten Persönlichkeiten. Pädagogisch gedachte und dennoch verletzende Maßnahmen von Lehrkräften vermögen bei vulnerablen Kindern die Ausformung psychischer Störungen zu begünstigen oder auch herbei zu führen. So wäre dem Verdacht nachzugehen, Lehrkräfte würden zumindest gelegentlich vulnerable Kinder als solche im speziellen schulischen Kontext verkennen und durch fehlerhafte Attribuierungen und nachfolgende unangemessene schulische Maßnahmen psychisch schädigen. In diesem Sinne sind entsprechende Berichte von betroffenen Kindern, von Mitschülern, von Eltern und auch von Bildungskritikern heranzuziehen, die sich zudem auf empirische Studien berufen könnten und auch berufen, nach denen Lehrkräfte, aus welchen Gründen auch immer, zu Risikofaktoren für die psychische und die somatische Gesundheit von Kindern werden. Die entstehenden Beeinträchtigungen können zur Minderung von Unterstützung führen. Fry (1983), sowie Hecht und Mit. (1998) fanden beispielsweise, dass Lehrkräfte ihre Unterstützung denjenigen Kindern entziehen, die Verhaltensauffälligkeiten oder emotionalen Schwierigkeiten erkennen lassen. Lehrkräfte investieren häufig weniger Anstrengungen und pädagogisches Bemühen in die Förderung solcher Kinder.

Die gegenüber der zuvor beschriebenen Sichtweise, nach der eine Befassung mit dem depressionsspezifischen Wissen der Lehrkräfte wegen der von denen zu erhoffenden Unterstützung anzuraten sei, nun geöffnete Perspektive auf das Verhalten von Lehrkräften als potentiell Risiko, mit dem psychische Störungen ausgeformt werden könnten, sobald entsprechend vulnerable Kinder den für sie kritischen Interaktionen ausgesetzt wären. Die kontroversen Perspektiven nötigen, wissenschaftliches Interesse vorausgesetzt, eine spezifische Befassung mit dem depressionsrelevanten Wissen der Lehrkräfte.

Die im Sinne der psychosozialen und gesundheitlichen Vorsorge und Fürsorge für Schülerinnen und Schüler durchaus wohlwollend zu betrachtende Perspektive auf Lehrkräfte als willkommene Helfer findet eine mit Besorgnis wahrzunehmende Spiegelung auf Lehrkräfte als potentielle Verursacher psychischer Schäden. Inhaltlich widersprechen diese Sichtweisen keineswegs der Erfahrung von Schule als durchaus janusköpfiger Institution. Die Eigenheit, potentiell sowohl dem pädagogisch Guten als auch dem Übel dienen zu können, hat Schule mit allen Institutionen gemein, die sich Menschen schaffen. Und wenn schon Lehrkräfte als potentielle Risikofaktoren für die psychische und die somatische Gesundheit von Schulkindern genannt werden, gebietet die wissenschaftliche Korrektheit dem Empiriker auch auf den entgegengesetzten Wirkzusammenhang zu verweisen, nach dem die Unterrichtstätigkeit und somit auch die Interaktion mit Schulkindern ein Risikofaktor für psychische und somatische Gesundheit vulnerabler Lehrerinnen und Lehrer darstellen. Auch die weltweit umfängliche Literatur zum Burnout Syndrom bei Lehrkräften könnte als Beleg für die Berechtigung solcher Hinweise herangezogen werden.

Eine auf beide Perspektiven gleichermaßen zu beziehende kritische Frage bezieht sich auf die Aussagekraft der über Lehrerurteile zitierten Befunde. In zahlreichen Fällen problematisiert die niedrige Übereinstimmung zwischen den gewonnenen Informationen deren Nutzung.

Die potentielle Nutzbarkeit aufgenommenener Informationen wird neben anderem auch durch die Untersuchungsstrategie und die Wahl der Messinstrumente beeinflusst. Bereits in einigen der historisch ersten oder frühen Untersuchungen zu Lehrerurteilen über depressive Symptome bei Schulkindern sind Hinweise auf ungelöste Validitätsprobleme gegeben worden. Eine zentrale Thematik bei Untersuchungen zum Auftreten depressiver Symptome bei Schulkindern ist die Übereinstimmung zwischen dem durch die Kinder selbst berichteten Auftreten depressiver Symptome und dem dazu eingeholten Urteilen der Lehrkräfte. In einer Studie von Lefkowitz und Tesiny (1984) haben die Autoren zwischen den Angaben der Kinder im CDI und den Urteilen der Lehrkräfte nach einem inhaltlich übereinstimmenden Verfahren nur eine sehr schwache Korrelation ($r = 0,14$) gefunden. Dieses Ergebnis haben Sacco und Graves (1985) zu der Frage veranlasst, ob die Verwendung eines anderen Verfahrens, das vermehrt spezifische Symptome depressiver Störungen abfragen müsste, zu anderen Ergebnissen und zum Nachweis einer höheren Informiertheit der Lehrkräfte führen könnte. Die Autoren konstruierten ein Verfahren, das aus vier spezifischen Skalen bestand, nämlich 1. aus depressionsrelevanten affektiven Symptomen, 2. aus kognitiven Aspekten und 3. Fragen nach sozialem Verhalten, allesamt 4. durch Aufnahme vegetativer Beschwerden ergänzt. Diese Veränderung der Methode zur Befragung der Lehrer führte trotzdem kaum zu besseren Ergebnissen. Eine signifikante Korrelation, die aber in negativer Richtung verlief, trat zwischen den CDI-Angaben und den Lehrurteilen in der Skala „Vegetative Symptoms“ auf. Alle anderen Korrelationen waren nicht signifikant. In der Diskussion ihrer Ergebnisse verwiesen die Autoren auf das Bestehen von gewichtigen Validitätsproblemen. Sie stuften die Validität der Lehrurteile als sehr wahrscheinlich niedriger als die der Selbstangaben seitens der Kinder ein. Sie begründeten das wie folgt: „...CDI scores correspond well with clinicians' diagnostic ratings...“ (S. 354). Das trifft für Lehrurteile nicht zu. So kommen Sacco und Graves zu der Hypothese, dass „...the findings of the present study seem most likely due to the fact that the majority of symptoms of depression are private, internal and not apparent to teachers through normal interactions in the classroom. Even the few overt symptoms of depression (e.g., social withdrawal) may go relatively unnoticed because such problems are viewed by teachers as less serious than acting-out behaviors“ (S. 355). Dieser Begründung kann kaum zwingend widersprochen werden. Aber immerhin ist auch der Sachverhalt gegeben, dass die klinischen Diagnosen die Selbstauskünfte der Mädchen und Jungen als eine wesentliche Informationsquelle nutzen. Klinisch besteht, mangels wählbarer Alternativen, ein diagnostisches Primat der Selbstauskünfte der Klienten. Darauf und auf das Fortbestehen dieses Primats ist explizit zu verweisen.

Aus der Studie von Ines und Sacco (1992) ergab sich unter anderem eine Test-Retest Reliabilität der Lehrerangaben, die von den Autoren als „excellent“ eingestuft wurde. Die Korrelation zwischen den Urteilen der Lehrkräfte und den Selbstauskünften der Jungen und Mädchen, wieder über den CDI, blieb an den beiden Messzeitpunkten dieser Untersuchung dennoch relativ schwach ausgeprägt. Daran änderte auch ein kurzes „Trainingsprogramm“ über Depressionen bei Kindern nichts, das für einige Lehrkräfte zwischen den zwei Messzeitpunkten durchgeführt worden war. Der Grad der persönlichen Vertrautheit („familiarity“) zwischen Lehrkräften und Schulkindern wurde von den Autoren als ein Faktor verstanden, der die Übereinstimmung zwischen Fremd- und Selbstangaben beeinflussen kann.

Die zuvor angegebenen niedrigen Korrelationen zwischen Selbstberichten und Fremdurteilen durch Lehrkräfte finden sich in der einschlägigen Literatur wieder und wieder, sie sind insofern als Aussage zu generalisieren und dürften danach als ein konsistenter Befund gelten, der sich, erwartungsgemäß, mit Studien in unterschiedlichen Populationen und nach Verwendung verschiedenartiger Messinstrumenten immer wieder ergibt (Almqvist et al., 1999; Groen et al.,

1997; Kashani et al., 1983; Moor et al., 2000; Reicher, 1998). Achenbach, McConaughy und Mit. (1987) berichten nach einer Meta-Analyse der Ergebnisse aus Studien über die Korrespondenz zwischen Selbstangaben und Fremdurteilen von Lehrkräften über eine mittlere Korrelation von etwa 0,28. Wie schon weiter oben in den Ausführungen zitiert wurde, bezieht sich eine in der einschlägigen Literatur häufig gegebene Erklärung zum wiederholten Auftreten der geringen Übereinstimmung zwischen Selbst- und Fremdurteilen auf den Umstand, dass viele der Hauptsymptome depressiver Störungen vor allem innere Zustände betreffen, die von außen nicht beobachtbar sind. „So sind beispielsweise Emotionen wie Traurigkeit und Verzweiflung, Kognitionen von Schuld und mangelndem Selbstwert sowie eine beeinträchtigte Motivationslage mit Antriebslosigkeit und Unlust konstituierende Anzeichen einer Depression, die von außen, sowohl für Bezugspersonen als auch für den professionellen Betrachter, nicht immer leicht erkennbar“ (Groen und Petermann, 2002, S. 139). Die diese Aussage abschließend moderierende Wortkette „...nicht immer leicht erkennbar...“ könnte als resignierende Umschreibung für ein Dilemma verstanden werden, das Beobachter in entsprechenden Situationen erleben. Die zitierte Wortwahl verhüllt die bestehenden komplizierten perzeptiven und kognitiven Bedingungen. Die gemeinten Symptome sind nicht wahrnehmbar. Im günstigsten Fall sind sie erschließbar.

Bei dem Versuch, die Gründe für die Divergenzen zwischen den interpersonal divergierenden Urteilen zu erkennen und zu benennen, finden Weisz und Weiss (1991) die nachstehenden Bedingungen relevant. Die durch Lehrkräfte oder Eltern abgegebenen Beurteilung von Kindern enthalten relativ häufig verzerrende oder verfälschende Beimengungen, wie etwa 1. Vorurteile der Lehrer oder Eltern, 2. inkorrekte Kriterien und Folgerungen bei der Bewertung „pathologischen“ Verhaltens, 3. Übergeneralisierungen von Ausschnitten aus einem variierenden Verhaltensfluss oder auch 4. unterschiedliche Kombinationen aus allen schon genannten Fehlerquellen. Da das auftretende Verhalten nicht identisch sein muss mit dem jeweils wahrgenommenen Verhaltensbild, können die abgegebenen Urteile sowohl das Verhalten selbst als auch dessen subjektiv konstituierte Repräsentation betreffen. Die Urteile können einerseits das Verhalten selbst betreffen, andererseits aber lediglich auf das für den Beobachter gültige Wahrnehmungsbild bezogen sein. So schlagen Kolko und Kazdin (1993) vor, die aus Fremdquellen gewonnenen Informationen unter Berücksichtigung der Merkmale des jeweiligen speziellen Kontextes zu interpretieren. Von Lehrern oder Lehrerinnen abgegebene Urteile müssten also unter Berücksichtigung schulischer Gegebenheiten zu verstehen versucht werden, und die von Eltern stammenden Aussagen hätten familiale Gegebenheiten einzubeziehen. Diese Zielsetzung ist noch keine Beschreibung der dazu einzuführenden Verfahrensweise, aber sie öffnet eine beachtenswerte Perspektive, auf die an anderer Stelle in diesen Ausführungen erneut eingegangen werden wird. Folgerichtig ist in diesem Zusammenhang dann auch die von Verhulst und Koot gewählte Sichtweise zu beachten: „...Instead of interpreting this low agreement between different informants as unreliability or error, it should be recognised that each informant may contribute valid by different data. Discrepancies between different informants' reports may be as informative as agreements between them.“ (Verhulst und Koot, 1992, S. 35)

Nach der historisch erstmals auftretenden Frage nach der Informiertheit von Lehrkräften über das in ihren Schülern und Schülerinnen bedeutsame Erleben depressiver Symptome hat sich im Verlaufe weniger Jahrzehnte ein Forschungsbereich entwickelt, der beachtenswert bleibt, obschon die auf depressive Symptome bei Kindern und Jugendlichen bezogenen und empirisch kontrollierten Übereinstimmungen zwischen Selbst- und Fremdurteilen von Untersuchung zu Untersuchung beharrlich niedrig zu bleiben scheinen.

12.2 Die Skala ESDM als Verfahren zur Erfassung depressiver Symptome aus der Perspektive der Lehrkräfte

Für die Untersuchung des depressionsspezifischen Wissens der an der hier beschriebenen Untersuchung beteiligten Lehrkräfte wurde das spanischsprachige Verfahren ESDM („Escala de Depresión para Maestros“, Domènech und Mit., 1985; Domènech & Polaino-Lorente, 1990) eingesetzt. Bei dem Versuch, Informationen aus anderen Studien über dieses Verfahren zu finden, stößt ein nicht des Spanischen kundiger Forscher auf die Schwierigkeit, dass fast alle relevanten Informationen nur in spanischer Literatur verfügbar sind. Die einzige von uns gefundene relevante Stelle in der englischsprachigen Literatur befindet sich in einer Veröffentlichung von Canals, Marti-Henneberg, Fernández-Ballart & Domènech (1995), wo die Ergebnisse aus einer „longitudinal study of depression in an urban Spanish pubertal population“ berichtet werden. Dort ist zu erfahren, dass „The Depressive Symptomatology Scale for Teachers (DSST; Domènech & Polaino-Lorente, 1990) was developed for the first Spanish epidemiological study of depression in children of 9 years of age (Polaino & Domènech, 1990)... and has demonstrated to be of high internal consistency (alpha coefficient = 0,88) and convergent validity of 0,41 and 0,44 on CDI and CDRS-R respectively“ (Canals, Marti-Henneberg, Fernández-Ballart & Domènech, 1995; S. 104).

DSST ist das englisch gefasste Äquivalent für die spanische Skala ESDM. Eine hilfreiche Sprachbrücke zu der spanischen Literatur über das ESDM stellt die Arbeit von Casanovas (2003) her. Maria Casanovas hat das ESDM ins Deutsche übersetzt, und diese Übersetzung in einer deutschen Untersuchung an einer Stichprobe von in Südwestdeutschland lebenden spanischen Migrantenkindern (N= 170) und ihren sowohl deutsch als auch spanisch sprechenden spanischen Lehrkräften in Deutschland erprobt. Die Anwendung der Skala ESDM in einer deutschen Übersetzung war von Domènech angeregt und beraten worden. Das spezielle Untersuchungsinteresse in dieser Studie war auf mutmaßliche Zusammenhänge zwischen depressiven Störungen und Migrationsproblemen gerichtet. Aus dieser Arbeit von Maria Casanovas sind sowohl die deutsche Version des Verfahrens als auch der Hintergrund der Entwicklung der Skala ESDM bekannt geworden.

Nach Casanovas liegt der Ansatzpunkt für die Entwicklung des ESDM durch Domènech und Mitarbeiter in den schwachen Zusammenhängen zwischen den Selbstauskünften der Schulkinder und den Urteilen der Lehrkräfte über ihre Kinder. Dieser immer wieder neu festgestellte Sachverhalt provoziert bei seiner Kenntnisnahme augenscheinlich leicht kognitive Widersprüche, weil hingenommen werden soll, dass pädagogisches Fachpersonal viele Wochen- und Monatsstunden mit den Schulkindern zusammen lebt und, den Befunden nach, dann doch nichts oder überaus wenig von den depressiven Episoden weiß, die diese Kinder nach den von ihnen gegebenen Selbstauskünften durchleben. Die entsprechenden Ergebnisse sind von zahlreichen Autoren und dann auch in unterschiedlicher Weise problematisiert worden. Die mangelnde Übereinstimmung der Selbst- und Fremdurteile kann viele Ursachen haben. Auch Domènech sah sich veranlasst, den aufkommenden Fragen nachzudenken. Sie richtete ihre kritische Aufmerksamkeit unter anderem auf die Untersuchungspläne und die Messinstrumente. Einen Hauptfehler sah Domènech in dem bei der Befragung von Lehrkräften genutzten Repertoire an Fragen. Da auch für Kinder und Jugendliche die Symptome der depressiven Störungen primär mit dem Erleben innerer Ereignisse verbunden sind, ist die Hauptquelle für Informationen darüber zwangsläufig die Selbstauskunft der Betroffenen. Für alle weiteren Informationsquellen ist zunächst zu prüfen, in wieweit sie zu validen Informationen aus dem subjektiven Erleben der Kinder und Jugendlichen kommen. Lehrkräfte als Informanten über das subjektive Erleben der von ihnen betreuten Schulkinder sind generell als sekundäre Informationsquellen zu

klassifizieren. Sie können über das persönliche Erleben der Kinder und Jugendlichen in ihren Schulklassen empirisch verlässliches Wissen über das Auftreten vieler depressiver Symptome bestenfalls durch Mitteilungen der Kinder gewinnen. Also sind die ihnen zukommenden Informationen sowohl in Qualität und Quantität mehrfach abhängig, so vom genutzten Informationskanal, von den Erfassungsmethoden, von der Informationsverarbeitung der Lehrkräfte und von den Bedingungen der von ihnen ausgehenden Weitergabe von Informationen an deren Abnehmer. Die Erwartung, Lehrkräfte müssten das Auftreten depressiver Symptome bei Schulkindern bemerken und Wissen darüber sammeln, übergeht die informationspsychologischen und kommunikationspsychologischen Bedingungen der zu betrachtenden und zu analysierenden Prozesse und ersetzt die implizit durch die Forderung, Lehrkräfte hätten das zu leisten. Domènech griff den mit zuvor eingesetzten Verfahren etablierten Widerspruch auf, der entsteht, wenn Lehrkräfte zu depressionsspezifischen Beobachtungen von nicht beobachtbaren inneren Prozessen befragt werden. Die Beobachtungs- und Beurteilungsleistungen der Lehrkräfte sollten sich verbessern, wenn man die Lehrerinnen und Lehrer nach an Schulkindern beobachtbaren depressionsspezifischen Phänomenen befragte. Diesem Ziel folgte Domènech mit der Entwicklung eines spanischen Fragebogens, der Skala ESDM (siehe Casanovas, 2003). Domènech nutzt darin 16 Items, die auf Phänomene bezogen sind, die im Schulalltag mit einiger Wahrscheinlichkeit auftreten, depressionsspezifische Bedeutung haben und beobachtbar sind. Die Items betreffen beobachtbare Äußerungen depressiver Symptome, denen die Lehrkräfte im Schulalltag begegnet sein könnten, wenn Kinder in ihren Klassen depressive Episoden erlebt haben sollten. Der Fragebogen konfrontiert die Lehrkräfte bei seiner Anwendung über seine Items mit solchen Phänomenen, die ihnen dann aus ihrer Erfahrung in der Schulklasse bekannt sein könnten. Die 16 im ESDM verwendeten Aussagen korrespondieren direkt oder modifiziert mit 16 der Items im CDI von Kovacs. Diese Übereinstimmung zu erreichen war eine der Zielsetzungen, denen Domènech bei der Entwicklung der Skala folgte.

Die Skala ESDM ist von Polaino und Domènech (1993) in Untersuchungen mit über 6000 Kindern und entsprechenden Lehrkräften angewendet worden. Die aus diesen Untersuchungen gefundenen Erkennensleistungen der Lehrkräfte stimmen mit Selbstberichten von Schülern und Schülerinnen im CDI und im CDRS-R tatsächlich in höherem Masse überein als bei Untersuchungen, die nicht den ESDM benutzt hatten. Die Übereinstimmung zwischen Selbst- und Fremdurteilen, zwischen Angaben von Mädchen und Jungen einerseits und ihren Lehrkräften andererseits, erhöhte sich sichtbar. Die entsprechenden Ergebnisse aus der Untersuchung von Maria Casanovas verlangen dann allerdings erneut speziellen Erklärungsaufwand. Casanovas konnte in ihrer Untersuchung an einer relativ kleinen Stichprobe von 170 spanischen Migrantenkindern und ihren Lehrkräften in Deutschland die von Domènech gefundenen Übereinstimmungen (ESDM-CDI: $r = 0,41$ und ESDM-CDRS-R: $r = 0,44$) nicht replizieren und blieb deutlich unter den ursprünglichen Findungen (ESDM-CDI: $r = 0,23$). Die hohe Zahl der von Domènech und Mitarbeitern untersuchten über 6000 spanischen Kinder und spanischen Lehrkräften legt nahe, die Ergebnisse aus der Untersuchung von Casanovas als Effekt der überaus kleinen Stichprobe, insbesondere der an Lehrkräften, aufzufassen.

Domènech und Mitarbeiter erreichten durch die Veränderung der Messmethode eine Erhöhung der depressionsspezifischen Übereinstimmungen der Selbstauskünfte von Schulkindern und den durch ihre Lehrkräfte gegebenen Fremdurteilen.

Die Kenntnisnahme der von Domènech und Mitarbeitern erreichten Erhöhung der Übereinstimmung von Selbstauskunft und Fremdurteil legte nahe, die Skala ESDM in der griechischen Untersuchung mit der Stichprobe 2034GR einzusetzen.

12.3 Die Datenbasis der Analysearbeit zum ESDM

Wie schon in vorangegangenen Teil erwähnt, besteht das Verfahren ESDM aus 16 Items, die sich auf Verhaltensweisen und Äußerungen beziehen, die an Schulkindern im täglichen Schulalltag beobachtbar werden könnten. Die ESDM-Items haben expliziten Bezug zu denen im CDI und korrespondieren mit einigen der im DSM-IV angegebenen Symptome depressiver Störungen.

In der hier beschriebenen Untersuchung 2034GR wurde die Skala ESDM in einer griechischen Übersetzung von Kampisiou eingesetzt. Wie in Kapitel 6 dieser Erörterungen eingehend beschrieben, waren für die griechische Version der Skala ESDM durchgehend gute Messqualitäten nachzuweisen. Den mit der griechischen Version aufgenommenen Informationen ist insoweit einschränkungslos zu vertrauen. Alle in den mit dieser Stichprobe untersuchten 112 Schulklassen als Klassenleiterin oder Klassenleiter tätigen Lehrkräfte wurden von der Untersuchungsleiterin persönlich gebeten, für jeden einzelnen Schüler und für jede Schülerin ihrer jeweiligen Klasse einen ESDM-Bogen, der gegebenen Instruktion folgend, auszufüllen. Für jede ESDM-Aussage bestehen drei Antwortmöglichkeiten, die in der deutschen Übersetzungen mit „selten“, „manchmal“ und „sehr oft“ wieder zu geben sind. Die Verrechnung der Antworten erfolgte, wie auch im CDI, durch Vergabe eines Punktwertes aus der bei „selten“ ansetzenden Skala 0, 1 und 2. So kann der einem Kind von seinem Lehrer oder der Lehrerin zugeschriebene Gesamtscore im ESDM zwischen 0 und $16 \cdot 2 = 32$ Punkten liegen.

Aus der Gesamtzahl von insgesamt 112 befragten schulischen Lehrkräften haben 2 Personen die ESDM-Bögen für ihre Klassenkinder nicht zurückgegeben. Alle anderen Lehrer und Lehrerinnen kamen der Bitte nach, wenn auch mit unterschiedlichen Reaktionen auf die ihnen mit dem Ausfüllens der ESDM Formulare gestellten zusätzlichen Aufgabe. Die individuellen Grade von Bereitschaft und Interesse variierten stark. Manche Lehrkräfte mögen nur teilgenommen haben, um nicht in offensichtliche Opposition gegen die Anweisungen des Leiters oder der Leiterin der Schule treten zu müssen. Aus dem beobachteten Zusammenhang ergibt sich die Mutmaßung, das Ausfüllen des ESDM Formulars könne subjektiv als erhebliche Belastung empfunden worden sein. Das wäre zum einen mit dem zu leistenden zeitlichen Arbeitsaufwand zu erklären. Die Einschätzung des mit dem Ausfüllen notwendigerweise zu erbringenden Zeiteinsatzes ist schwierig, weil das Setzen von 16 Antwortkreuzen einerseits mit einem Minimaleinsatz von wenigen Minuten pro Schulkind zu leisten wäre, andererseits aber durch eingehendes Nachdenken über die jeweils zu beurteilenden Kinder zu Maximalwerten von mehreren Arbeitsstunden anwachsen würde. Erkennbar wird, dass der von einer jeden Lehrkraft einzubringende motivationale Einsatz wie auch die kognitive Leistung interpersonal erheblich variieren dürfte, was die Qualität der abzugebenden Urteile beeinflusst haben wird. In der Literatur zur Problematik der depressionsspezifischen Urteile von Lehrkräften über die von ihnen betreuten Schulkinder ist der Reliabilität der Urteilsfindung durch die Lehrer und Lehrerinnen bislang wenig Bedeutung zugemessen worden. Auch in der hier zu besprechenden Untersuchung kann über die potentielle Einwirkung von individueller Zufriedenheit oder Unzufriedenheit mit der übernommenen Aufgabe auf die Qualität der mit der Skala ESDM abgegebenen Urteile nur spekuliert werden. Mit welchem eigenen Qualitätsanspruch die ESDM Formulare durch die beteiligten Lehrkräfte bearbeitet wurden, ist nicht bekannt. So bleibt die als durchaus gewichtig einzuschätzende Einflussgröße der Urteilsbereitschaft der Lehrkräfte zwar nicht unbeachtet, aber dennoch verdeckt. Ohne Berücksichtigung dieser – keineswegs unwichtigen – Variablen - stellen die Antworten der Lehrkräfte auf die Fragen der Skala ESDM die Datenbasis der für die griechische Stichprobe durchzuführenden Analyse dar. Die mit der Skala ESDM abgegebenen Urteile der Lehrkräfte über ihre Schüler und Schülerinnen sind in einer Matrix von 2015 Zeilen in 16 auf die ESDM Items bezogenen Variablen durch die Werte 0, 1 und 2 symbolisiert.

12.4 Verteilung der ESDM-Punktwerte

Die in der Stichprobe 2034GR verwendete griechische Form der Skala ESDM entspricht sprachlich der spanischen Version. Nachfolgend sind deutschsprachige Charakterisierungen der Items aufgelistet. Die Lehrerinnen und Lehrer werden in der Untersuchung aufgefordert zu entscheiden, ob die mit den Items 1 bis 16 gegebenen Aussagen auf das jeweils zu beurteilende Kind in welchem Maße zutreffen.

Tabelle 12.1: Deutschsprachige Entsprechungen der Items 1 bis 16 in der griechischen und der spanischen Version der Skala ESDM

Itemnummer und Iteminhalt	
Das Kind....	Das Kind...
1 „fühlt sich traurig“	9 „fühlt sich gegenüber anderen weniger Wert“
2 „ihm fällt schwer, aufmerksam zu sein“	10 „widerspricht und streitet“
3 „ist in der Pause allein“	11 „neigt zu Selbstvorwürfen“
4 „weint“	12 „kann sich nicht entscheiden“
5 „versteh die Inhalte des Unterrichts“ ¹	13 „hat schwankende Stimmungslagen“
6 „wird in der Klasse akzeptiert“ ¹	14 „zeigt in der Schule aktive Mitarbeit“ ¹
7 „sieht müde aus“	15 „erreicht zufrieden stellende Leistung“ ¹
8 „hat Spaß in der Schule“ ¹	16 „spielt mit den anderen“ ¹

Zeichenerklärung: ¹= umgepolte Items

Die Lehrkräfte geben zu jedem Item eine auf das Kind nach ihrer Meinung zutreffende Einschätzung. Trifft die in der Skala ESDM vorgegebene Aussage nur „selten“ auf das Kind zu, wird dies mit dem Wert „0“ symbolisiert. Falls die Aussage „manchmal“ zutreffen sollte, steht dafür der Wert „1“. Würde die Behauptung in der Skala ESDM auf das Kind „sehr oft“ zutreffen, wird der Wert „2“ verwendet. Daraus ergibt sich eine formal definierte Abstufungen der durch jedes Kind erreichbaren Gesamtpunktwerte. Der aus 16 Items aufsummierte Gesamtpunktwert „0“ stünde für ein Kind, auf das alle 16 Aussagen in der Skala ESDM nur selten zutreffen. Die Klasse der Werte zwischen 1 und 16 wird von Kindern besetzt, für die ihre Lehrkräfte Kombinationen aus den Werten 0, 1 und 2 verwendeten. Der Wert 16 ist, wie auch immer die individuelle Konfiguration der als Urteile gegebenen Werte zusammengesetzt sein mag, identisch mit einem arithmetischen Mittel von 1, dem die Verbalisierung „manchmal“ zukäme. Alle Gesamtpunktwerte über 16 enthalten in ihren Konfigurationen aus 16 Einzelwerten das Urteil „sehr oft“. Es kann abhängig von dem Auftreten des Urteils „manchmal“ in der individuellen Konfiguration unterschiedliche Frequenzen haben.

Informationen über die von den Lehrkräften in der Untersuchung 2034GR mitgeteilten Einschätzungen zum Auftreten der über ESDM abgefragten Items sind der nachfolgenden Tabelle zu entnehmen. In einem ersten Analyseschritt wird festgestellt, welche Verteilung die von den Lehrkräften für die Mädchen und Jungen der Stichprobe vergebenen Gesamtpunktwerte einnehmen. Diesen Gesamtpunktwerten kommt ein etwas ungewöhnlicher Status zu. Der Gesamtpunktwert ist ein jedem Mädchen und Jungen zugerechneter Punktwert, der aber von den Lehrkräften vergeben wurde. Der Gesamtpunktwert der Skala ESDM ist somit ein den Schülerinnen und Schülern von ihren Lehrkräften zugeschriebener Punktwert. Tabelle 12.2 stellt die Frequenzliste der von den Lehrkräften ausgeteilten Gesamtpunktwerte der Skala ESDM

(ESDM-GPW) dar. Diese Liste betrifft eine Stichprobe von 1825 Mädchen und Jungen. Für 209 Schulkinder konnte kein ESDM GPW ermittelt werden.

Tabelle 12.2: Häufigkeits- und Prozentualverteilung der ESDM-Gesamtpunktwerte aus den Angaben der Lehrkräfte für eine Stichprobe von 1825 griechischen Kindern und Jugendlichen.

ESDM-GPW	Frequenz	Prozent	Kum. Prozent	ESDM-GPW	Frequenz	Prozent	Kum. Prozent
0	268	14,7	14,7	13	47	2,6	88,8
1	144	7,9	22,6	14	46	2,5	91,3
2	134	7,3	29,9	15	38	2,1	93,4
3	123	6,7	36,7	16	33	1,8	95,2
4	139	7,6	44,3	17	19	1,0	96,3
5	116	6,4	50,6	18	13	,7	97,0
6	135	7,4	58,0	19	17	,9	97,9
7	104	5,7	63,7	20	15	,8	98,7
8	99	5,4	69,2	21	5	,3	99,0
9	98	5,4	74,5	22	9	,5	99,5
10	73	4,0	78,5	23	4	,2	99,7
11	71	3,9	82,4	24	3	,2	99,9
12	70	3,8	86,2	25	2	,1	100,0
				Total	1825	100,0	

Die in Tabelle 12.2 dargestellte Verteilung der ESDM-Gesamtpunktwerte beruht auf ESDM-Fragebögen, die von den Lehrern vollständig ausgefüllt worden sind (n= 1825). Für die übrigen 209 (10,3%) Kinder der Stichprobe vermochten die Lehrkräfte entweder eine oder mehrere Fragen des ESDM nicht zu beantworten.

Die oben dargestellte Häufigkeitsverteilung der ESDM-Gesamtpunktwerte wird von folgenden statistischen Werten gekennzeichnet.

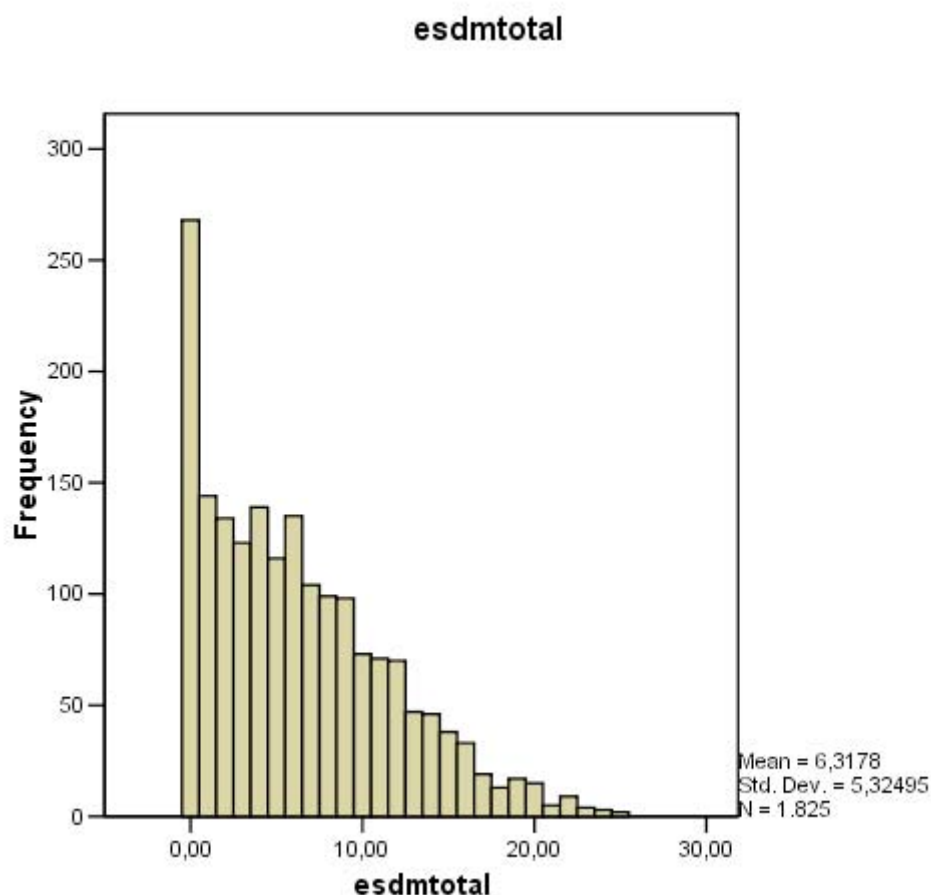
Mean	6,3178	Variance	28,355	Skewness	,818
Std. Error of Mean	,12465	Range	25,00	Std. Error of Skewness	,057
Median	5,0000	Minimum	,00	Kurtosis	,123
Mode	,00	Maximum	25,00	Std. Error of Kurtosis	,115
Std. Deviation	5,32495	Sum	11530,00		

Die theoretisch zu erwartende Abweichung der Produktion von ESDM-GPW gegenüber einer Normalverteilung war nach der Durchführung des Kolmogorov-Smirnov-Tests zu bestätigen (Kolmogorov-Smirnov Z = 5,029; p = 0,000). Aus der Position des arithmetischen Mittels (Mean = 6,31), des Medianwertes (Median = 5,00) und des Modalwertes (Mode = 0,00) darf diese asymmetrische Verteilung als linkssteil charakterisiert werden, worauf auch die gefundene positive Schiefe (Skewness = 0,818) verweist. Zusätzlich ist die gefundene Verteilung durch den relativ kleinen Exzesswert (Kurtosis = 0,123) eher als schmalgipflig zu bezeichnen (Bortz, 1999, 5. Aufl.; S. 47).

Die hier für die untersuchte griechische Stichprobe gefundenen deskriptiven statistischen Werte der ESDM-Verteilung weisen Ähnlichkeiten mit den Werten auf, die Casanovas (2003) nach der Anwendung des Verfahrens mit Lehrkräften von 170 spanischen Kindern in Deutschland berichtet hat (Mittelwert = 8,24; Varianz = 27,77; Standardabweichung = 5,27; Minimum = 0; Maximum = 24).

Die nachstehende Abbildung veranschaulicht die für die untersuchte griechische Stichprobe gefundene Verteilung der ESDM-Gesamtpunktwerte.

Abbildung 12.1: Verteilung der für 1825 griechische Schulkinder erhobenen ESDM Gesamtpunktwerte



Aus der ermittelten Verteilung der ESDM-Gesamtpunktwerte sind einige erste Informationen über die Urteile der beteiligten Lehrkräfte abzuleiten. Zu verweisen ist darauf, dass die Lehrkräfte mit ihren Antworten auf Items der Skala ESDM eingehen. Ob sie explizit Angaben zum Auftreten depressiver Symptome bei ihren Schulkindern geben, ist nicht sicher und eher zu bezweifeln. Die bis hier mit dem Einsatz der Skala ESDM gewonnenen Informationen sind in den nachstehenden Aussagen zusammengefasst.

12.1 AS: Die Verteilung der ESDM-Gesamtpunktwerte in der Untersuchung 2034GR erweist sich als linksasymmetrisch. Die sich ergebende Verteilung entspricht den theoretischen

Erwartungen gegenüber dem Einsatz von Messverfahren zur Erfassung des Auftretens von Störungsbildern in einer ansonsten eher unauffälligen Population.

12.2 AS: Die sich für diese Verteilung ergebenden statistischen Werte stimmen weitgehend mit entsprechenden Werten überein, die Casanovas (2003) nach Anwendung des Verfahrens in einer Stichprobe mit 170 spanischen Kindern in Deutschland gefunden hatte.

12.3 AS: Die dargestellte linksasymmetrische, schmalgipflige Verteilung der ESDM-Gesamtpunktwerte verweist darauf, dass die Mehrheit der Kinder nach Beobachtungen seitens ihrer Lehrer und Lehrerinnen durch eine jeweils eher geringe Zahl an depressiven Symptomen, erfasst durch die Items des ESDM, belastet ist.

12.4 AS: Immerhin 14,7% der Schüler und Schülerinnen weisen, den Angaben der Lehrkräfte und den Items der Skala ESDM zufolge, keine depressiven Symptom aus der ESDM-Liste auf.

12.5 AS: Die in Aussage 2 gegebene Information über die kaum belasteten Schüler und Schülerinnen stellt einen ersten Hinweis auf bestehende Divergenzen zwischen Selbst- und Fremdwahrnehmung depressiver Symptome innerhalb der vorliegenden Stichprobe dar. Laut Selbstangaben im DTK-Skala 1+2 „Depressive Belastung“ ist der Prozentsatz der nach DTK von depressiven Symptomen kaum belasteten Kinder nur mit 1,1% veranschlagen.

12.6 AS: Aus der Verteilung der DTK-Gesamtpunktwerte in der untersuchten griechischen Stichprobe war auch zu erfahren, dass der in der ESDM Teilstichprobe für den Punktwert 0 festgestellte Prozentsatz im DTK in etwa durch eine Teilstichprobe von 13,4% eingestellt wird, die sich aber bereits mit bis zu 4 DTK Symptomen belastet erweist.

12.7 AS: Der Prozentsatz der kaum belasteten Kinder im CDI ist ebenfalls nur 1,1%.

12.8 AS: In der Verteilung der CDI Gesamtpunktwerte aus der griechischen Stichprobe berichten 16,8% der Kinder ebenfalls schon darüber, von bis zu 4 CDI-Symptomen belastet zu sein. Andererseits berichten die Lehrkräfte, dass bei 14,7% ihrer Schüler kein ESDM-Symptom auftritt.

12.9 AS: In den zu vergleichenden Verteilungen (CDI, DTK und ESDM) aus der untersuchten griechischen Stichprobe treten rein numerisch für die als symptomfrei geltenden Jungen und Mädchen beachtenswerte Unterschiede zwischen Selbst- und Fremdanangaben zum Auftreten depressiver Symptome auf.

12.10 AS: Damit kann aber nicht nachgewiesen werden, dass die Wahrnehmung der Lehrkräfte eingeschränkt sei. Möglich wäre immerhin, dass die von Kindern im DTK und CDI gegebenen Mitteilungen mit der ESDM-Itemliste nicht synonym erfasst werden.

12.11 AS: In einer theoretisch verfügbaren Variationsbreite von 0 bis zu 32 Gesamtpunktwerten in der Skala ESDM erreichen die von Lehrkräften den Kindern zugeschriebenen individuellen Gesamtwerte bis 25 Punkte.

12.12 AS: Der höchste innerhalb dieser Untersuchung vergebene ESDM-Gesamtwert entspricht mit 25 Punkten dem Prozentsatz von 78,13% der vom ESDM theoretisch repräsentierten maximalen Belastung mit depressiven Symptomen. Den Urteilen der beteiligten Lehrer zufolge, reichen aber schon 50% der Skalenbreite des ESDM (16 Punkte) aus, um die Belastung von 95,2% Schüler und Schülerinnen nach ESDM Items treffend zu kennzeichnen. Die andere Hälfte der ESDM-Skala, die von den Gesamtpunktwerten 17 bis 32 gebildet wird, betrifft nach Angaben der Lehrer nur eine Minderheit von 4,8% der zu beurteilenden Kinder.

12.13 AS: Daraus ergibt sich, dass auch in der Wahrnehmung der in dieser Untersuchung beteiligten Lehrkräfte das Auftreten höherer depressiven Belastungen - repräsentiert durch höhere ESDM-Gesamtwerte - ein für nur maximal 4,8% der Kinder festzustellendes seltenes Ereignis ist.

12.14 AS: Die Verteilung der ESDM Gesamtpunktwerte für die untersuchte griechische Stichprobe steht in guter Übereinstimmung zu der in der Untersuchung von Casanovas (2003). Das ist kein Nachweis, immerhin aber im Zusammenwirken mit übereinstimmenden Messqualitäten, ein Indiz für sprachübergreifende identische Messleistungen.

Die Urteile der befragten Lehrkräfte manifestieren sich über die Verteilung der Gesamtpunkte hinaus in weiteren Häufigkeiten, die Einzelheiten darüber erkennen lassen, wie die Lehrkräfte der untersuchten griechischen Stichprobe das Auftreten der einzelnen ESDM-Symptome bei ihren Schülern und Schülerinnen wahrnehmen.

Während der Gesamtpunktwert die Menge der den einzelnen Schulkindern zugeschriebenen Punkte präsentiert, ist aus dem Prozentsatz der Nutzung eines Items zu ersehen, für wie viele Kinder in der Stichprobe die Lehrkräfte das jeweilige Item nutzten. Wenn sicher wäre, dass die Nennung eines Items durch die Lehrkräfte in jedem Fall mit dem wahrgenommenen Ausprägungsgrad des Auftretens am Kind übereinstimmte, wäre der Nutzungsgrad ein Indikator für die Stärke der itembezogenen Wahrnehmung der Lehrkräfte. Das einzelne Urteil einer jeden Lehrkraft könnte aber auch den Status von Vorurteilen haben oder ohne speziellen Bezug zu einem Kriterium und in diesem Sinne dann auch zufällig zustande gekommen sein. Wie das jeweilige Urteil entstand, ist aber ungewiss, so verweist der Prozentsatz der Nutzung vor allem anderen auf die von den Lehrkräften vorgenommenen Ankreuzungen.

Die nachfolgende Tabelle 12.3 gibt für jedes ESDM Item die auf den Grundwert der untersuchten Kinder bezogenen Prozentsätze an, mit denen über die Urteile der Lehrkräfte eine itemspezifische Zahl an Kindern als vom genannten Symptom betroffen erklärt wird. Der im Lehrerurteil enthaltene Ausprägungsgrad ist auch in der Tabelle repräsentiert.

Tabelle 12.3: Frequenzen für die auf insgesamt 48 wählbaren Positionen des Verfahrens ESDM gegebenen Lehrerurteile

	n	„selten“	„manchmal“	„sehr oft“	Miss	R „sehr oft“
1 „fühlt sich traurig“	1947	57,5	40,0	2,52	87 (4,3%)	5.
2 „fällt schwer, aufmerksam zu sein“	1947	44,7	40,9	14,4	87 (4,3%)	
3 „ist in der Pause allein“	1895	82,7	15,3	2,0	139 (6,8%)	4.
4 „weint“	1905	80,6	18,7	,7	129 (6,3%)	1.
5 „versteht den Unterricht“ ¹	1941	57,9	34,4	7,7	93 (4,6%)	
6 „wird in der Klasse akzeptiert“ ¹	1918	77,3	18,8	3,9	116 (5,7%)	
7 „sieht müde aus“	1940	59,9	37,5	2,6	94 (4,6%)	6.
8 „hat Spaß in der Schule“ ¹	1880	77,3	19,3	3,4	154 (7,6%)	
9 „fühlt sich weniger Wert“	1868	62,0	33,6	4,4	166 (8,2%)	
10 „widerspricht und streitet“	1945	69,0	25,5	5,5	89 (4,4%)	
11 „neigt zu Selbstvorwürfen“	1873	83,3	15,8	,9	161 (7,9%)	2.
12 „kann sich nicht entscheiden“	1913	51,4	43,9	4,7	121(5,9%)	
13 „schwankende Stimmungslage“	1914	56,9	40,2	2,8	120 (5,9%)	7.
14 „aktive Mitarbeit in der Schule“ ¹	1942	54,2	33,1	12,7	92 (4,5%)	
15 „zufrieden stellende Leistung“ ¹	1943	50,9	34,2	14,9	91 (4,5%)	
16 „spielt mit den anderen“ ¹	1881	82,0	16,2	1,8	153 (7,5%)	3.

Zeichenerklärung: „selten“, „manchmal“, „sehr oft“ = ankreuzbare Vorgaben im ESDM; miss = auf die Gesamtzahl der beteiligten Kinder bezogene Mengen an fehlenden Lehrerurteilen (ESDM-Urteile), ¹= umgepolte Frage

Aus der Tabelle ist zu ersehen, dass nach Meinung der Lehrkräfte die Zahl der von einem ESDM Item betroffenen Kinder und damit dann auch die Zahl der nicht betroffenen stark variiert. Nicht betroffen von einem ESDM Item sind jeweils zwischen 44,7% und 83,3% der Kinder.

Das nächsthäufigste vom ESDM erfragte Auftreten depressiver Symptome, die Einstufung „manchmal“, verrechnet mit 1 Punkt, stellen die Lehrkräfte, variierend mit den Items, bei 15,3%-43,9% ihrer Schulkinder fest.

Letztlich ist ein Prozentsatz zwischen 0,7 und 14,9 der beteiligten Kinder laut Urteilen der Lehrer und Lehrerinnen von einzelnen ESDM-Symptomen sehr häufig (2 Punkte) belastet. Die nachstehende Tabelle zeigt wieder eine prozentuelle Verteilung der von Lehrkräften pro Item abgegebenen Urteile, diesmal aber nur die auf die Angabe „sehr oft“ bezogenen Wahlen. Die Tabelle ist ergänzt durch eine Umrechnung der Prozentsätze in absolute Zahlen. Sie bestimmen die Menge der „sehr oft“-Urteile. Diese Zusammenfassung entspricht kategorial einer „hohen Betroffenheit durch das spezielle Symptom“. Sie stellt solche Kinder implizit denen gegenüber, die nach den Urteilen der Lehrkräfte als „nicht oder weniger betroffen“ anzusehen sind. Diese Angaben sind in den weiteren Spalten durch χ^2 -Werte ergänzt.

Tabelle 12.4: Frequenzen der von Lehrkräften im ESDM für die Antwort „sehr oft“ pro Item abgegebenen Antworten

	N	% sehr oft	n von „sehr oft“	Exp. „sehr oft“	Chi ²
1 „fühlt sich traurig“	1947	2,52	49,1	102,4	27,8
2 „fällt schwer, aufmerksam zu sein“	1947	14,4	280,4	102,4	309,3
3 „ist in der Pause allein“	1895	2	37,9	102,4	40,6
4 „weint“	1905	0,7	13,3	102,4	77,5
5 „versteh den Unterricht“ [†]	1941	7,7	149,5	102,4	21,6
6 „wird in der Klasse akzeptiert“ [†]	1918	3,9	74,8	102,4	7,4
7 „sieht müde aus“	1940	2,6	50,4	102,4	26,4
8 „hat Spaß in der Schule“ [†]	1880	3,4	63,9	102,4	14,5
9 „fühlt sich weniger Wert“	1868	4,4	82,2	102,4	3,9
10 „widerspricht und streitet“	1945	5,5	106,9	102,4	0,2
11 „neigt zu Selbstvorwürfen“	1873	0,9	16,9	102,4	71,5
12 „kann sich nicht entscheiden“	1913	4,7	89,9	102,4	1,5
13 „schwankende Stimmungslage“	1914	2,8	53,6	102,4	23,3
14 „aktive Mitarbeit in der Schule“ [†]	1942	12,7	246,6	102,4	203,2
15 „zufrieden stellende Leistung“ [†]	1943	14,9	289,5	102,4	341,9
16 „spielt mit den anderen“ [†]	1881	1,8	33,6	102,4	45,9
Spaltensummen			1638,8	1638,4	1216,4
					Df = 15

Unter Kenntnisnahme der in der voran stehenden Tabelle wiedergegebenen Verteilung der Lehrerurteile für die einzelnen Items der Skala ESDM ist auf das von Item zu Item recht unterschiedliche Aufkommen hinzuweisen. Die Spalte „Exp. „sehr oft“ gibt die statistisch erwartete Zahl der Schüler oder Schülerinnen an, die von den Lehrkräften bei dem jeweiligen Item mit dem Urteil „sehr oft“ bedacht werden sollten. Die als beobachtet genannten Mengenangaben sind Umrechnung des entsprechenden Prozentsatzes in der benachbarten Spalte. Alle Mengenangaben der Tabelle sind gerundet, wurden aber in den statistischen Berechnungen

ungerundet verarbeitet. Unter Annahme einer Gleichverteilung des Gesamtaufkommens an Nennungen auf die einzelnen Items ergibt sich eine hochsignifikante Abweichung der beobachteten Verteilung von den, unter Annahme der Gleichverteilung, erwarteten Häufigkeiten an Lehrerurteilen für das jeweilige Item. Die für die verschiedenen Items beobachteten Frequenzen an Urteilen nach der Einstufung „sehr oft“ sind nicht mit zufällig auftretenden Varianzen zu erklären. Aus dieser Feststellung wird aber auch nicht ersichtlich, womit diese nicht zufällig variierende Verteilung in Verbindung stehen könnte. Ein in diesem Zusammenhang nahe liegender Erklärungsversuch wäre, von einer Unterschiedlichkeit der Items auszugehen. Im Folgenden wird der Hypothese nachgegangen, dass drei Kategorien von ESDM Items existieren. Eine erste Kategorie umfasste danach Items, die innere Zustände (IZ) der Kinder betreffen, womit eine eher schwierige Beobachtbarkeit verbunden wäre. Zu dieser Kategorie „Innere Zustände“ werden die Items 1, 9 und 11 gerechnet. Eine zweite Kategorie betrifft „beobachtbare Eigenheiten der Kinder“ (BE). Dazu gehören die Items 3, 4, 6, 7, 13 und 16. Eine dritte Kategorie fasst Items zusammen, die „schulübliche Zuschreibungen an persönlichen Eigenheiten“ (SZ) zum Inhalt haben. Dazu wurden die Items 2, 5, 8, 10, 12, 14 und 15 gezählt.

Tabelle 12.5: In drei Itemgruppen angeordnete und auf das jeweilige Item bezogene Frequenzen der Lehrerurteile zur ESDM Kategorie „sehr oft“

Item	IZ	Item	BE	Item	SZ
1	49	3	38	2	280
9	82	4	13	5	149
11	17	6	75	8	64
		7	50	10	107
		13	54	12	90
		16	34	14	247
				15	290
Summe	148		264		1227

Zeichenerklärung: IZ= Itemgruppe „Innere Zustände“ mit den Items 1, 9 und 11; BE= Itemgruppe BE, „beobachtbare Eigenheiten der Kinder“, Items 3, 4, 6, 7, 13 und 16; SZ= Itemgruppe „schulübliche Zuschreibungen an persönlichen Eigenheiten“ mit den Items 2, 5, 8, 10, 12, 14, 15

Für die drei definierten Itemgruppen IZ, BE und SZ ergeben sich unterschiedliche Gesamtsummen, die das gruppenbezogene Aufkommen an Lehrerurteilen repräsentieren. Unter Einführung einer Gleichverteilungshypothese ergibt sich für jedes Item ein Erwartungswert von 1/16 des Gesamtaufkommens. Entsprechend der in jeder Gruppe zusammengefassten Zahl an Items ist der Erwartungswert pro Item mit der Zahl der Items in einer Itemgruppe zu multiplizieren. So ergeben sich die auf der Basis der Annahme von Gleichverteilung zu berücksichtigten Erwartungswerte. Der danach berechnete χ^2 – Wert ist mit 645,2 bei $df = 2$ hoch signifikant. Die in der Verteilung der Lehrerurteile auf die definierten Itemgruppen aufgetretenen Unterschiede dürften nicht als durch den Zufall verursacht angesehen werden. Die Wahrscheinlichkeit dafür läge niedriger als $p = 0,001$. Die nachstehende Tabelle fasst die Werte zusammen, die in der Berechnung des χ^2 – Wertes zu beachten waren.

Tabelle 12.6: Für die Berechnung des Chi-Quadrats der vorangegangenen Tabelle benutzten Werte

	IZ	BE	SZ	Zeilensumme	Chi ²
Obs.	148	264	1227	1639	
Exp. pro Item	102,44	102,44	102,44		
Exp. pro Itemgruppe	307,32	614,64	717,08	1639,04	
Chi ²	82,5	200,0	362,6	645,2	*** = hoch signifikant

Zeichenerklärung: IZ= Itemgruppe „Innere Zustände“ mit den Items 1, 9 und 11; BE= Itemgruppe BE, „beobachtbare Eigenheiten der Kinder“, Items 3, 4, 6, 7, 13 und 16; SZ= Itemgruppe „schulübliche Zuschreibungen an persönliche Eigenheiten“ mit den Items 2, 5, 8, 10, 12, 14, 15; Obs.= beobachtet; Exp.= erwartete Werte

Nach diesem Analyseschritt sollte, weil die entsprechenden Wahrscheinlichkeiten dafür sprechen, davon ausgegangen werden, dass die unterschiedlichen Lehrerurteile mit den unterschiedenen Itemgruppen variieren. Das berechtigt zu den nachstehenden Aussagen:

12.15 AS: Die in der Anwendung der Skala ESDM abgegebenen Lehrerurteile variieren mit drei Kategorien von ESDM Items.

12.16 AS: Die niedrigste Frequenz des Aufkommens an Lehrerurteilen wurde für die Gruppe von Items beobachtbar, die innere psychische Zustände der Kinder betreffen (Itemgruppe IZ).

12.17 AS: Ein etwa 1,5-faches Aufkommen war für beobachtbare Eigenheiten der Kinder festzustellen. (Itemgruppe BE).

12.18 AS: Gegenüber dem Aufkommen an Lehrerurteilen in der Itemgruppe IZ lagen die Frequenzen für schulübliche und explizit auf Unterrichtsverhalten bezogene Zuweisungen von Eigenheiten der Kinder über die Itemgruppe SZ etwa 8,3 mal höher.

12.19 AS: Selbst gegenüber einem durch Zusammenlegung der Frequenzen aus den Itemgruppen IZ und BE erhöhte Menge ist das Aufkommen an Urteilen zur Itemgruppe SZ noch immer um das 2,7 fache höher.

12.20 AS: In der Gesamtmenge an den im ESDM durch Lehrkräfte abgegebenen Urteilen über die Schulkinder in den eigenen Klassen dominieren auf schulübliche Kategorien von Verhaltenseinheiten, wie Aufmerksamkeit und Leistung, bezogenen Aussagen der Lehrkräfte, wenn der Vergleich auf die im ESDM wählbare Angabe „sehr oft“ angesetzt wird.

12.21 AS: Die beobachtete Dominanz könnte mit einem faktisch gegebenen Übergewicht des Auftretens der über die Itemgruppe SZ beschriebenen Verhaltensweisen zusammenhängen. Dafür sprechen die aus vielen Untersuchungen international bekannt gewordenen Schätzungen über depressive Kinder in Schulklassen. Die Gesamtfrequenz der Urteile über die Itemgruppe IZ könnte mit einer um 10% liegenden Rate dem entsprechen.

12.22 AS: Die vorangegangene Aussage entspräche einer Abweisung der benutzten Gleichverteilungshypothese.

12.23 AS: Der in den beiden vorangegangenen Aussagen benutzten Argumentation ist insoweit nicht zu folgen, als alle 16 Items in der Skala ESDM depressionsspezifischen Aussagewert haben sollen. Die Skala ist unter dieser Vorgabe konstruiert worden.

12.24 AS: Dann aber ist festzustellen, dass die Lehrkräfte unter den durch die Skala ESDM angebotenen Items eben die überzufällig hoch frequentieren, die in der Itemgruppe SZ expliziten Bezug zur alltäglichen Schulpraxis der durch Lehrkräfte präferierten Beurteilungskategorien haben.

12.25 AS: Die in der griechischen Stichprobe 2034GR beobachtete Dominanz spezieller schulbezogenen Items in der Vergabe des Urteils „sehr oft“ steht in hoher inhaltlicher Übereinstimmung zu Findungen, über die Casanovas (2003) in ihrer Untersuchung mit spanischen Lehrkräften berichtet.

12.26 AS: Die beobachtete Dominanz berechtigt zur Einführung einer auf sie bezogenen speziellen Hypothese. Angenommen werden darf, dass Lehrkräfte durch ihre schulische Praxis einer berufsspezifischen und unterrichtsrelevanten Akzentuierung in ihrer Wahrnehmung und in entsprechenden Wahrnehmungsurteile folgen.

Das, was die voranstehenden Aussagen verbalisieren, wird von der nachfolgenden Tabelle veranschaulicht. In ihr findet sich eine Spalte, in der die Prozentsätze für die Urteile „manchmal“ und „sehr oft“ addiert worden sind. Kategorial entspricht dieses einer Gruppenbildung von „betroffenen Kindern“ und „nicht betroffenen“. In dieser Tabelle sind die ESDM-Items nicht nach ihrer Nummerierung im ESDM aufgeführt, sondern in einer der Höhe des Prozentsatzes in der Spalte „1+2 Punkte“ folgenden Rangreihe. Der erste Platz kommt dem Item mit dem höchsten Prozentsatz in Spalte „1+2 Punkte“ zu.

Tabelle 12.7: Lehrerurteile über Häufigkeiten des Auftretens der ESDM Items bei den Kindern in ihren Schulklassen.

	0 Punkte	1 Punkt	2 Punkte	1+2 Punkte
2 „fällt schwer, aufmerksam zu sein“	44,7	40,9	14,4	55,3
15 „zufrieden stellende Leistung“ ¹	50,9	34,2	14,9	49,1
12 „kann sich nicht entscheiden“	51,4	43,9	4,7	48,6
14 „aktive Mitarbeit in der Schule“ ¹	54,2	33,1	12,7	45,8
13 „schwankende Stimmungslage“	56,9	40,2	2,8	43,0
1 „fühlt sich traurig“	57,5	40,0	2,5	42,5
5 „verstehen den Unterricht“ ¹	57,9	34,4	7,7	42,1
7 „sieht müde aus“	59,9	37,5	2,6	40,1
9 „fühlt sich weniger wert“	62,0	33,6	4,4	38,0
10 „widerspricht und streitet“	69,0	25,5	5,5	31,0
6 „wird in der Klasse akzeptiert“ ¹	77,3	18,8	3,9	22,7
8 „hat Spaß in der Schule“ ¹	77,3	19,3	3,4	22,7
4 „weint“	80,6	18,7	,7	19,4
16 „spielt mit den anderen“ ¹	82,0	16,2	1,8	18,0
3 „ist in der Pause allein“	82,7	15,3	2,0	17,3
11 „neigt zu Selbstvorwürfen“	83,3	15,8	,9	16,7

Die in dieser Tabelle abzusehende Rangfolge der Urteile nach dem Prozentsatz in der Spalte „1+2 Punkte“ („manchmal + sehr oft“) wird in der folgenden Tabelle weiter bearbeitet. Die Tabelle zeigt, welche Rangplätze auf welche Itemgruppen entfallen.

Tabelle 12.8: Rangplätze pro Itemgruppe nach der dem Prozentsatz in der Spalte „1+2 Punkte“ folgenden Rangreihe

R	RfI	RfSZ	RfBE	RfIZ
1	2	1		
2	15	2		
3	12	3		
4	14	4		
5	13		5	
6	1			6
7	5	7		
8	7		8	
9	9			9
10	10	10		
11	6		11	
12	8	12		
13	4		13	
14	16		14	
15	3		15	
16	11			16
Sp-Summe		39	66	31
Mittlerer Rang		5,57	11	10,33

Zeichenerklärung: R = Rangplatz, RfI = Rangplatz für Item Nr. n; RfSZ = Rangplatz für Itemgruppe SZ; RfBE = Rangplatz für Itemgruppe BE; RfIZ = Rangplatz für Itemgruppe IZ.

Aus der obigen Tabelle ist ersichtlich, dass auch bei einer Zusammenlegung der Urteile „manchmal“ und „sehr oft“ die Dominanz der Itemgruppe SZ erhalten bleibt. Der für sie geltende durchschnittliche Rangplatz 5,57 ist, der hier zuvor geäußerten und begründeten Hypothese entsprechend, niedriger als der für die Itemgruppen BE und IZ, die durchschnittliche Rangplätze von 11 und 10,33 erreichen.

Auch so bestätigt sich der Erklärungswert der Hypothese von der Dominanz der explizit unterrichtsbezogenen Items der Itemgruppe SZ.

Die Darstellung der prozentualen Betroffenheit der Schüler und Schülerinnen dieser Stichprobe durch die jeweiligen ESDM Symptome kann auch als Information über die Rangreihe der von den Lehrern beobachteten Symptome gelesen werden. So ist aus Tabelle 12.7 zu erfahren, dass die Lehrkräfte bei 55,3% ihrer Schüler und Schülerinnen manchmal oder sehr oft Aufmerksamkeitsschwierigkeiten (Item 2) beobachten. Darüber hinaus berichten die Lehrkräfte, dass etwa 50% ihrer Schulkinder manchmal oder sehr oft sowohl Schwierigkeiten damit haben,

zufriedenstellende schulische Leistungen zu erbringen als auch Entscheidungen zu treffen. Sie haben dementsprechend Probleme, aktiv in der Schule mitzuarbeiten (Items 15, 12 und 14).

Von den insgesamt 16 in der Skala ESDM verwendeten Symptomen werden andererseits 6 relativ selten von den Lehrkräften an ihren Schulkindern beobachtet. Die Lehrkräfte geben an, dass, umgerechnet auf durchschnittliche Schülerzahlen in den Klassen, etwa 8 von 10 Kindern in den Klassen selten zu Selbstvorwürfen neigen, immer mit Mitschülern in der Pause zusammen sind und mit denen spielen, selten weinen, fast immer Spaß in der Schule haben und von ihren Mitschülern akzeptiert werden (Items 11, 3, 16, 4, 8 und 6).

Die über die Skala ESDM aufgenommenen Informationen sind mit den vorangegangenen Ausführungen zur Kenntnis gebracht worden. Gegenstand der mit dieser Untersuchung geführten Analyse kann dann zum einen das Beziehungsgeflecht zwischen den aufgenommenen ESDM-Daten werden. Zum anderen bietet sich an, die Relationen zwischen den ESDM-Daten und den Daten aus den ansonsten in dieser Untersuchung benutzten Messinstrumenten zu erfassen und zu analysieren. Dem wird im Folgenden nachgegangen.

Die Zahl der potentiell zu betrachtenden Beziehungen zwischen den in dieser Untersuchung gesammelten Daten ist hoch. Sie könnte bereits auf der Itemebene ansetzen und die dort aufzufindenden Ergebnisse vergleichen.

In einem solchen Vergleich ergäbe sich beispielsweise die quantitativ beachtenswerte Übereinstimmung zwischen der von den Lehrkräften im ESDM bemängelten schulischen Leistungen einer Teilmenge von Schulkindern und den im CDI im Selbstbericht mitgeteilten schulischen Leistungsdefiziten. Die Lehrkräfte sehen nach ESDM bei 49,1% ihrer Schulkinder defizitäre Schulleistungen. Immerhin berichten 40,3% der Kinder in Selbstberichten über eigene Leistungsprobleme. Die Übereinstimmung ist nicht punktgenau, aber immerhin dürfte sie als relativ gut angesehen werden.

Dieser Sachverhalt rechtfertigt das Überdenken von Gründen für die Übereinstimmung. Die sollte nach den bis hier bereits mitgeteilten Ergebnissen erwartet werden dürfen. Schulleistungen werden primär durch die Lehrkräfte festgestellt und mitgeteilt. Wenn nun nachfolgend die Lehrkräfte danach über die Skala ESDM befragt werden, sollten sie sich, vor allen in der verlangten verallgemeinernden Form, ebenso erinnern können, wie die von ihnen täglich beurteilten Schüler und Schülerinnen. Nicht die Übereinstimmung der Angaben von Lehrkräften und Schulkindern ist verwunderlich, ihr Ausbleiben wäre es.

Ob dann diese partielle Übereinstimmung auch schon weitere konsonante Angaben ankündigt, wäre eingehend zu prüfen. Weil die durchgeführte Untersuchung 2034GR nach Hinweisen auf das Bestehen depressiver Störungen bei den Schulkindern der Stichprobe sucht, wird die Analyse der Beziehungen zwischen ESDM-Daten und denen aus den weiteren eingesetzten Messinstrumenten auf depressionsspezifische Aspekte zentriert.

12.5 Die Informationen über das Auftreten depressiver Störungen nach den Einschätzungen der Lehrkräfte und den Selbstauskünften ihrer Schulkinder

Die in der durchgeführten griechischen Untersuchung erhobenen Daten über das selbst berichtete Auftreten depressiver Symptome bei den Schulkindern der Stichprobe kommen aus mehreren Quellen. Die Unterschiedlichkeit der Daten ergibt sich aus den eingesetzten Messinstrumenten und den über sie von zwei Personengruppen eingeholten Informationen. Zum einen gehen die erhobenen Daten auf Angaben der Mädchen und Jungen aus den beteiligten Schulen zurück und zum anderen auf die der diese Schulkinder unterrichtenden Lehrkräfte.

Die in den vorangegangenen Abschnitten über die ESDM-Daten mitgeteilten Befunde geben Informationen über das in der Perspektive der Lehrkräfte zu bemerkende Auftreten depressiver Symptome bei den Schulkindern der untersuchten Stichprobe von griechischen Kindern. Zwar waren die Lehrkräfte nicht um diagnostische Einschätzungen gebeten worden, wohl aber haben sie mit ihren Angaben zu Eigenheiten der ihnen bekannten Schulkinder zumindest implizit über das Auftreten von depressiven Symptomen berichtet. Die Beobachtungen der Lehrenden sind bereits als solche von speziellem psychologischen Interesse. Da die Untersuchung aber insgesamt auf das Erfassen von Indizien für das Auftreten depressiver Störungen bei den untersuchten Kindern gerichtet worden ist, bliebe vordringlich zu entscheiden, ob die Einschätzungen der Lehrkräfte dazu einen Beitrag geleistet haben.

Ein Urteil darüber ließe sich empirisch verlässlich erbringen, wenn dazu eindeutige externe Kriterien über das Bestehen oder Nichtbestehen depressiver Störungen zu nutzen wären. Solche Kriterien sind bislang nicht bekannt. Entscheidungen über das Vorliegen depressiver Störungen beruhen auf Entscheidungsprozessen, die Informationen aus unterschiedlichen Quellen einbeziehen müssen. Das zu kennzeichnende Problem kann auch als Problem der eingeschränkten Validität aller bislang bekannten depressionsspezifischen diagnostischen Entscheidungsprozesse bezeichnet werden.

Die in dieser Untersuchung aufgenommenen Daten wurden alle über Messinstrumente mit eingeschränkter Validität für das zweifelsfreie Erkennen depressiver Störungen aufgenommen. Für das auf depressive Störungen bezogene Screening kann dies als Standardsituation bezeichnet werden.

Bezogen auf die über die Skala ESDM aufgenommenen Urteile der Lehrkräfte bleibt festzustellen, dass zunächst ungewiss ist, welche Validität ihnen zukommt.

Aber auch für die Selbstausskünfte der Jungen und Mädchen über ihre durch das Auftreten depressiver Symptome hergestellte eigene Betroffenheit gilt diese Einschränkung.

Der Versuch einer Entscheidung über den durch die Urteile der Lehrkräfte in der Skala ESDM erbrachten Beitrag zur Erfassung depressiver Störungen in der untersuchten Stichprobe wird durch den dargestellten Sachverhalt problematisiert.

Die nachfolgenden Analysen dienen einer Annäherung an die erwünschte und kaum erreichbare zweifelsfreie Beantwortung der Frage nach dem Aussagewert der Lehrerurteile in der durchgeführten Untersuchung.

Ein erster Analyseschritt versucht zu klären, ob die von den Lehrkräften in der Untersuchung 2034GR mitgeteilten Beobachtungen mit den Selbstangaben der Kinder und Jugendlichen in den eingesetzten depressionsspezifischen Messverfahren übereinstimmen.

12.5.1 Korrelationen auf der Ebene von Gesamtpunktwerten aus den eingesetzten Verfahren

Zunächst wird auf der Ebene von Gesamtscores überprüft, ob Übereinstimmungen zwischen Lehrerurteilen und Selbstausskünften der Kinder nachweisbar sind. In der Untersuchung 2034GR könnten solche Übereinstimmungen zwischen der Skala ESDM einerseits und den Verfahren CDI, DTK und SDQ bestehen. Für DTK und SDQ ist die Existenz verschiedener Skalen in diesen Verfahren zu berücksichtigen.

Aus unterschiedlichen Publikationen entnommene Informationen über Zusammenhänge zwischen Lehrerurteilen auf der Skala ESDM und den Angaben der von ihnen unterrichteten Kinder im CDI

berichtete Maria Casanovas (2003). Danach sind von den entsprechenden Autoren die nachstehend aufgeführten Korrelationen ermittelt worden: del Barrio (1989) $r = 0,38$, Domènech und Polaino-Lorente (1990) $r = 0,41$, Ezpeleta (1987) $r = 0,33$, Mestre, Garcia, Frias & Moreno (1993) $r = 0,31$, und Monreal (1988) $r = 0,45$. Casanovas (2003) selbst ermittelte in einer eigenen Studie einen Zusammenhang zwischen ESDM und CDI von $r = 0,23$.

Die nachstehende Tabelle zeigt für die griechische Untersuchung 2034GR die Korrelationen nach Pearson zwischen Lehrerurteilen in der Skala ESDM einerseits und von Schulkindern über Selbsteinschätzungsverfahren gegebene Angaben andererseits. Die Korrelationen betreffen sowohl die Gesamtstichprobe (dritte Spalte) als auch jeweils die Teilstichproben der Grundschulkindern und weiterhin die der Gymnasiasten. Gesondert ausgewiesen werden Korrelationen für die nach dem Geschlecht der Lehrkräfte unterschiedenen Teilstichproben (vierte bis siebte Spalte). Die Pearson Korrelationen wurden wegen der durch sie hergestellten Vergleichbarkeit mit anderen Studien ausgewählt. Aus der Durchführung von Korrelationen nach Spearman ergeben sich im übrigen sehr ähnliche Ergebnisse.

Tabelle 12.9: Korrelationen (r) zwischen Urteilen der Lehrkräfte (ESDM) und Selbstangaben der Kinder (CDI, DTK-Skalen und SDQ-Skalen) für die Gesamtstichprobe und für zwei Altersstufen (Kinder an Grundschulen versus Jugendliche von Gymnasien).

Korrelationen zwischen ESDM und...		Gesamt- stichprobe	Grundschule		Gymnasium	
			Männer	Frauen	Männer	Frauen
cdi (26 items) ¹	r	,270 ***	,192 **	,272 ***	,152 **	,288 ***
	N	1804	308	601	297	598
„Dysphorie“ ²	r	,126 ***	,081 n.s.	,096 *	,021 n.s.	,156 ***
	N	1808	310	600	298	600
„Agitiertes Verhalten“ ²	r	,211 ***	,116 *	,102 *	,116 *	,275 ***
	N	1809	311	603	298	597
„Müdigkeit“ ²	r	,082 **	,064 n.s.	,093 *	-,028 n.s.	,124 **
	N	1816	312	604	298	602
DTK „Skala 1+2“ ²	r	,186 ***	,111 n.s.	,108 **	,068 n.s.	,249 ***
	N	1798	309	596	297	596
„Total Difficulties“ ³ (im SDQ)	r	,230 ***			,179 **	,256 ***
	N	898			297	601
„Emotional problems“ ³ (im SDQ)	r	,122 ***			,100 n.s.	,130 **
	N	901			299	602
„Hyperactivity“ ³ (im SDQ)	r	,158 ***			,117 *	,186 ***
	N	900			298	602
„Conduct problems“ ³ (im SDQ)	r	,183 ***			,115 *	,216 ***
	N	899			298	601
„Peer problems“ ³ (im SDQ)	r	,206 ***			,190 **	,215 ***
	N	900			298	602
„Prosocial Behaviour“ ³ (im SDQ)	r	-,111 **			-,107 n.s.	-,108 **
	N	901			299	602

Zeichenerklärung: r = Pearson's Korrelation; N = Teilstichprobezahl; ¹ = CDI-Version 26 Items; ² = Subskalen des Verfahrens DTK; ³ = Subskalen des Verfahrens SDQ (das SDQ wurde nur von den Gymnasiasten ausgefüllt).
*** = $p < 0,001$; ** = $p < 0,01$; * = $p < 0,05$

Wie aus der Tabelle 12.9 zu ersehen ist, ergeben sich insgesamt schwache bis sehr schwache Korrelationen zwischen den Urteilen der Lehrkräfte und den Selbsteinschätzungen der von ihnen unterrichteten Jungen und Mädchen.

Damit aber ist eine für die durchgeführte Untersuchung wesentliche Aussage zu vermerken.

12.27 AS: In der griechischen Untersuchung mit der Stichprobe 2034GR bestehen zwischen den Urteilen der Lehrkräfte in der Skala ESDM und den Selbsteinschätzungen der von ihnen unterrichteten Schulkinder bezogen auf die Gesamtscores aus diesen Messverfahren insgesamt nur schwache Korrelationen, die den Wert $r = 0,30$ nicht übersteigen.

12.28 AS: Die für die Stichprobe 2034GR in den Gesamtscores insgesamt festgestellten signifikanten aber nur schwach ausgeprägten Zusammenhänge zwischen den Urteilen der Lehrkräfte (ESDM) über das Auftreten depressiver Symptome bei den von ihnen unterrichteten Schulkindern einerseits und den Selbsteinschätzungen der Schulkinder (CDI, DTK, SDQ) andererseits berechtigt zu der Feststellung, dass beide Personengruppen die Aufforderung zu Aussagen über beobachtete depressive Symptome bei den Schulkindern in unterschiedlicher Weise nachkommen. Diese Unterschiede können das jeweils gemeinte Objekt der Urteile oder die zur Urteilsfindung benutzte Verfahrensweise oder auch beides betreffen.

Der höchste Zusammenhang nach den Ergebnissen für die Gesamtstichprobe besteht zwischen den Urteilen der Lehrkräfte (ESDM) und den Angaben der Kinder im CDI. Aber auch dieser Zusammenhang ist schwach ausgeprägt ($r = 0,27$). Theoretisch war die im Vergleich zu anderen noch relativ hohe Übereinstimmung von ESDM und CDI insoweit zu erwarten, als das Verfahren ESDM auf Items des CDI basiert. Diese Verfahren benutzen mehrere weitgehend übereinstimmende Items.

Im Hinblick auf die Korrelationen zwischen ESDM und den DTK-Subskalen ergibt sich interessanterweise, dass die Gesamtbewertung der Lehrkräfte höher mit dem selbstberichteten „Agitierten Verhalten“ (DTK-Skala 2) zusammenhängt und niedriger mit den Selbstangaben der Kinder über die von ihnen erlebte „Dysphorie“ oder ihre gesamte „Depressive Belastung“ (Skala 1+2). Ähnliche Ergebnisse ergeben sich auch im Hinblick auf die Zusammenhänge zwischen ESDM und SDQ-Subskalen. Dort ist ebenfalls zu beobachten, dass die Skala ESDM stärker mit den Skalen des SDQ korreliert, die verhaltensbezogene Probleme messen sollten, so vor allem mit der Skala „Peer problems“ und eben weniger mit der eher depressionsrelevanten Skala des Verfahrens „Emotional symptoms“ (siehe Tabelle 12.9).

Bei dem Versuch, dieses Ergebnis zu erklären, scheinen zwei Sichtweisen von Bedeutung. Entweder besteht das Verfahren ESDM eher aus Items, die verhaltensbezogene depressive Phänomene umfassen, oder die Gesamtbewertung der Lehrer bezieht sich vornehmlich auf Wahrnehmungen von „externalisierten“ und weniger auf das Bemerkte „internalisierter“ depressiver Belastungen der Kinder.

In die Überlegungen dazu ist, im Rückgriff auf weiter oben mitgeteilte Analyseergebnisse, einzubeziehen, dass spezifische Besonderheiten der Urteile von Lehrkräften über depressive Symptome bei den von ihnen unterrichteten Schulkindern bereits aufgezeigt wurden. Danach unterscheiden sich Lehrkräfte und Schulkinder insoweit nachweislich in den Objekten ihrer Aussagen, als die Lehrkräfte die Itemgruppe der schulisch üblichen Zuschreibungen von unterrichtsrelevanten Eigenheiten ihrer Schulkinder massiv akzentuieren und, was als fast zwangsläufig auftretend gesehen werden könnte, die in der Perspektive der Kinder explizit subjektiv akzentuierten Items unübersehbar vernachlässigen. Sie beachten das subjektive Erleben der Kinder kaum. Und sehr wahrscheinlich wissen sie darüber auch wenig. Da solch Erleben nicht

beobachtbar ist, könnten die Lehrkräfte ein relevantes Wissen über subjektive Inhalte im Erleben der Kinder nur über entsprechend akzentuierte Kommunikationen erwerben. Die Schüler fühlen die sie treffenden Stimmungslagen jeweils selbst, auch empathische Lehrkräfte vermögen das nicht.

Als besonders niedrig erweist sich in einer Liste ohnehin schwacher Korrelationen der Zusammenhang zwischen Gesamtscores der Skala ESDM und der DTK-Subskala „Müdigkeit / autonome Reaktionen“ ($r = 0,082$). Bei der Beurteilung dieses Wertes ist die Tatsache zu berücksichtigen, dass die Skala ESDM nur ein Item zu skalenspezifischen körperlichen depressogenen Beschwerden enthält. Dieser Sachverhalt dürfte am Zustandekommen der genannten niedrigen Korrelation beteiligt sein, dennoch wäre die Erwartung höherer Korrelationen nach den relevanten Störungsbildern zu rechtfertigen. Das erwies sich im Übrigen auch innerhalb dieser Studie 2034GR an anderen Stellen der Analysearbeit über Korrelationen der DTK-Subskala „Müdigkeit/ autonome Reaktionen“ mit anderen Skalen, in denen kein Item explizit der DTK-Subskala entsprach. Die Erwartung höherer Korrelationen folgt der Annahme, dass depressionsspezifische Subskalen untereinander korrelieren müssten, weil die von einer depressiven Störung betroffenen Persönlichkeiten die von den jeweiligen Skalen vertretenen depressionsspezifischen Phänomene mit höherer oder geringerer Ausprägung in sich vereinen.

Die bisher diskutierten und im Sinne der Zielsetzungen dieser Untersuchung als eher unbefriedigend einzustufenden Zusammenhänge zwischen ESDM und den innerhalb dieser Untersuchung eingesetzten Selbsteinschätzungsverfahren rechtfertigen die Erörterung der sich anbietenden Folgerung, der Skala ESDM fehle gegenüber den die Selbsteinschätzung der Kinder nutzenden Verfahren eine hinreichende konvergente Validität.

Weil die auf ein identisches Objekt, nämlich auf depressionsspezifische Symptome bei Kindern, bezogenen Aussagen von unterschiedlichen Personengruppen in verschiedenartigen Perspektiven gebildet worden sind, werden Erklärungsversuch auch dort anzusetzen sein. Die Skala ESDM hat in der durchgeführten Untersuchung mit der Stichprobe 2034GR eine niedrige konvergente Validität gegenüber CDI, DTK und SDQ. Dies gilt aber nur insoweit, als die gefundenen Korrelationen als hinreichender Beleg für die Einstufung der konvergenten Validität mit den auf Selbstangaben beruhenden Verfahren angesehen werden dürfte. Korrelationen als Beleg für die Einstufung einer konvergenten Validität werden in dieser Erörterung aber nicht als hinreichend angesehen. Insoweit ist die mit Korrelationen begründete Einstufung der genannten konvergenten Validität, obschon durchaus gebräuchlich, mit entsprechenden Vorbehalten zu versehen. Gesichert werden müsste die bei Korrelationen der Gesamtscores implizit unterstellte Erwartung, der Gesamtscore der verglichenen Verfahren hätte identische Bedeutung als Indikator für depressive Störungen. Die mit Korrelationen eingeführte Unterstellung setzt implizit den Idealfall einer 1:1 Relation zwischen den Gesamtscores als zutreffend ein. Andere mathematisch definierbare Beziehungen sind denkbar und könnten vorliegen. Insofern ist die Feststellung einer niedrigen konvergenten Validität von ESDM einerseits und CDI, DTK sowie SDQ andererseits mit Vorbehalten zu versehen.

Die getroffene Feststellung ist weiterhin ausdrücklich mit dem Hinweis zu verbinden, sie sei auf Gesamtscores in der Gesamtverteilung von Lehrkräften und Kindern bezogen. Die Verfahren aber sind, wie sich nachweisen lässt, sensibel gegenüber systematischen Variationen in den Stichproben, was gleichbedeutend mit systematisch vorgenommen Modifikationen der Zusammensetzung der Stichproben ist. Der Begriff der modifizierten Stichprobe erweist sich in der Regel als oberbegriffliche Umschreibung für die Einbeziehung unterschiedlicher Personengruppen in die jeweilige Stichprobe. Damit wird nicht der nach Zufallsprinzipien

verlaufende Austausch von Persönlichkeiten verbalisiert, sondern der Austausch kategorial unterschiedener Personengruppen. Die benutzte Kategorie A zeigt an, dass die jeweils aus unterschiedlichen Persönlichkeiten gebildete Personengruppe bei denkbarer individueller Unterschiedlichkeit zumindest in einem Merkmal, nämlich dem von der kategorialen Benennung gemeinten, interindividuell übereinstimmt, und dies im Gegensatz zur Gruppe von Persönlichkeiten der Kategorie B. Die Gesamtstichprobe umfasst beispielsweise Vertreter beider Geschlechter. Da dies sowohl auf die Lehrkräfte als auf die von ihnen unterrichteten Schulkinder zutrifft, ergeben sich allein aus diesem Sachverhalt, sobald die Verfahren sensibel gegenüber der geschlechtsspezifischen Zusammensetzung von Stichproben sein sollten, die nachstehend unterschiedenen Varianten von Teilstichproben. Die wären relativ bedeutungslos, wenn die jeweils erhobenen Messdaten keine geschlechtsspezifische Abhängigkeit hätten. Falls aber die Messungen geschlechtsspezifische Sensibilität zeigten, würden korrelative Beziehungen zwischen zwei Gesamtstichproben A und B nicht zwangsläufig mit Korrelationen zwischen den von ihnen eingeschlossenen Teilstichproben $A_{(\text{männlich})}$, $A_{(\text{weiblich})}$, $B_{(\text{männlich})}$ und $B_{(\text{weiblich})}$ identisch sein. Für die zu betrachtenden Beziehungen zwischen Gesamtscores aus ESDM und CDI oder DTK gelten die nachstehenden Bedingungen.

Tabelle 12.10: Geschlechtsspezifische Varianten der Beziehungen zwischen Urteilen von Lehrkräften und Selbsteinschätzungen der Schulkinder in ihren Klassen

Schulkinder	Lehrkräfte		
	Lehrer	Lehrerinnen	nuB
Mädchen	Beziehung A	Beziehung B	2
Jungen	Beziehung C	Beziehung D	2
Summen	2	2	4

Zeichenerklärung: nuB = Zahl unterschiedlicher Beziehungen

Die unterschiedenen Varianten an Beziehungen erhöhen sich, wenn zusätzlich zwischen Kindern an Grundschulen und Schülerinnen und Schüler an Gymnasien unterschieden wird. Für diesen Fall sind zwei eigenständige Tabellen mit dann insgesamt 8 Variationen der betrachteten Beziehungen zu verwenden. Nachweise für die geschlechts- und altersspezifische Sensibilität der über ESDM, CDI und DTK eingeholten Fremd- und Selbsteinstellungen finden sich in den entsprechenden Tabellen zur Analyse der Daten aus CDI und DTK in den voran gegangenen Kapiteln und für die Skala ESDM in der hier in diesem Kapitel weiter oben angeführten Übersichtstabelle zu den in der Untersuchung 2034GR gefundenen Korrelationen. Betrachtet man dort (siehe Tabelle 8.10) die Spalten, in denen die Zusammenhänge zwischen geschlechtsspezifisch unterschiedenen Urteilen der Lehrkräfte im ESDM einerseits und Selbsteinschätzungsverfahren andererseits sowohl für Grundschulkinder als auch für die Gymnasiasten dargestellt sind, wird erkennbar, dass Unterschiede auftreten. So findet sich für die geschlechtsspezifisch unterschiedenen Korrelationen zwischen ESDM und DTK-Subskala „Müdigkeit“ für die Teilstichprobe der Gymnasiasten ein Zusammenhang von $r = -0,128$ bei männlichen Lehrkräften und $r = 0,124$ für weibliche Lehrkräfte. Im Überblick erreichen die Korrelationen für Urteile der männlichen Lehrkräfte numerisch niedrigere Korrelationen zu den Selbstangaben der Kinder, sowohl an Grundschulen als auch an Gymnasien. Darüber hinaus bestehen für die Teilstichprobe der weiblichen Lehrkräfte numerisch stärkere Zusammenhänge mit den Angaben der Kinder aus Gymnasien als mit den Angaben der Kinder aus Grundschulen. Diese Hinweise finden sich hier, weil sie die Vorbehalte gegenüber verallgemeinernden Urteilen über die konvergente Validität des Skala ESDM unterstützen. Zu verdeutlichen ist, dass die von

Domènech und Mit. (1985) sowie von Domènech & Polaino-Lorente (1990) berichteten Zusammenhänge nicht repliziert werden konnten. Die bestehenden Divergenzen zwischen den berichteten Korrelationen werden hier vor allem als Argumente für weitere Nachprüfungen gesehen. Für solche Überprüfungen wäre zu bedenken, ob sie nicht, anders als bei den häufig zu finden Beschränkungen auf Korrelationen zwischen Gesamtstichproben, von der Annahme interindividueller Unterschiede in den Leistungen von Lehrkräften beim Erkennen depressionsspezifischer Symptome ausgehen sollten. Dazu wäre auch die Einbeziehung einer Theorie dieser Urteilsbildungen hilfreich, die zu erklären versuchte, wodurch die unterschiedlichen Leistungen der Lehrkräfte beim Erkennen depressiver Symptome bedingt sein könnten. Explizite Annahmen dazu würden die Überprüfbarkeit der Hypothesen herstellen. Denkbar wären beispielsweise Unterschiede in den Erkennensleistungen, die mit der Anteilnahme der Lehrkräfte an der Befindlichkeit der Schüler und Schülerinnen variierten. Ein relativ einfach zu handhabender Indikator dafür könnte die Gesprächszeit sein, die Lehrkräfte für die nicht auf den jeweiligen Lernstoff bezogene Kommunikation mit Schülern und Schülerinnen aufwendeten. Aber auch der Umfang und die Ausdifferenzierung des Wissens über Schulkinder könnte indirekt mit den Erkennensleistungen variieren. Einen Zugang zu dieser Analyseperspektive öffnete die Berechnung von individuell auf die einzelne Lehrkraft bezogenen Zusammenhänge zwischen ihren ESDM-Urteilen und den Selbstauskünften der von ihnen unterrichteten Schulkinder im CDI oder dem DTK. Wenn die dargestellte Analyseperspektive mehr als eine zu rechtfertigende Spekulation sein sollte, müssten die für die einzelnen Lehrkräfte berechneten Korrelationen mit den Selbstauskünften ihrer jeweiligen Schulklassen untereinander Variationen aufweisen, die in ihrer Ausprägung nicht mit dem Zufall zu erklären wären. Da nicht angenommen werden sollte, dass alle Lehrkräfte gleiche Erkennensleistungen erreichten, läge nahe, die dann auftretenden signifikanten Unterschiede als auf lehrerspezifische oder klassenspezifische und zugleich unterschiedliche Bedingungen zurückzuführen. Falls dafür hinreichende empirische Indizien nachzuweisen wären, bestünde Anlass, die das Verfahren ESDM speziell betreffende Theorie zu prüfen und eventuell einer Revision zu unterziehen. Die Skala ESDM gilt als ein standardisiertes Verfahren, das geeignet ist, objektive Informationen über das Auftreten depressionsspezifischer Symptome bei Schulkindern zu erhalten. Objektiv sind sie dann auch, weil unabhängig von der urteilenden Persönlichkeit, und das ist die jeweils befragte Lehrkraft, in ihren messtheoretischen Qualitäten ungeschmälert erhalten bleibt, wenn von statistisch kontrollierten üblichen Toleranzbreiten gegenüber dennoch auftretenden interindividuellen Abweichen, also als zufällig aufzufassenden Unterschieden, abgesehen wird. Wenn demgegenüber sich aber erweisen sollte, dass die Variation der interindividuellen Unterschiede in den Urteilen der Lehrkräfte solche Toleranzbreiten überschreiten und also nicht mehr als zufällig auftretend interpretiert werden dürften, veränderte das nachfolgend und zwangsläufig die Entscheidungen über den Status der Skala ESDM. Indem die Skala keine interindividuell objektiven Lehrerurteile erfasste, wäre eine niedrige Validität zu erwarten. Nicht das Verfahren ESDM sicherte die Qualität der Wahrnehmungsurteile, ihre variierende Validität wäre im Gegenteil die Folge einer jeweils persönlichkeitsabhängigen und interindividuell unterschiedlichen Erkennensleistung der Lehrkräfte.

Das Verfahren ESDM unterliegt, anders als etwa CDI und DTK, dieser Anfälligkeit, weil die auf das jeweilige Verfahren zu beziehende Theorie für die Skala ESDM von der für CDI oder DTK geltenden Theorie im zu beschreibenden Urteilsprozess abweicht. Mit dem Verfahren ESDM werden Urteile über Verhaltensaspekte bei Schulkindern aufgenommen. Danach sind zwei potentiell auftretende Variationen zu beachten. Zum einen können die Lehrerurteile variieren. Des Weiteren sind Variationen im Schülerverhalten zu erwarten. Das nach der Durchführung des Verfahrens ESDM vorliegende Ergebnis präsentiert Werte, die durch zwei unterschiedliche

Variationen verursacht sein können. Aus den Werten selbst ist nicht ableitbar oder erkennbar, welche Unterschiede in den Werten auf welche den Urteilsprozess determinierenden Variablen zurück zu führen sind. Die Verhältnisse bringen eine Situation hervor, die der eines Autofahrers gleicht, der während seiner Fahrt fortlaufend die Geschwindigkeit seines Fahrzeugs kontrolliert, den Verkehrsvorschriften entsprechend sensibel auf veränderte Anzeigen des Tachometers reagiert und dennoch bei einer Geschwindigkeitskontrolle wegen Überschreitung der zulässigen Höchstgeschwindigkeit protokolliert wird. Da der vom Tachometer angezeigte Wert zumindest auf zwei potentiell auftretende Variationen zurück geht, auf Variationen in der Geschwindigkeit des Fahrzeugs und auf Variationen in der Funktionsweise des Tachometers, können die auf dem Tachometer erscheinenden unterschiedlichen Messwerte eben auch durch zwei verschiedenartige Variationsquellen verursacht worden sein. Welche Quelle im aktuellen Fall einen Wert produziert ist an dem selbst nicht erkennbar. Die Anzeige 112 Stundenkilometer kann jederzeit durch die aktuelle Geschwindigkeit oder aber durch eine nicht geplante und demnach also gestörte Funktionsweise im Tachometersystem, durch einen Defekt, verursacht worden sein. Bei einer immer in identischer Weise ablaufenden Funktionsstörung kann sogar eine bemerkenswert hohe Reliabilität entstehen. Das Verfahren ESDM steht aus analogen Gründen ständig im Verdacht, dass die von diesem Verfahren mitgeteilte Messwerte entweder durch qualitätsverändernde Eigenheiten in der Messeinheit Lehrkraft oder durch Variationen in oder an dem beurteilten Schulkind bedingt gewesen sein könnten. Zudem wäre möglich, dass beides miteinander oder gegeneinander auf die Messwerte einwirkte. Im Interesse der auf solche Messwerte beruhenden Folgearbeiten ist zu versuchen, die jeweils bestehende Ungewissheit durch Gewissheit verschaffende Erkenntnisse aufzuheben.

Daraus entsteht eine spezifische Verfahrensproblematik. Die hier problematisierte Situation in der Anwendung des Verfahrens ESDM ist nicht auf dieses Instrument beschränkt, sondern wird in ähnlicher Weise auch für CDI und DTK wirksam. Der Unterschied zwischen ESDM einerseits und CDI und DTK andererseits liegt in der personalen Verdopplung der Abhängigkeit der Messergebnisse. Bei der Anwendung des Verfahrens ESDM sind zwei Personen in den Urteilsprozess einbezogen, Lehrkraft und Schulkind. Bei der Anwendung von CDI oder DTK ist eine Person, eben das Schulkind, die personale Quelle von Variationen der Messwerte, weil Urteile über die eigene Person gegeben werden.

Falls das Verfahren ESDM, was anzunehmen nach den bislang bekannten Daten nicht abwegig wäre, zu Messwerten führte, die in ihrer Validität von der persönlichen Leistung der jeweils urteilenden Lehrkraft abhängig wäre, sollte es nicht als Verfahren zur Erfassung von depressiven Symptomen bei Schulkindern, sondern besser als Verfahren zur Erfassung der Identifikationsleistungen von Lehrkräften beim Erkennen depressiver Symptome an Schulkindern bezeichnet werden.

Das Verfahren könnte nicht, wie angenommen, die relative interpersonale Konstanz der Lehrerurteile als Verfahrensqualität garantieren, sondern anstelle einer interpersonalen Konstanz der Urteile bestenfalls nur eine nachweisbare intrapersonale Konstanz bieten. Die unterschiedliche Ausprägung der beispielsweise korrelativen Übereinstimmung zwischen dem Urteil einer Lehrerin oder eines Lehrers in der Skala ESDM mit Selbstauskünften der von ihnen unterrichteten Schulkindern wäre eine persönliche Leistung der jeweiligen Lehrkraft. Das Verfahren ESDM erfasse die unterschiedlichen Leistungen der Lehrer und Lehrerinnen und brächte die zur Kenntnis.

Wenn diesen Mutmaßungen nachgegangen werden sollte, erleichterte die Bezugnahme auf ein valides Außenkriterium die notwendige empirische Arbeit. Valide externe Kriterien für das Bestehen von depressiven Störungen bei griechischen Schulkindern sind innerhalb der

Untersuchung 2034GR aber nicht verfügbar. So kann auch nicht die Frage beantwortet werden, ob die mit Gesamtpunktwerten aus dem Verfahren ESDM nachgewiesenen Erkennensleistungen der Lehrkräfte, bezogen auf ein externes Kriterium für depressive Störungen bei Schulkindern, zufällig oder signifikant variierten. Verfügbar ist andererseits aber das externe Kriterium der Selbsteinstufungen der Schulkinder. Demnach kann nachgeprüft werden, ob die individuellen Leistungen der Lehrkräfte im ESDM, gemessen als korrelative Übereinstimmung mit den Selbstangaben der Kinder im CDI oder DTK, als zufällig variierend aufgefasst werden dürfen. Falls dies so sein sollte, dürfte nicht von individuell variierenden Leistungen der Lehrkräfte gesprochen werden. Sollten die Variationen dieser Zusammenhänge sich aber nicht als zufällig erweisen, wäre das Verfahren ESDM als Instrument für die Erfassung der unterschiedlichen Erkennensleistungen von Lehrkräften zu diskutieren.

Ob eine solche Perspektive mit den in der Untersuchung 2034GR ermittelten Ergebnissen zu begründen und zu rechtfertigen wäre, wird nun nachfolgend geprüft.

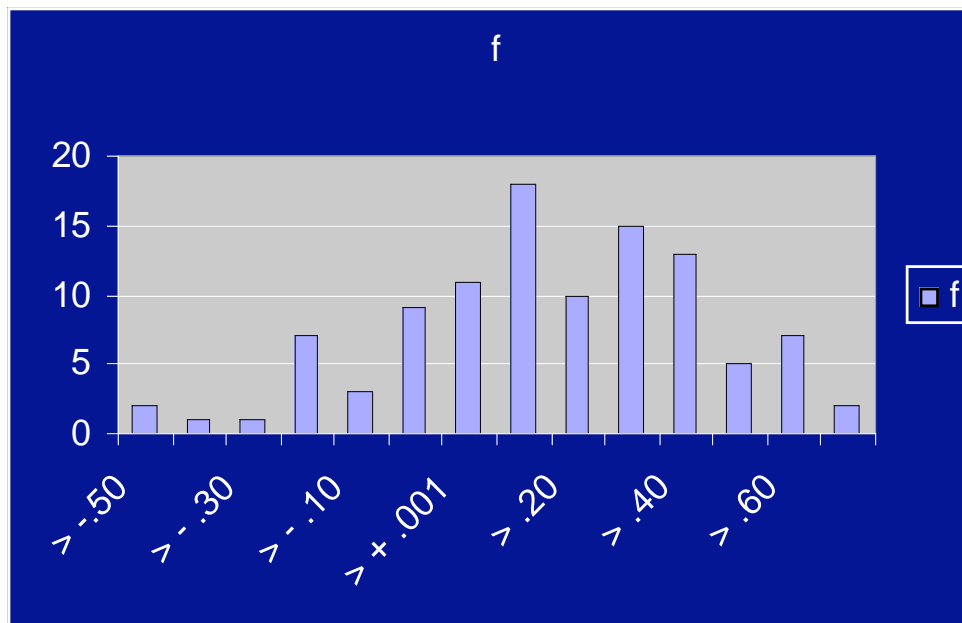
Ein erster Analyseschritt kann durch die Berechnung der für jede der insgesamt 112 in der Untersuchung besuchten Schulklassen feststellbaren Korrelationen zum Zusammenhang zwischen dem im Verfahren ESDM gegebenen Urteil der unterrichtenden Lehrkraft und der Selbsteinstufung der Kinder dieser Klasse über CDI oder DTK vollzogen werden.

Tabelle 12.11: Spearman's Korrelationen (rho) zwischen Urteilen der Lehrkräfte (ESDM) und Selbstangaben der Schulkinder (im CDI) in jeder der in die Untersuchung 2034GR einbezogenen Schulklassen

Klasse Nr.	N	Rho	Klasse Nr.	N	Rho	Klasse Nr.	N	Rho	Klasse Nr.	N	Rho
1	14	,155	29	8	,173	57	25	,020	85	29	-,040
2	18	,374	30	9	-,561	58	13	-----	86	24	-,154
3	18	,085	31	7	,607	59	20	-,215	87	26	-----
4	20	-,203	32	19	,142	60	26	,229	88	10	-,100
5	13	-,136	33	19	-----	61	24	-,052	89	14	,156
6	14	-,217	34	19	,705 **	62	22	-----	90	14	,364
7	18	,094	35	9	,415	63	26	,573 **	91	18	,289
8	16	-,206	36	14	,441	64	21	,384	92	22	,339**
9	18	,643 **	37	10	-,196	65	10	,352	93	28	,001
10	27	,493 **	38	14	,297	66	15	,576 *	94	31	,374 *
11	28	-,247	39	25	,072	67	24	,318	95	19	,591 **
12	15	,167	40	23	,514 *	68	21	-,321	96	17	-----
13	13	-,046	41	21	,461 *	69	30	,408 *	97	14	,232
14	14	,769 **	42	20	,093	70	9	,338	98	16	-,041
15	20	,453 *	43	24	,282	71	11	,213	99	10	,616*
16	29	,306	44	17	,608 **	72	13	,490*	100	8	,319
17	24	,021	45	15	,320	73	7	,468	101	9	-,567
18	17	,173	46	20	,474 *	74	18	,621 **	102	5	-,447
19	17	,091	47	19	,356	75	21	,211	103	13	,458
20	24	-----	48	20	,402*	76	13	-,096	104	8	,084
21	23	,220	49	15	,155	77	18	,166	105	13	,442
22	24	,151	50	18	-----	78	14	,229	106	13	,503*
23	12	-,046	51	17	,204	79	17	,578 *	107	16	,433*
24	25	-----	52	22	,188	80	13	,694 **	108	17	-,218
25	25	,620 **	53	11	,122	81	16	-,034	109	21	,137
26	17	,376	54	23	,023	82	17	,142	110	22	,104
27	18	-,202	55	24	,012	83	19	,181	111	19	,369
28	9	,134	56	13	,326	84	22	-,005	112	19	-----

Aus der obigen Tabelle sind die Korrelationen für 112 über ESDM von den Lehrkräften in 112 Schulklassen gegebenen Urteile mit den über CDI mitgeteilten Selbstausskünften der von ihnen unterrichteten Schulkinder zu ersehen. Für 8 von 112 Schulklassen fehlen die Berechnungen, weil ein Teil der zur Berechnung erforderlichen Messdaten nicht vorlag. Für 104 Schulklassen konnten die Korrelationen berechnet werden. In einigen Schulklassen ist die Zahl der beteiligten Schulkinder relativ gering. Die berechneten Korrelationen variieren von -0,56 bis +0,77. In der nachstehenden Abbildung sind die errechneten Korrelationen auf 14 Werteklassen verteilt worden.

Abbildung 12.2: Auf 14 Werteklassen verteilte Korrelationen zwischen 104 über ESDM abgegebene Urteile von Lehrkräften und den über CDI erfassten Selbstausskünften der Schulkinder in 104 Schulklassen



Gegenüber dem Verfahren ESDM war die Frage aufgeworfen worden, ob die Korrelationen zwischen den über ESDM abgegebenen Lehrerurteilen in ihrer interindividuellen Variation zufällige oder signifikante Unterschiede aufweisen. Zu den Schwierigkeiten der Signifikanzberechnungen wird hier auf Bortz (1999, S. 259 u.f.) und auf Bortz und Döring (2002) verwiesen. In der vorangegangenen Tabelle 12.11 finden sich Angaben zu den festgestellten Signifikanzen. Nicht als zufällig zu bezeichnende Unterschiede zwischen auftretenden Korrelationen weisen ihren signifikanten Abstand zum Vergleichswert durch entsprechende Konfidenzbereiche aus. Unter den mit der dargestellten Verteilung der Korrelationen geltenden Verhältnisse kann insoweit auf eine Auflistung der einzelnen Unterschiedsprüfungen verzichtet werden, als nach der dargestellten Verteilung der rho-Werte drei Gruppen von Korrelationen unterscheidbar sein könnten. Eine Gruppe von Lehrkräften, die hier mit GG gekennzeichnet wird, erreicht mit den für sie geltenden Übereinstimmungen zwischen ESDM und den über CDI mitgeteilten Selbsteinschätzungen der von ihnen unterrichteten Schulkinder keine signifikante Absetzung von einer mit dem rho-Wert 0,000 als unterliegend gedachten Grundgesamtheit GG. Die von diesen Lehrkräften erreichten rho-Werte sollten demnach als zufällige Abweichungen von einer mit der Korrelation $\rho = 0,000$ gekennzeichneten Grundgesamtheit (Gg) gelten.

Gegenüber dieser Teilgruppe von Lehrerurteilen haben sich in Richtung auf den Rho-Wert 1.00 einzelne Lehrkräfte mit ihren individuellen Korrelationen in signifikanter Weise abgesetzt. Diese Lehrkräfte der Gruppe Plus erreichen signifikant höhere Leistungen im Benennen der Selbsteinschätzungen ihrer Schülerinnen und Schüler als die zuvor aufgeführte Gruppe GG.

Eine dritte Gruppe, die Gruppe Minus, setzte sich mit der Ausprägung ihrer Korrelationen in Richtung auf die Korrelation -1.00 ab. Die Minus-Gruppe der Lehrkräfte ist mit ihren Korrelationen numerisch von den Lehrkräften unterschieden, die der Gruppe GG angehören. Aber obschon immerhin drei dieser Rho-Werte über -.40 liegen, kann doch für keine die Signifikanz der Abweichung von der Gruppe GG nachgewiesen werden. Der Hinweis darauf, dass vermutlich die geringe Gruppengröße der einbezogenen Schulkinder in den entsprechenden Klassen das Ausbleiben der Nachweise bewirkte, schafft selbst keine Signifikanzen. Ebenso kann auch der Hinweis auf nur knapp das erforderliche Signifikanzniveau verfehlende t-Werte keine Änderung der Datenlage herbeiführen. Für die Gruppe Minus kann in dieser Untersuchung keine Signifikanz für die anzunehmende Unterscheidung von der Gruppe GG vorgelegt werden.

Demgegenüber ist die Plus-Gruppe der Lehrkräfte mit ihren Korrelationen signifikant und teilweise sehr deutlich von den Angehörigen der Gruppe GG unterschieden. Weiterhin ist für die drei höchsten positiven Korrelationen und die drei niedrigsten auf der Gegenseite der Verteilung nachzuweisen, dass sie im Hinblick auf die ihnen zukommenden Korrelationen keiner gemeinsamen Grundgesamtheit angehören können. Die Irrtumswahrscheinlichkeit für diese Aussage liegt unter $p = 0,01$. Aus diesem Sachverhalt geht hervor, dass die in der Untersuchung 2034GR zu beobachtenden Urteile der Lehrkräfte über die Selbsteinschätzungen der von ihnen unterrichteten Schulkinder mit den Lehrerpersönlichkeiten zwischen $\rho = -0,567$ und $\rho = 0,779$ variieren. Bezogen auf die Selbsteinschätzungen der Kinder im CDI variieren die ESDM-Werte mit den Lehrerpersönlichkeiten. Die Informiertheit der Lehrkräfte über die Selbstzuschreibung depressionsspezifischer Symptome im CDI durch die von ihnen unterrichteten Schulkinder variiert mit der Persönlichkeit. Über die an den Korrelationen orientierten Abschätzung der unterschiedlichen Leistungen verweisen darauf, dass einzelne Lehrkräfte eine fast 60%ige Übereinstimmung mit den Selbstzuschreibungen ihrer Schulkinder erreichen, während auf der Gegenseite der Verteilung andere Lehrkräfte eine etwa 32%ige geschätzte lineare Umkehrung der für ihre Kinder geltenden Rangreihen herstellen.

Damit aber ist insgesamt nachgewiesen, dass die Skala ESDM die unterschiedlichen Erkennensleistungen der Lehrkräfte gegenüber den Selbsteinschätzungen der durch die Lehrkräfte unterrichteten Schulkinder im CDI misst. Die Variation der ESDM-Werte wird sehr wahrscheinlich in Abhängigkeit von der unterschiedlichen Leistungsstärke der Urteile der Lehrkräfte determiniert. Insoweit fehlt dem Verfahren ESDM auch interpersonale Objektivität, wenn die einzelnen Lehrkräfte als die Personen betrachtet werden, die Beobachtungsaufgaben durchführen und deren Resultate vorlegen.

Diese Feststellungen treffen uneingeschränkt zu, wenn die Selbstzuschreibungen der Kinder im CDI als Kriterium benutzt werden.

Ob die Leistungsstärke der urteilenden Lehrkräfte bei der Identifizierung depressionsspezifischer Symptome an ihren Schulkindern bei Einbeziehung externer diagnostischer Kriterien insgesamt höher und interpersonal homogener ist, konnte wegen des Fehlens der dafür erforderlichen Außenkriterien nicht geprüft werden.

Die Plus-Gruppe der Lehrkräfte besteht aus den 24 Lehrerinnen und Lehrern der Klassen 9, 10, 14, 15, 25, 34, 40, 41, 44, 46, 48, 63, 66, 69, 72, 74, 79, 80, 92, 94, 95, 99, 106 und 107. Die mit

den Urteilen im Verfahren ESDM abgegebenen Aussagen über die von ihren Schülerinnen und Schülern im CDI mitgeteilten Selbsteinstufungen variieren hochwahrscheinlich nicht zufällig. Die Erkennensleistungen dieser Lehrkräfte sind signifikant höher als die der Lehrkräfte aus der Gruppe GG und ebenfalls signifikant unterschieden von den drei Lehrkräften, die mit Korrelationen über $\rho = -0,40$ als potentielle Angehörige der erwarteten Minusgruppe gelten könnten.

Die GG genannte Gruppe der Lehrkräfte umfasst dann 80 Lehrkräfte. Die für sie geltenden Korrelationen sind nicht signifikant von einer Grundgesamtheit mit dem Rho-Wert = 0 abgesetzt. Ihre über die Urteile in der Skala ESDM nachgewiesenen Erkennensleistungen gegenüber den Selbsteinstufungen ihrer Schulkinder im CDI variieren zufällig.

Obschon unter den beteiligten Lehrkräften nur zwei Leistungsgruppen signifikant voneinander getrennt werden konnten, erweist sich die Skala ESDM in der Untersuchung 2034GR mit den über sie erreichten Messwerte als von unterschiedlichen Erkennensleistungen der Lehrkräfte in starkem Maße abhängig. Die Einstufung der Abhängigkeit des Verfahrens ESDM von den unterschiedlichen Erkennensleistungen der Lehrkräfte darf als stark bezeichnet werden, weil die von dem Verfahren erreichte Übereinstimmung mit den Urteilen der Kinder zu fast 80% keine Signifikanz erreicht. Demgegenüber variieren die von etwa 20% der Lehrkräfte erreichten Übereinstimmungen mit den Selbsteinstufungen der Schülerinnen und Schüler im Bereich mittlerer und hoher Korrelationen.

Die nachstehenden Aussagen fassen diesen Aspekt der Analyse der ESDM Daten innerhalb der Untersuchung 2034GR zusammen.

12.29 AS: Die über das Verfahren ESDM von den beteiligten Lehrkräften nachgewiesenen Erkennensleistungen für die über CDI durch die von ihnen unterrichteten Kinder selbst berichteten depressionsspezifischen Symptomen zeigen signifikante Unterschiede zwischen Gruppen von Lehrkräften und einzelnen Lehrern und Lehrerinnen.

12.30 AS: Bezogen auf die Selbsteinschätzungen der Kinder im CDI unterscheidet das Verfahren ESDM besser zwischen den Erkennensleistungen der Lehrkräfte als zwischen den Selbsteinschätzungen der Kinder.

12.31 AS: Die auf die Selbsteinschätzung der Kinder bezogene Validität der Lehrerurteile im ESDM variiert mit den Lehrkräften. Die so nachgewiesene Abhängigkeit des ESDM von der stark unterschiedlichen individuellen Leistung der Lehrkräfte muss sich bei einer auf ein diagnostisches Außenkriterium bezogenen Analyse der Validität nicht zwangsläufig wiederholen.

12.32 AS: Die starke Abhängigkeit der auf die Selbsteinschätzung von Kindern im CDI bezogenen Validität der Lehrerurteils berechtigt allerdings zu der Hypothese, diese Abhängigkeit werde sich bei Validitätsprüfungen mit einem diagnostischen Außenkriterium fortsetzen.

Falls die Varianz in den festgestellten Korrelationen vornehmlich auf die Persönlichkeit einer jeden Lehrkraft und die ihr eigene Wahrnehmung zurückzuführen wäre, dann könnte man behaupten, dass drei Gruppen von Lehrkräften in der untersuchten Stichprobe auftreten. Eine erste Gruppe umfasste alle diejenigen, die sich über die von ihnen im Vergleich von ESDM und CDI bewiesenen Leistungen als empathisch gegenüber den Kindern in ihrem Unterricht darstellen. Sie sind über die Selbstwahrnehmungen dieser Kinder partiell informiert. Diese Gruppe empathischer Lehrkräfte wird über positive und signifikante Korrelationen zwischen ESDM und CDI gebildet. Eine zweite Gruppe von Lehrkräften erreicht keine signifikanten Korrelationen und weist damit eine relative Uninformiertheit über die Selbstwahrnehmungen der Kinder nach. Die Gruppe der relativ uninformatierten Lehrkräfte wird durch alle diejenigen gebildet, die über den Vergleich von ESDM und CDI keine signifikanten Korrelationen erreichen. Die dritte Gruppe von Lehrkräften

benutzt verzerrte Informationen über die Selbstwahrnehmung der von ihnen unterrichteten Kinder. Diese dritte Gruppe von Lehrkräften umfasst diejenigen, die mit negativen Korrelationen von $\rho = 0,40$ und höher auffallen.

Die Unterscheidung dieser Lehrergruppen unter Benutzung der Konzepte von Empathie und Informiertheit über die depressionsspezifische Selbstwahrnehmung der von ihnen unterrichteten Kinder orientiert sich an den über den Vergleich zwischen ESDM und CDI nachgewiesenen Leistungen der Lehrkräfte. Hohe und signifikante Korrelationen können Lehrkräfte erreichen, die in ihrem Urteil mit der Selbsteinstufung ihrer Schulkinder übereinstimmen. Sie beweisen sich damit als empathisch und informiert. Die entsprechenden psychischen Ausprägungen beider Konzepte in diesen Lehrkräften könnten einander wechselseitig bedingen. Empathische Lehrkräfte nehmen Informationen über die von ihnen betreuten Kinder möglicherweise interessierter auf als andere Lehrkräfte. Höhere Informiertheit über die Selbstwahrnehmung der Kinder begünstigt aber auch empathische Stellungnahmen. Die zweite Gruppe der relativ uninformierten Lehrkräfte beweist ihre Uninformiertheit durch das Verfehlen signifikanter Korrelationen. Die dritte Gruppe stellt die Verzerrung Ihrer Urteilsbasis ebenfalls selbst und dies durch mittlere oder hohe negative Korrelationen dar.

Diese Gruppierung der Lehrkräfte spekuliert nicht mit speziellen und für die Ausprägung der Leistung spezifischen persönlichen Eigenheiten. Die könnten allerdings durchaus vorliegen. Andererseits würden sich solche Leistungsunterschiede auch ergeben, wenn die Möglichkeiten der Aufnahme von Informationen über die Selbstwahrnehmung der Kinder durch externe Bedingungen erschwert oder verhindert wird. Eine Lehrkraft mit 2 Wochenstunden Fachunterricht in einer x. Schulklasse ist mit ihren Möglichkeiten zur Informationsnahme in einer anderen Situation als eine Lehrkraft mit 14 Stunden.

Die dargestellte Unterscheidung von drei über unterschiedliche Ausprägungen von Korrelationen zwischen ESDM und CDI gebildeten Gruppen von Lehrkräften lässt nach weiteren Informationen über diese Teilgruppen suchen. Da Faktorenanalysen als hypothesengenerierende Verfahren genutzt werden, bot sich die Durchführung einer auf diese drei Gruppen bezogene Faktorenanalyse an. Dazu musste allerdings die Definition der dritten Gruppe geändert werden. Für sie schien geboten, eine Zugehörigkeit zu dieser Gruppe vom Erreichen oder Überschreiten einer negativer Korrelation mit einer Ausprägung von $-0,40$ abhängig zu machen. Das erreichen in der untersuchten Stichprobe aber nur drei Lehrkräfte. Für die Faktorenanalyse und nur für sie wurde als Kriterium für die Gruppenzugehörigkeit das Auftreten einer negativen Korrelation eingesetzt. Das ebnet die Unterschiede zu den Mitgliedern der zweiten Gruppe ein. Das schien im Hinblick auf die Verteilung der negativen Korrelationen bei einem hypothesengenerierenden Verfahren hinnehmbar zu sein.

Wie aus Tabelle 12.12 zu ersehen ist, organisieren alle drei unterschiedenen Gruppen der beteiligten Lehrkräfte ihre Urteile auf der Skala ESDM in 4 Bereichen (Faktoren). Nach der Durchführung der getrennt für jede Gruppe angesetzten Hauptkomponentenanalyse mit anschließender Varimax-Rotation nach dem Eigenwertkriterium (Eigenwert > 1) haben sich vier Faktoren ergeben. Der erste Faktor (Eigenwert = 6,971) erklärt für die 1. Gruppe 43,568% Prozent der Gesamtvarianz, für die 2. Gruppe 33,71% und für Gruppe 3 dann 36,46% der Gesamtvarianz (Eigenwerte 5,39% und 5,76% entsprechend).

Alle vier Faktoren erbringen zusammen eine Aufklärungsrate von 70,05 der Varianz in der ersten Gruppe, von 62,25% in der zweiten Gruppe und 61,63% in der dritten Gruppe.

Tabelle 12.12: Faktorenanalyse der mit ESDM abgegebenen Urteile der Lehrkräfte für drei Gruppen von Lehrkräften

	Gruppe 1 signifikante positive rho-Werte	Gruppe 2 nicht signifikante positive rho-Werte	Gruppe 3 nicht signifikante negative rho-Werte
<i>Faktor 1</i>	(15) zufrieden stellende Leistung (14) aktive Mitarbeit in der Schule (5) versteht den Unterricht (2) fällt schwer, aufmerksam zu sein	(15) zufrieden stellende Leistung (5) versteht den Unterricht (14) aktive Mitarbeit in der Schule 2 fällt schwer, aufmerksam zu sein 9 fühlt sich weniger Wert	(15) zufrieden stellende Leistung (5) versteht den Unterricht 2 fällt schwer, aufmerksam zu sein (14) aktive Mitarbeit in der Schule 9 fühlt sich weniger Wert 12 kann sich nicht entscheiden 11 neigt zu Selbstvorwürfen
<i>Faktor 2</i>	11 neigt zu Selbstvorwürfen 7 sieht müde aus 12 kann sich nicht entscheiden 13 schwankende Stimmungslage 1 fühlt sich traurig 9 fühlt sich weniger Wert	1 fühlt sich traurig 13 schwankende Stimmungslage 4 weint 7 sieht müde aus 11 neigt zu Selbstvorwürfen 12 kann sich nicht entscheiden	(16) spielt mit den anderen 3 ist in der Pause allein (8) hat Spaß in der Schule (6) wird in der Klasse akzeptiert
<i>Faktor 3</i>	(3) ist in der Pause allein (16) spielt mit den anderen (6) wird in der Klasse akzeptiert (8) hat Spaß in der Schule	(8) hat Spaß in der Schule (6) wird in der Klasse akzeptiert (16) spielt mit den anderen 3 ist in der Pause allein	13 schwankende Stimmungslage 1 fühlt sich traurig 4 weint 7 sieht müde aus
<i>Faktor 4</i>	(10) widerspricht und streitet (4) weint	10 widerspricht und streitet	10 widerspricht und streitet

Extraction Method: Principal Component Analysis. Rotations Method: Varimax with Kaiser Normalization.

Aus der vorangegangenen Tabelle ist zu ersehen, welche Items der Skala ESDM von den drei unterschiedenen Gruppen an Lehrkräften in welche, durch die gruppenspezifische Nutzung manifestierte, assoziativ verbundene Itemgruppen gestellt worden sind. Mit vier solchen Faktoren wird das Auftreten der Items im jeweiligen Gesamtaufkommen in den drei Gruppen von Lehrkräften erfasst. Da die drei Gruppen von Lehrkräften durch ihre signifikant unterschiedlichen Leistungen im Benennen der Selbstzuweisungen der Kinder im CDI mehr oder weniger erfolgreich sind, könnte vermutet werden, die drei unterschiedlich erfolgreichen Gruppen von Lehrkräften nutzten unterschiedlich strukturierte Urteilsprozesse. Dem entsprechend wird hier der Hypothese gefolgt, die unterschiedlichen Leistungen der drei Gruppen von Lehrkräften träten im Zusammenhang mit den aus unterschiedlichen Items gebildeten Faktoren auf. Solche Unterschiede müssten in der Itembesetzung durch die identifizierten Faktoren zu finden sein. Diese Itembesetzungen finden sich in der nachstehenden Tabelle aufgelistet. Da die Gruppe 1 der Lehrkräfte den Selbstzuweisungen der Kinder mit ihren Urteilen näher gekommen ist als die

beiden anderen Gruppen, wird die Itemnutzung der Gruppe 1 auf der für sie gefundenen Itemebene als Modell einer relativ erfolgreichen Beurteilungsstruktur bei der Arbeit mit der Skala ESDM gesetzt.

Die Gruppen 2 und 3 könnten mit dieser Struktur übereinstimmen, oder auch von der abweichen. Bestimmt wird, als Abweichung jeden Fall zu zählen, bei dem in der Itemliste der nachgeordneten Gruppen von Lehrkräften ein Item nicht den als Modell aus Gruppe 1 genommenen Faktoren angehört, sondern stattdessen in einem anderen platziert ist. Die auf diese Weise festgestellten Abweichungen finden sich in den die Abweichungen aufführenden Zeilen der Tabelle als abweichende Itemmenge.

Tabelle 12.13: Abweichende Itemmengen in der Faktorenstruktur der ESDM-Urteile dreier Gruppen von Lehrkräften

Lehrkräfte	Nummer der ESDM Item in vier Faktoren				Summe
	Faktor 1	Faktor 2	Faktor 3	Faktor 4	
Gruppe 1	15, 14, 5, 2	11, 7, 12, 13, 1, 9	3, 16, 6, 8	10, 4	16
Gruppe 2	15, 5, 14, 2, 9,	1, 13, 4, 7, 11, 12	8, 6, 16, 3	10	16
Gruppe 3	15, 5, 2, 14, 9, 12, 11	16, 3, 8, 6	13, 1, 4, 7	10	16
<i>Abweichungen</i>					
Gruppe 2 von 1	1	2	0	1	4*(25%)
Gruppe 3 von 1	3	4	4	1	12*(75%)

* = Das beobachtete Verhältnis ist signifikant gegenüber einer Gleichverteilungshypothese

Qualitative Unterschiede bestehen zwischen den definierten Gruppen von Lehrkräften in der Nutzung der Items in den für die Teilgruppen analysierten Faktoren. Da die Gruppe 1 signifikante Korrelationen zwischen ESDM und CDI erreichte, wird deren Nutzung der Items in den Faktoren 1, 2 und 3 als Kriterium gesetzt. Bestimmt wurde die Summe der Items, die in Gruppen 2 und 3 in ihren Itemlisten pro Faktor anders platziert sind als in Gruppe 1. Diese Zahl wird als Abweichung gegenüber Gruppe 1 bezeichnet. Aus der obigen Tabelle ergibt sich, dass die Gruppe 2 gegenüber der Gruppe 1 auf 3 Abweichungen kommt, während für die Gruppe 3 insgesamt 11 Abweichungen festgestellt wurden. Der mitgeteilte Sachverhalt kann als Hinweis auf unterschiedliche Strukturen in den Prozessen der Urteilsfindung verstanden werden. Die aus den einzelnen Faktoren durch die Gruppen 1 und 2 pro Faktor beachteten Items gleichen einander weitgehend. Die von der Gruppe 3 benutzte Verteilung der Items auf die Faktoren zeigt demgegenüber 11 Abweichungen.

Die für die Realisierung von ESDM Urteilen jeweils erforderliche Itemwahl ist in der Gruppe 3 der Lehrkräfte möglicherweise anders organisiert als in den Vergleichsgruppen.

Die Hypothese ist gerechtfertigt, dass die Lehrerinnen und Lehrer der Gruppe 3 ihre Urteile mit einem anders als in Gruppe 1 und 2 strukturierten Urteilsprozess herstellen. Wenn dann aber Gruppe 1 und 2 untereinander in der Zahl der pro Faktor beachteten Items ähneln, ist die Nachfrage, warum sie dennoch, bezogen auf das Kriterium der Übereinstimmung mit den Selbsteinstufungen der Kinder über den CDI, signifikant unterschiedliche Urteile produzieren, notwendig. Die Antwort darauf sollte sich, hinreichende Informationen vorausgesetzt, über die Betrachtung der dem Urteilsprozess im ESDM unterliegenden Handlungsstruktur erfassen lassen.

Die genannte Voraussetzung wird allerdings nicht erfüllt. So ist zu prüfen, welche Bedingungen erschlossen werden könnten.

Geleitet von der jeweils individuell übernommenen Aufgabe, organisiert sich die Handlungsstruktur, den allgemeinen Prinzipien der psychischen Handlungsregulation, beispielsweise sensu Hacker (1986), folgend, über die Teilziele einbeziehenden Zielhierarchien, über die Auswahl von Wegen, Mitteln und weiteren psychischen Instanzen, zur Planung, Ausführung und Evaluation. Für die über ESDM abzugebenden Urteile der Lehrkräfte folgt die Handlungsregulation mit Sicherheit sowohl allgemeinen Prinzipien als auch individuellen Bedingungen. Von daher wird hier zum einen der These gefolgt, die zu betrachtenden Unterschiede in der Qualität der ESDM Urteile variierten nicht unwesentlich mit interpersonalen Unterschieden in der Urteilsituation. Das Urteilsverhalten wird mit der Interaktion von Person und Situation ausgebildet. Aber gemeinsam werden alle Lehrkräfte mit der Zielsetzung konfrontiert, aus dem mit dem ESDM vorgegebenen 48 Urteilsmöglichkeiten das für jeweils ein bestimmtes Kind zutreffende Urteilsmuster von 16 ESDM Aussagen zu suchen. Die den Selbstzuschreibungen der Kinder entsprechende korrekte Wahl aus den 48 möglichen Entscheidungen hat für jedes Item eine Wahrscheinlichkeit von 0,33 für erratene Übereinstimmungen (Treffer). Schon nach 5 Items wäre das Erraten der für ein Kind zutreffenden Kombination statistisch extrem unwahrscheinlich. So ließe sich eine Setzung denken, nach der Lehrkräfte, die pro Kind 6 oder mehr richtige ESDM Urteile abgeben, als - graduell unterschiedlich – über die Erlebenseinhalte des Kindes informiert zu bezeichnen wären.

Die von den Lehrkräften in der Gruppe 1 erreichten signifikanten Korrelationen mit den Urteilen der Kinder erreichen Korrelationen, die eine bis an 50% geschätzte Aufklärung der Varianz ihrer Urteile bieten. Dieser Prozentsatz erfasst die korrekten Urteile. Damit ist die Aussage unabweisbar, dass die Lehrkräfte in Gruppe 1 individuell unterschiedliches Wissen über depressionsspezifische Befindlichkeiten ihrer Schüler und Schülerinnen verfügten. Auch Casanovas (2003) war auf die persönlichkeitspezifische Abhängigkeit der ESDM Urteile aufmerksam geworden, vermochte sie aber wegen eines dafür zu geringen Umfanges ihrer Stichprobe an Lehrkräften nicht zu sichern.

Da dies handlungspsychologisch, von Zufallstreffern abgesehen, nicht anders sein kann, ist die Hypothese empirisch gestützt, dass ein wesentlicher Unterschied zwischen den drei Gruppen der Lehrkräfte ein signifikanter Unterschied in der Qualität der treffsicheren Zuordnung von ESDM Urteilen zu Schulkindern ist. Die Gruppe 1 der Lehrkräfte zeichnet sich durch ein signifikant höheres Maß an Treffsicherheit bei der Zuordnung des rechten Urteils zum rechten Kind aus. Die Gruppe 2 erreicht dieses Qualitätsniveau nicht. Die Gruppe 3 bleibt noch tiefer unter dem Qualitätsniveau der anderen Gruppen und fällt durch ein beachtenswertes Maß an Umkehrung der bei den Kindern geltenden Selbstzuweisungen auf, was hier als Verzerrung und Verkehrung der für die Kinder geltenden Realität bezeichnet wird.

Als Ergebnis dieses Analyseschrittes werden die nachfolgenden Aussagen vermerkt:

12.33 AS: Die unterschiedlichen Erkennensleistungen der Lehrkräfte bei der Aufgabe, über die Skala ESDM die von den Schulkindern ihrer eigenen Schulklassen im CDI gegebenen Selbstzuweisungen wiederzugeben, variieren mit den Lehrkräften und lassen sich in der Studie 2034GR drei Leistungsklassen zuordnen.

12.34 AS: An Ursachen für die unterschiedlichen Leistungen der Lehrkräfte sind vermutlich zwei funktionale Besonderheiten zu beachten.

12.35 AS: Negative Korrelationen zu den Selbstzuweisungen der Kinder stellen sich da ein, wo die Lehrkräfte zum einen im Prozess der Zuordnung ihrer Urteile einer um über 50% von der faktoriellen Itemstruktur der erfolgreicherer Lehrer und Lehrerinnen abweichenden Vorgehensweise folgen. Dazu tritt zum anderen ein deutliches Defizit an zutreffendem Wissen

über die subjektive Befindlichkeit der Kinder auf, dessen funktionale Schwäche durch realitätsverzerrende Umkehrungen der Aussagen der Schulkinder auffällt.

12.36 AS: Für die Leistungsunterschiede zwischen der Gruppe der nur schwach informierten Lehrkräfte und den besser informierten Lehrerinnen und Lehrern ist der Wissensstand über die Befindlichkeit der Schülerinnen und Schüler hoch wahrscheinlich die den Unterschied hervorbringende Variable.

12.37 AS: Warum sich die signifikanten Wissensunterschiede zwischen den Lehrkräften ergeben, war mit dem Datenmaterial der Untersuchung 2034GR nicht zu analysieren.

12.38 AS: Der Anteil der Lehrkräfte, die in der Untersuchung 2034GR nachweislich über auf die Kinder ihrer Schulklasse bezogenes depressionsspezifisches Wissen verfügen, umfasst etwa 20% der gesamten Lehrerschaft in dieser Studie.

12.39 AS: Die mit dem Verfahren ESDM durch die Lehrkräfte abgegebenen Urteile über das Auftreten spezieller depressionsspezifischer Symptome bei den von ihnen unterrichteten Schulkindern, erreichen insgesamt, wenn die Selbsteinschätzungen der Kinder im CDI als Kriterium eingesetzt werden, für die gesamte untersuchte Stichprobe und über Korrelationen bestimmt, nur niedrige Übereinstimmungen.

12.40 AS: Die Annahme, dass die Lehrkräfte insgesamt keine signifikante höhere Übereinstimmung erreichen, erweist sich bei näherer Analyse als nicht zutreffen.

12.41 AS: Die Gesamtverteilung der Korrelationen zwischen ESDM und CDI variiert von rho -0,57 über den Nullpunkt hinweg bis + 0,77.

12.42 AS: Die Ergebnisse der für die ESDM Daten dieser Untersuchung durchgeführten Analysen sprechen zwingend für die Annahme von die Validität des Verfahrens beeinflussenden Besonderheiten der Skala ESDM. Solche Besonderheiten dürften mutmaßlich auch für andere Verfahren nachweisbar sein, über die eine Fremdbeurteilung von Kindern treffende psychische Symptome versucht wird. Diese Besonderheit entsteht durch die Nichtbeachtung der zwangsläufig entstehenden Verdoppelung des Problems der Messqualität.

12.43 AS: Im Falle der Anwendung des ESDM wird die Messqualität dieses Instrumentes geprüft. Ungeprüft bleibt die Messqualität, die jeder an einer ESDM Studie beteiligte Lehrer und jede Lehrerin in die Messungen einbringt.

12.44 AS: Die Skala ESDM könnte statt als Verfahren zur Identifizierung von an Kindern auftretenden Symptomen im Verbund mit dem CDI oder dem durch Kinder ausgefüllten Doppel des ESDM als Verfahren zur Messung der individuellen Leistungsunterschiede von Lehrkräften im Erkennen von Selbstwahrnehmungen der von ihnen unterrichteten Kinder genutzt werden.

12.45 AS: Durch teilnehmende Beobachtungen bei der Durchführung der Untersuchung scheint die Annahme gerechtfertigt, dass ein Teil der Lehrkräfte die von ihnen verlangte Bearbeitung der A Skala ESDM als unerwünschte Belastung oder aber als relativ bedeutungslos auffasste und sich nicht motiviert erlebte, zutreffende ESDM Urteile zu suchen.

12.5.2 Übereinstimmung zwischen Urteilen der Lehrkräfte im ESDM und Selbstangaben der Kinder zu ausgewählten CDI Items

Die im Arbeitsteil 12.5.1 dargestellten Korrelationswerte zwischen Urteilen der Lehrkräfte im ESDM und Selbstangaben der Kinder können als Nachweise dafür verstanden werden, dass es in der Gesamtstichprobe der Lehrkräfte ein interpersonal unterschiedliches Wissen über die symptombezogene Selbstwahrnehmung der von den Lehrkräften unterrichteten Kinder besteht.

Die direkte Korrespondenz von 9 ESDM-Aussagen mit 9 CDI-Fragen erlaubt, die Kompetenz der Lehrkräfte an der Übereinstimmung ihrer ESDM Urteile mit den Selbsteinstufungen der Kinder bei diesen Items zu prüfen.

Die nachstehende Tabelle kreuzt die Selbsteinstufungen der Kinder bei 9 CDI-Items mit den im Verfahren ESDM korrespondierenden 9 Items.

Tabelle 12.14: Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 1

CDI 1 "Sadness"		ESDM 1 "fühlt sich traurig"			Gesamt
		"selten"	"manchmal"	"sehr oft"	
"selten"	% within CDI 1	59,2%	39,0%	1,7%	100,0%
	% within ESDM 1	88,3%	83,8%	60,4%	85,8%
	Adjusted Residual	3,7	-2,1	-5,1	
"manchmal"	% within CDI 1	48,6%	45,3%	6,1%	100,0%
	% within ESDM 1	10,7%	14,3%	31,3%	12,6%
	Adjusted Residual	-3,0	1,8	3,9	
"sehr oft"	% within CDI 1	38,7%	48,4%	12,9%	100,0%
	% within ESDM 1	1,1%	1,9%	8,3%	1,6%
	Adjusted Residual	-2,1	1,0	3,8	
Gesamt	% within CDI 1	57,5%	40,0%	2,5%	100,0%
	% within ESDM 1	100,0%	100,0%	100,0%	100,0%

Chi-Square

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	39,238(a)	4	,000
Likelihood Ratio	29,221	4	,000
Linear-by-Linear Association	25,307	1	,000
N of Valid Cases	1941		

a 1 cells (11,1%) have expected count less than 5. The minimum expected count is ,77

Für 59,2% aller Kinder, die bei Frage 1 des CDI angegeben haben, dass sie "selten" traurig sind, haben ihre Lehrkräfte dies bestätigt. Weitere 39% dieser Kinder beschreiben die Lehrkräfte aber als „manchmal“ traurig. Schließlich werden auch etwa 2% dieser fast immer frohen CDI-Kinder als „sehr oft“ traurig eingestuft.

Von allen Kindern, die im CDI geantwortet haben, „manchmal“ traurig zu sein, nehmen das 45,3% ihrer Lehrkräfte auch an. Nach deren Wahrnehmung ist aber etwa die andere Hälfte dieser Kinder eher „selten“ traurig.

Die Gruppe der Kinder, die sich im CDI selbst als sehr oft traurig einordnet, sind nur 12,8% der Lehrkräfte übereinstimmend dieser Meinung.

Die nachstehende kurze Tabelle fasst die richtigen Urteile der Lehrkräfte für die CDI-Frage 1 zusammen.

Tabelle 12.15: Übersicht über die bei CDI Frage 1 durch die Lehrkräfte richtig zugeordneten Prozentsätze an Kindern (prozentuale Trefferliste)

CDI-Kategorie	selten	manchmal	sehr oft
PrZ	59,2%	45,3%	12,9%

Zeichenerklärung: PrZ = Prozentsatz richtiger Zuordnungen durch die Lehrkräfte zum Grundwert der durch die Kinder erfolgten Selbstzuschreibung

Nachfolgend finden sich für jede der CDI-Fragen eine ausführliche Übersichtstabelle und eine als prozentuale Trefferliste ausgelegte Kurztabelle.

Tabelle 12.16: Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 10

CDI 10 "crying spells"		ESDM 4 "weint"			Gesamt
		"selten"	"manchmal"	"sehr oft"	
"selten"	% within CDI 10	81,7%	17,7%	,7%	100,0%
	% within ESDM 4	85,2%	79,2%	84,6%	84,1%
	Adjusted Residual	2,7	-2,8	,1	
"manchmal"	% within CDI 10	76,3%	22,9%	,8%	100,0%
	% within ESDM 4	13,3%	17,2%	15,4%	14,0%
	Adjusted Residual	-1,9	1,9	,1	
"sehr oft"	% within CDI 10	63,9%	36,1%	,0%	100,0%
	% within ESDM 4	1,5%	3,7%	,0%	1,9%
	Adjusted Residual	-2,6	2,7	-,5	
Gesamt	% within CDI 10	80,6%	18,7%	,7%	100,0%
	% within ESDM 4	100,0%	100,0%	100,0%	100,0%

Chi-Square

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	11,637(a)	4	,020
Likelihood Ratio	10,594	4	,032
Linear-by-Linear Association	8,674	1	,003
N of Valid Cases	1894		

a 2 cells (22,2%) have expected count less than 5. The minimum expected count is ,25.

Tabelle 12.17: Übersicht über die bei CDI Frage 10 durch die Lehrkräfte richtig zugeordneten Prozentsätze an Kindern (prozentuale Trefferliste)

CDI-Kategorie	selten	manchmal	sehr oft
PrZ	81,7%	22,9%	0 %

Zeichenerklärung: PrZ = Prozentsatz richtiger Zuordnungen durch die Lehrkräfte zum Grundwert der durch die Kinder erfolgten Selbstzuschreibung

Tabelle 12.18: Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 17

CDI 17 "fatigue"		ESDM 7 "sieht müde aus"			Gesamt
		"selten"	"manchmal"	"sehr oft"	
"selten"	% within CDI 17	62,9%	35,1%	2,0%	100,0%
	% within ESDM 7	53,1%	47,4%	39,2%	50,6%
	Adjusted Residual	2,7	-2,2	-1,7	
"manchmal"	% within CDI 17	58,3%	38,7%	3,0%	100,0%
	% within ESDM 7	41,5%	44,1%	49,0%	42,7%
	Adjusted Residual	-1,3	1,0	,9	
"sehr oft"	% within CDI 17	48,5%	46,9%	4,6%	100,0%
	% within ESDM 7	5,4%	8,4%	11,8%	6,7%
	Adjusted Residual	-2,8	2,3	1,5	
Gesamt	% within CDI 17	59,9%	37,4%	2,6%	100,0%
	% within ESDM 7	100,0%	100,0%	100,0%	100,0%

Chi-Square

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	13,328(a)	4	,010
Likelihood Ratio	13,036	4	,011
Linear-by-Linear Association	12,401	1	,000
N of Valid Cases	1932		

a 1 cells (11,1%) have expected count less than 5. The minimum expected count is 3,43

Tabelle 12.19: Übersicht über die bei CDI Frage 17 durch die Lehrkräfte richtig zugeordneten Prozentsätze an Kindern (prozentuale Trefferliste)

CDI-Kategorie	selten	manchmal	sehr oft
PrZ	62,9%	38,7%	4,6 %

Zeichenerklärung: PrZ = Prozentsatz richtiger Zuordnungen durch die Lehrkräfte zum Grundwert der durch die Kinder erfolgten Selbstzuschreibung

Tabelle 12.20: Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 21

CDI 21 "school dislike"		ESDM 8 "hat Spaß in der Schule"*			Gesamt
		"selten"	"manchmal"	"sehr oft"	
"selten"	% within CDI 21	79,1%	18,3%	2,6%	100,0%
	% within ESDM 8	78,8%	72,7%	60,3%	77,0%
	Adjusted Residual	3,4	-2,1	-3,2	
"manchmal"	% within CDI 21	72,9%	22,1%	5,0%	100,0%
	% within ESDM 8	18,0%	21,8%	28,6%	19,1%
	Adjusted Residual	-2,2	1,4	1,9	
"sehr oft"	% within CDI 21	63,0%	27,4%	9,6%	100,0%
	% within ESDM 8	3,2%	5,5%	11,1%	3,9%
	Adjusted Residual	-3,0	1,8	3,0	
Gesamt	% within CDI 21	77,3%	19,4%	3,4%	100,0%
	% within ESDM 8	100,0%	100,0%	100,0%	100,0%

*umgepolte Frage

Chi-Square

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	21,643(a)	4	,000
Likelihood Ratio	18,424	4	,001
Linear-by-Linear Association	19,742	1	,000
N of Valid Cases	1873		

a 1 cells (11,1%) have expected count less than 5. The minimum expected count is 2,46.

Tabelle 12.21: Übersicht über die bei CDI Frage 21 durch die Lehrkräfte richtig zugeordneten Prozentsätze an Kindern (prozentuale Trefferliste)

CDI-Kategorie	selten	manchmal	sehr oft
PrZ	79,1%	22,1%	9,6 %

Zeichenerklärung: PrZ = Prozentsatz richtiger Zuordnungen durch die Lehrkräfte zum Grundwert der durch die Kinder erfolgten Selbstzuschreibung

Tabelle 12.22: Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 24

CDI 24 "Self depreciation"		ESDM 9 "fühlt sich weniger Wert"			Gesamt
		"selten"	"manchmal"	"sehr oft"	
"selten"	% within CDI 24	68,0%	29,0%	3,0%	100,0%
	% within ESDM 9	33,5%	26,4%	21,0%	30,6%
	Adjusted Residual	3,5	-2,8	-1,9	
"manchmal"	% within CDI 24	60,8%	35,3%	3,9%	100,0%
	% within ESDM 9	55,8%	59,7%	50,6%	56,9%
	Adjusted Residual	-1,2	1,8	-1,2	
"sehr oft"	% within CDI 24	52,8%	37,3%	9,9%	100,0%
	% within ESDM 9	10,7%	13,9%	28,4%	12,5%
	Adjusted Residual	-3,1	1,3	4,4	
Gesamt	% within CDI 24	62,0%	33,7%	4,4%	100,0%
	% within ESDM 9	100,0%	100,0%	100,0%	100,0%

Chi-Square

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	31,367(a)	4	,000
Likelihood Ratio	27,557	4	,000
Linear-by-Linear Association	23,524	1	,000
N of Valid Cases	1859		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 10,15

Tabelle 12.23: Übersicht über die bei CDI Frage 24 durch die Lehrkräfte richtig zugeordneten Prozentsätze an Kindern (prozentuale Trefferliste)

CDI-Kategorie	selten	manchmal	sehr oft
PrZ	68%	35,3%	9,9 %

Zeichenerklärung: PrZ = Prozentsatz richtiger Zuordnungen durch die Lehrkräfte zum Grundwert der durch die Kinder erfolgten Selbstzuschreibung

Tabelle 12.24: Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 27

CDI 27 "fighting"		ESDM 10 "widerspricht und streitet"			Gesamt
		"selten"	"manchmal"	"sehr oft"	
"selten"	% within CDI 27	70,7%	24,7%	4,6%	100,0%
	% within ESDM 10	83,6%	78,8%	69,5%	81,6%
	Adjusted Residual	3,4	-1,9	-3,3	
"manchmal"	% within CDI 27	63,2%	28,8%	8,0%	100,0%
	% within ESDM 10	15,2%	18,8%	24,8%	16,7%
	Adjusted Residual	-2,5	1,5	2,3	
"sehr oft"	% within CDI 27	47,1%	35,3%	17,6%	100,0%
	% within ESDM 10	1,2%	2,4%	5,7%	1,8%
	Adjusted Residual	-2,8	1,3	3,2	
Gesamt	% within CDI 27	69,1%	25,5%	5,4%	100,0%
	% within ESDM 10	100,0%	100,0%	100,0%	100,0%

Chi-Square

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	23,155(a)	4	,000
Likelihood Ratio	19,336	4	,001
Linear-by-Linear Association	20,002	1	,000
N of Valid Cases	1939		

a 1 cells (11,1%) have expected count less than 5. The minimum expected count is 1,84

Tabelle 12.25: Übersicht über die bei CDI Frage 27 durch die Lehrkräfte richtig zugeordneten Prozentsätze an Kindern (prozentuale Trefferliste)

CDI-Kategorie	selten	manchmal	sehr oft
PrZ	70,7%	28,8%	17,6 %

Zeichenerklärung: PrZ = Prozentsatz richtiger Zuordnungen durch die Lehrkräfte zum Grundwert der durch die Kinder erfolgten Selbstzuschreibung

Tabelle 12.26: Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 8

CDI 8 "self-blame"		ESDM 11 "neigt zu Selbstvorwürfen "			Gesamt
		"selten"	"manchmal"	"sehr oft"	
"selten"	% within CDI 8	82,8%	16,5%	,8%	100,0%
	% within ESDM 11	69,6%	72,6%	62,5%	70,0%
	Adjusted Residual	-,9	1,1	-,7	
"manchmal"	% within CDI 8	84,3%	14,8%	,9%	100,0%
	% within ESDM 11	23,2%	21,3%	25,0%	22,9%
	Adjusted Residual	,7	-,7	,2	
"sehr oft"	% within CDI 8	84,8%	13,6%	1,5%	100,0%
	% within ESDM 11	7,2%	6,1%	12,5%	7,1%
	Adjusted Residual	,5	-,7	,8	
Gesamt	% within CDI 8	83,3%	15,9%	,9%	100,0%
	% within ESDM 11	100,0%	100,0%	100,0%	100,0%

Chi-Square

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1,998(a)	4	,736
Likelihood Ratio	1,902	4	,754
Linear-by-Linear Association	,376	1	,540
N of Valid Cases	1865		

a 2 cells (22,2%) have expected count less than 5. The minimum expected count is 1,13.

Tabelle 12.27: Übersicht über die bei CDI Frage 8 durch die Lehrkräfte richtig zugeordneten Prozentsätze an Kindern (prozentuale Trefferliste)

CDI-Kategorie	selten	manchmal	sehr oft
PrZ	72,8%	72,6%	1,5 %

Zeichenerklärung: PrZ = Prozentsatz richtiger Zuordnungen durch die Lehrkräfte zum Grundwert der durch die Kinder erfolgten Selbstzuschreibung

Tabelle 12.28: Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 13

CDI 13 "indeciviness"		ESDM 12 "kann sich nicht entscheiden"			Gesamt
		"selten"	"manchmal"	"sehr oft"	
"selten"	% within CDI 13	51,7%	44,4%	3,9%	100,0%
	% within ESDM 12	25,5%	25,7%	21,1%	25,4%
	Adjusted Residual	,1	,3	-1,0	
"manchmal"	% within CDI 13	51,8%	43,5%	4,7%	100,0%
	% within ESDM 12	69,2%	68,1%	67,8%	68,7%
	Adjusted Residual	,5	-,4	-,2	
"sehr oft"	% within CDI 13	46,0%	45,1%	8,8%	100,0%
	% within ESDM 12	5,3%	6,1%	11,1%	5,9%
	Adjusted Residual	-1,2	,3	2,1	
Gesamt	% within CDI 13	51,4%	43,8%	4,7%	100,0%
	% within ESDM 12	100,0%	100,0%	100,0%	100,0%

Chi-Square

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	5,538(a)	4	,236
Likelihood Ratio	4,762	4	,313
Linear-by-Linear Association	1,362	1	,243
N of Valid Cases	1905		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 5,34.

Tabelle 12.29: Übersicht über die bei CDI Frage 13 durch die Lehrkräfte richtig zugeordneten Prozentsätze an Kindern (prozentuale Trefferliste)

CDI-Kategorie	selten	manchmal	sehr oft
PrZ	51,7%	25,7%	8,8 %

Zeichenerklärung: PrZ = Prozentsatz richtiger Zuordnungen durch die Lehrkräfte zum Grundwert der durch die Kinder erfolgten Selbstzuschreibung

Tabelle 12.30: Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 23

CDI 23 "school performance decrement"		ESDM 15 "zufrieden stellende Leistung"			Gesamt
		"selten"	"manchmal"	"sehr oft"	
"selten"	% within CDI 23	67,1%	26,2%	6,7%	100,0%
	% within ESDM 15	77,3%	45,1%	26,8%	58,9%
	Adjusted Residual	16,9	-8,9	-11,9	
"manchmal"	% within CDI 23	31,3%	46,9%	21,8%	100,0%
	% within ESDM 15	19,5%	43,7%	47,2%	31,9%
	Adjusted Residual	-11,9	8,1	6,0	
"sehr oft"	% within CDI 23	17,3%	41,3%	41,3%	100,0%
	% within ESDM 15	3,1%	11,2%	26,1%	9,3%
	Adjusted Residual	-9,5	2,1	10,6	
Gesamt	% within CDI 23	51,1%	34,2%	14,7%	100,0%
	% within ESDM 15	100,0%	100,0%	100,0%	100,0%

Chi-Square

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	354,649(a)	4	,000
Likelihood Ratio	350,926	4	,000
Linear-by-Linear Association	329,526	1	,000
N of Valid Cases	1933		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 26,30.

Tabelle 12.31: Übersicht über die bei CDI Frage 23 durch die Lehrkräfte richtig zugeordneten Prozentsätze an Kindern (prozentuale Trefferliste)

CDI-Kategorie	selten	manchmal	sehr oft
PrZ	67,1%	46,9%	41,3 %

Zeichenerklärung: PrZ = Prozentsatz richtiger Zuordnungen durch die Lehrkräfte zum Grundwert der durch die Kinder erfolgten Selbstzuschreibung

Die für jede der 9 verwendeten CDI-Fragen vorangehend dargestellte prozentuale Trefferliste wird nachfolgend in eine für alle 9 CDI-Fragen geltende prozentuale Trefferliste eingefügt.

Tabelle 12.32: Prozentualer Trefferliste für alle im ESDM verwendeten 9 CDI-Items

CDI-Kategorie	Selten	manchmal	sehr oft
Item 1, PrZ	59,2%	45,3%	12,9%
Item 10, PrZ	81,7%	22,9%	0 %
Item 17, PrZ	62,8%	38,7%	4,6 %
Item 21, PrZ	79,1%	22,1%	9,6 %
Item 24, PrZ	68%	35,3%	9,9 %
Item 27, PrZ	70,7%	28,8%	17,6 %
Item 8, PrZ	72,8%	72,6%	1,5 %
Item 13, PrZ	51,7%	25,7%	8,8 %
Item 23, PrZ	67,1%	46,9%	41,3 %

Zeichenerklärung: PrZ = Prozentsatz richtiger Zuordnungen durch die Lehrkräfte zum Grundwert der durch die Kinder erfolgten Selbstzuschreibung

Aus der Tabelle 12.32 ist zu ersehen, mit welchen Prozentsätzen die Gesamtgruppe aller Lehrkräfte mit ihren ESDM Urteilen die Selbstzuweisungen von Symptomen durch die Kinder im CDI trifft. Die nachstehenden Aussagen fassen die Ergebnisse zusammen.

12.46 AS: Die Übereinstimmung der über ESDM Urteile von den Lehrkräften gegebenen Einschätzungen des Auftretens von durch CDI Items verbalisierten depressiver Symptome mit den Selbstzuweisungen der Kinder im CDI variiert von Item zu Item und weiterhin auch mit den Ausprägungsgraden „selten“, „manchmal“ und „sehr oft“.

12.47 AS: Der Ausprägungsgrad „selten“ wird bei allen 9 benutzten CDI Items mit Prozentsätzen zwischen 50% und etwa 80% zutreffend eingeschätzt.

12.48 AS: Der Ausprägungsgrad „manchmal“ wird signifikant seltener richtig angegeben. Itemübergreifend variieren die korrekten Lehrerurteile zwischen 22% und 72%.

12.49 AS: Die Selbsteinschätzungen der Kinder zum Ausprägungsgrad „sehr oft“ werden durch die Lehrerurteile signifikant seltener als bei den anderen Ausprägungsgraden eingestellt. Die Übereinstimmungsquoten zwischen ESDM und CDI variieren zwischen 0% und 41,3%, dabei bleiben aber 8 der 9 von der Gruppe aller Lehrkräfte abgegebenen Urteilen unter 20% Übereinstimmung.

12.6 Weitere Einflussfaktoren auf die Lehrurteile im ESDM

Bei der Suche nach weiteren bislang noch nicht in Betracht gezogenen potentiellen Einflussfaktoren auf das Urteil der Lehrkräfte im Verfahren ESDM wurden die nachstehend dargestellten Überprüfungen durchgeführt.

Den bisher durchgeführten Analysen über Zusammenhänge zwischen Urteilen der Lehrkräfte in der Skala ESDM und Selbsteinschätzungen der Kinder im CDI können noch einige hinzugefügt werden, die möglicherweise weitere Aufschlüsse darüber bieten, wovon die Urteile der Lehrkräfte beeinflusst werden. Anzusetzen sind solche Analysen allerdings nur an Daten, die in dieser Untersuchung ohnehin vorliegen und eben nicht an eventuell zu wünschendem zusätzlichen Datenmaterial.

Aus den unterschiedlichen Informationen, die mit der Durchführung der Untersuchung bekannt geworden sind, wurden solche ausgewählt, die sich den nachstehenden Gruppen zuordnen lassen:

1. Informationen über spezielle Merkmale der beteiligten Lehrkräfte, so über deren Geschlecht und über ihre Unterrichtsfächer
2. Informationen über persönliche Merkmale der beteiligten Schulkinder, verfügbar sind Angaben über deren Geschlecht und das Alter
3. Schulisch wirksame Bedingungen, die indirekt oder direkt die Interaktionen zwischen Lehrkräften und Schülern beeinflussen könnten. In dieser Untersuchung können Angaben über die Zahl der Schüler und Schülerinnen in einer jeden Klasse, über die Leistung der Kinder und über spezielle Kommentare der Lehrkräfte zu den einzelnen Schulkindern genutzt werden, aber auch die Bewertung der Schule durch die Kinder.

In einem ersten Schritt wird durch nonparametrische statistische Verfahren ein jeweils denkbarer isolierter Einfluss der ausgewählten Variablen auf die Urteile der Lehrkräfte untersucht. Anschließend wird diese Analyse durch eine Regressionsanalyse der Daten ergänzt.

12.6.1 Potentieller Einfluss des Geschlechts der Lehrkräfte auf die Urteile in der Skala ESDM

In einem ersten Analyseschritt werden die Gesamtwerte aus den ESDM Urteilen der Lehrkräfte für zwei Schulformen, für Grundschule und Gymnasium, geschlechtsspezifisch auf Unterschiede geprüft.

Tabelle 12.33: Geschlechtsspezifisch unterschiedene ESDM Urteile der Lehrkräfte aus Grundschulen über die bei den Grundschulkindern (N= 923) zu beobachtenden Symptome (Mann-Whitney Test)

Ranks				Test Statistics	
Geschlecht	N	Mean Rank	Sum of Ranks	Mann-Whitney U	84206,000
Männer	312	497,61	155254,00	Wilcoxon W	271172,000
Frauen	611	443,82	271172,00	Z	-2,912
				Asymp. Sig. (2-tailed)	,004

Tabelle 12.34: Gesamtbeurteilung der Gymnasiasten (N= 902) im ESDM getrennt nach dem Geschlecht der Lehrkräfte (mittels Mann-Whitney Test)

Ranks				Test Statistics	
Geschlecht	N	Mean Rank	Sum of Ranks	Mann-Whitney U	80844,500
Männer	300	419,98	125994,50	Wilcoxon W	125994,500
Frauen	602	467,21	281258,50	Z	-2,571
				Asymp. Sig. (2-tailed)	,010

Die in den Tabellen 33 und 34 dargestellten Ergebnisse berechtigen zu den folgenden relevanten Arbeitsaussagen:

12.50 AS: Die von Lehrkräften an Grundschulen über ESDM Gesamtpunktwerte vollzogene Beurteilung des Auftretens depressionsspezifischer Symptome bei Grundschulkindern weist geschlechtsspezifische Unterschiede auf. Männer und Frauen urteilen signifikant unterschiedlich. Die männlichen Lehrkräfte berichten von einer signifikant höheren Belastung der Grundschulkindern als ihre weiblichen Kolleginnen.

12.51 AS: Die durch die Lehrkräfte über die Skala ESDM vollzogenen Gesamtbewertungen der Gymnasiasten ist ebenfalls geschlechtsspezifisch signifikant unterschiedlich, allerdings

gegenüber den Ergebnissen aus den Grundschulen nun in umgekehrter Richtung: die Frauen schätzen über die Gesamtpunktwerte aus der Skala ESDM die „depressive Belastung“ ihrer Schüler und Schülerinnen in Gymnasien höher ein als ihre männlichen Kollegen.

12.52 AS: Dieser gegensinnig zwischen Grundschulen und Gymnasien wechselnde geschlechtsspezifische Urteilsunterschied gegenüber der Auftretenshäufigkeit von depressiven Symptomen bei Schulkindern kann auch unter Benutzung des Begriffes der Toleranz gegenüber dem Auftreten der fraglichen Symptome ausgedrückt werden. In dieser Perspektive haben Lehrerinnen an Grundschulen eine höhere Toleranz gegenüber dem Auftreten depressionsspezifischer Symptome bei den Kindern als Lehrer. Entsprechend gilt, dass Lehrer an Gymnasien in der griechischen Stichprobe höhere Toleranz gegenüber dem Auftreten depressiver Symptome zeigen als ihre Kolleginnen.

12.53 AS: Zweifelsfrei gilt, dass die Urteile über die Skala ESDM geschlechtsspezifische Abhängigkeiten aufweisen, deren Akzente von Grundschule zu Gymnasium gegensinnig wechseln.

Dieser Sachverhalt ergibt sich aus der statistischen Überprüfung. Hinweise auf Bedingungen, nach denen sich dieser geschlechtsspezifische Aspekt ergibt, sind mit seinem Nachweis nicht bekannt geworden. Das veranlasste zur Suche nach weiteren Informationen. Möglicherweise könnten diese Unterschiede auf itemspezifische Variationen in den Urteilen zurückgehen, über die sich dann der beobachtete Wechsel aufbaute. Dem nachgehend war nach geschlechtsspezifischen Unterschieden in den Urteilen zu den verschiedenen Items der Skala ESDM zu fahnden.

Gepprüft wird, ob die zuvor beobachteten geschlechtsspezifisch auftretenden Unterschiede alle vom ESDM beachteten Symptome betreffen, oder ob sie auf spezielle Symptome begrenzt bleiben. Die itembezogene Übersicht trennt erneut die Kinder aus Grundschulen von denen aus den Gymnasien und ebenso die Lehrerinnen und Lehrer der beiden Schultypen. Die nachstehende Tabelle (Tabelle 12.35) zeigt die auf diese Verteilungen bezogenen Ergebnisse aus der Durchführung des Mann-Whitneys Tests.

Wie aus den in Tabelle 12.35 dargestellten Ergebnissen zu ersehen ist, ergeben sich für die Teilstichprobe der Grundschulkinde in 6 (von 16) Items des ESDM und für die Stichprobe der Gymnasiasten in 8 ESDM-Items signifikante Unterschiede beim Vergleich der Urteile von Lehrerinnen und Lehrern.

In 5 von 6 Aussagen, die in der Stichprobe der Grundschulkinde sehr signifikante oder tendenziell signifikante geschlechtsspezifische Abweichungen aufweisen, stufen die männliche Lehrkräfte die diesbezüglichen Schwierigkeiten der Kinder - vor allem „Traurigkeit“, „Widerspruch und Streit“, „Unentschlossenheit“, „schwankende Stimmungslage“ und tendenziell auch „Neigung zu Selbstvorwürfen“ - höher ein als ihre Kolleginnen. Einzige Ausnahme stellen die geschlechtsspezifischen Antworten der Lehrer auf das Item 8 („hat Spaß in der Schule“, umgepolte Frage) dar. Hier tendieren die weiblichen Lehrkräfte dazu, für ihre Schulkinder einen höheren Belastungsgrad anzusetzen als die Lehrer.

Die signifikant auftretenden Unterschiede bei den Lehrkräften der Grundschule tendieren also mehrheitlich in die Richtung des zuvor festgestellten geschlechtsspezifischen Unterschiedes für alle 16 ESDM Items. Zugleich zeigt sich aber, dass die beobachtete höhere Toleranz der Lehrerinnen gegenüber auftretenden Symptomen sich speziell auf 5 Items bezieht.

Das sind die Items 1 („fühlt sich traurig“), 10 („widerspricht und streitet“), 11 („neigt zu Selbstvorwürfen“), 12 („kann sich nicht entscheiden“) und 13 („schwankende Stimmungslage“).

Der Akzent der geschlechtsspezifischen Unterschiede bei einzelnen Items rechtfertigt die Annahme, dass die Unterschiede auf geschlechtsspezifisch verschiedenartige Sensibilitäten in der Unterrichtspraxis zurückgehen könnten. Grundschullehrer scheinen in dieser griechischen

Stichprobe gegenüber anderen Varianten des Schülerverhaltens sensibel zu reagieren als Lehrerinnen und dies trifft dann für andere Items auch auf die Gymnasien zu.

In der Stichprobe der Kinder an Gymnasien treten bei 8 ESDM-Aussagen geschlechtsspezifisch signifikante Abweichungen auf. Bemerkenswert ist wiederum, dass nicht alle Items von den Unterschieden betroffen sind. Die weiblichen Lehrkräfte bewerten Verhaltensäußerungen wie Traurigkeit, Konzentrationsprobleme, Selbstwertprobleme, Neigung zu Selbstvorwürfen und schwankende Stimmungslage hochsignifikant höher und andere Probleme wie Müdigkeit und Unentschlossenheit zumindest tendenziell höher als ihre männlichen Kollegen. Das wiederum gestattet die Wiederholung der zuvor getroffenen Feststellung. Die Vermutung ist gerechtfertigt, dass die Lehrerinnen an Gymnasien gegenüber speziellen Auffälligkeiten bei den Schülern und Schülerinnen sensibler reagieren als die Lehrer.

Tabelle 12.35: Beurteilung der Grundschul Kinder und der Gymnasiasten in 16 ESDM-Fragen durch Lehrerinnen und Lehrer (Mann-Whitney Test)

ESDM-Item	G	Grundschule				Gymnasium			
		N	Mean Rank	Z	p	N	Mean Rank	Z	p
1 „fühlt sich traurig“	M	312	515,22	-3,890	,000	324	463,26	-3,330	,001
	F	633	452,19			678	519,77		
2 „fällt schwer, aufmerksam zu sein“	M	312	476,58	-,315	,753	324	467,64	-2,758	,006
	F	633	471,23			678	517,68		
3 „ist in der Pause allein“	M	312	470,26	-,348	,728	322	478,72	-,378	,705
	F	633	474,35			628	473,85		
4 „weint“	M	312	483,52	-1,150	,250	313	435,10	-5,492	,000
	F	633	467,82			647	502,46		
5 „versteh den Unterricht“ ¹	M	312	492,13	-1,831	,067	324	482,20	-1,372	,170
	F	633	463,57			672	506,36		
6 „wird in der Klasse akzeptiert“ ¹	M	312	467,77	-,636	,525	324	501,36	-1,431	,152
	F	633	475,58			649	479,83		
7 „sieht müde aus“	M	312	487,21	-1,363	,173	318	472,65	-2,182	,029
	F	633	465,99			677	509,91		
8 „hat Spaß in der Schule“ ¹	M	312	455,41	-2,132	,033	316	461,17	-,702	,483
	F	633	481,67			619	471,49		
9 „fühlt sich weniger wert“	M	312	467,40	-,534	,594	323	441,11	-2,972	,003
	F	611	459,24			622	489,56		
10 „widerspricht und streitet“	M	312	508,20	-3,375	,001	324	502,60	-,200	,841
	F	633	455,65			676	499,50		
11 „neigt zu Selbstvorwürfen“	M	312	488,35	-1,983	,047	309	438,45	-3,085	,002
	F	633	465,43			619	477,50		
12 „kann sich nicht entscheiden“	M	312	517,58	-4,051	,000	314	459,49	-2,172	,030
	F	633	451,03			654	496,51		
13 „schwankende Stimmungslage“	M	312	510,89	-3,545	,000	324	455,23	-2,677	,007
	F	633	454,32			645	499,95		
14 „aktive Mitarbeit in der Schule“ ¹	M	312	482,24	-,863	,388	324	484,51	-1,192	,233
	F	633	468,45			673	505,97		
15 „zufrieden stellende Leistung“ ¹	M	312	487,82	-1,355	,175	324	493,99	-,449	,653
	F	633	465,69			674	502,15		
16 „spielt mit den anderen“ ¹	M	312	479,42	-,956	,339	323	467,61	-,096	,923
	F	633	469,84			613	468,97		

Zeichenerklärung: M = Lehrer; F = Lehrerinnen, N = Zahl der beurteilten Schulkinder, Mean-Rank = durchschnittlicher gruppenspezifischer Rang im Mann-Whitney Test; Z = z-Wert aus dem Mann-Whitney Test; p = die sich aus dem Z-Wert ergebende Irrtumswahrscheinlichkeit

Für den Vergleich zwischen Grundschule und Gymnasium wäre dann allgemein zu bemerken, dass die Lehrkräfte an den Grundschulen bezogen auf die Menge anfallender Beobachtungen nicht die gleiche Chance haben, sich gegenüber dem sensibel zu beweisen, was die Lehrerinnen an den Gymnasien deutlich erkennbar bewegt. Die Möglichkeit für Wahrnehmungen, denen gegenüber sie sensibel sind, nimmt altersbedingt an den Gymnasien im Vergleich zu davor liegenden Schuljahren zu. So darf die Hypothese verteidigt werden, Lehrer und Lehrerinnen seien gegenüber unterschiedlichen Symptomen an den Schulkindern geschlechtsspezifisch sensibel. Da sich die Frequenz der vom ESDM beachteten Symptomen von der Grundschule zum Gymnasium insbesondere für einzelne, emotionsnahe Bekundungen seitens der Schulkinder erhöht, haben Lehrerinnen und Lehrer im Wechsel von Grundschule zum Gymnasium unterschiedliche Chancen auf das zu treffen, was für sie jeweils besonders beachtenswert ist.

Was das dann jeweils ist, fasst die nachstehende Tabelle zusammen.

Tabelle 12.36 gibt, basierend auf die in Tabelle 12.35 dargestellten Daten, die zuvor mitgeteilten Informationen zusammengefasst wieder.

Tabelle 12.36: Überblick auf die signifikante Abweichungen bei der nach Lehrkräften geschlechtsspezifisch unterschiedenen Beurteilung der Grundschulkindern und der Gymnasiasten in den einzelnen ESDM-Fragen

ESDM-„Symptom“	Grundschule		Gymnasium	
	Signifikanz	Höhere Einschätzung	Signifikanz	Höhere Einschätzung
1 „fühlt sich traurig“	,000 (s.s.)	Männer	,001 (s.s.)	Frauen
2 „fällt schwer, aufmerksam zu sein“			,006 (s.s.)	Frauen
4 „weint“			,000 (s.s.)	Frauen
7 „sieht müde aus“			,029 (t.s.)	Frauen
8 „hat Spaß in der Schule“ ¹	,033 (t.s.)	Frauen		
9 „fühlt sich weniger Wert“			,003 (s.s.)	Frauen
10 „widerspricht und streitet“	,001 (s.s.)	Männer		
11 „neigt zu Selbstvorwürfen“	,047 (t.s.)	Männer	,002 (s.s.)	Frauen
12 „kann sich nicht entscheiden“	,000 (s.s.)	Männer	,030 (t.s.)	Frauen
13 „schwankende Stimmungslage“	,000 (s.s.)	Männer	,007 (s.s.)	Frauen

Signifikanz = 2-tailed p; s.s. = sehr signifikant, t.s. = tendenziell signifikant

Worauf die beobachteten geschlechtsspezifischen Unterschiede in den Urteilen der Lehrkräfte auf der Skala ESDM zurückzuführen sind, kann anhand der in dieser Untersuchung verfügbaren Informationen nicht zweifelsfrei entschieden werden.

Die beobachteten Unterschiede würden aber mit den nachstehenden Vermutungen in Übereinstimmung stehen.

12.54 AS: Die über die Skala ESDM abgegebenen Urteile von Lehrkräften unterliegen geschlechtsspezifischen Einwirkungen.

12.55 AS: Diese Einwirkungen gehen auf geschlechtsspezifisch unterschiedliche Sensibilitäten gegenüber dem Verhalten von Schülerinnen und Schülern zurück.

12.56 AS: Da die depressionsrelevanten Symptome in ihrer Ausprägung und Häufigkeit im Wechsel zwischen Grundschule und Gymnasium eine Steigerung erfahren, unterscheiden sich die in der Grundschule als auffällig darstellenden Symptome in ihrer Art von denen, die als Auffälligkeiten in Gymnasien zu erkennen sind.

12.57 AS: Lehrer erweisen sich als aufmerksamer gegenüber den in Grundschulen beobachtbaren Symptomausprägungen und beurteilen die als bemerkenswert. Im Gegensatz dazu

bewertet die Gruppe der Lehrer die an Gymnasien beobachtbaren Ausprägungen depressiver Symptome als weniger bedeutungsvoll. Ein Grund dafür könnten subjektive Theorien der Lehrer gegenüber pubertären Phänomenen sein, die trotz ihrer depressionsspezifischen Einschlägigkeit als erwartbar und temporär angesehen werden und von daher eher zu übergehen sind.

12.58 AS: Lehrerinnen zeigten gegenüber den mit der Pubertät anderen Ausprägungsgraden depressionsspezifischer Symptome eine ungleich höhere Sensitivität als Lehrer.

12.59 AS: Die mitgeteilten Annahmen haben rein hypothetischen Charakter und dienen dazu, eine bestehende Erklärungslücke versuchsweise zu füllen.

Von Interesse ist die Beobachtung, dass die umgekehrte Richtung der geschlechtsspezifischen Effekte bei den ESDM-Aussagen auftritt, die in beiden Schulstichproben zu signifikanten Unterschieden führen, aber eben bei zwei unterschiedlichen Gruppen von Lehrkräften. (Aussagen 1, 11, 12 und 13). Bei den Grundschulkindern wird Traurigkeit, Neigung zu Selbstvorwürfen, Unentschlossenheit und schwankende Stimmungslage von den männlichen Lehrkräften als höher eingeschätzt, bei den Gymnasiasten von den weiblichen Lehrkräften. Die zuvor mitgeteilten Hypothesen bieten zumindest probeweise eine Erklärung für das beobachtete Phänomen.

Andere Hypothesen wären denkbar und einsetzbar. Wären signifikant abweichende Urteile an sich schon Indizien für realitätsnahe Aussagen, würde das bedeuten, dass in der Grundschule die Männer und im Gymnasium die Frauen ihren Schülerinnen und Schülern mehr Aufmerksamkeit schenken. Wären dagegen diese Angaben eher Indizien für Übertreibungen seitens der Informationsquellen, wäre zu folgern, dass männliche Lehrkräfte in der Grundschule und weibliche Lehrkräfte in Gymnasien weniger tolerant gegenüber Verhaltensproblemen bei ihren Schülern und Schülerinnen sind. Aber im Widerspruch dazu ließe sich die Einschätzung verteidigen, der über ESDM gegebene Hinweis wäre Ausdruck von Fürsorglichkeit und Mitgefühl. Dann praktizierten das die Lehrer in den Grundschulen und die Lehrerinnen am Gymnasium signifikant höher als die jeweiligen Gegenpartien.

Hypothesen verlangen nach empirischen Prüfungen und auf deren Ergebnissen basierenden Entscheidungen.

12.6.2. Die Unterrichtsfächer der Lehrkräfte als potentielle Determinanten oder Moderatoren von Urteilen über die Skala ESDM

Ein neben dem Geschlecht weiteres Merkmal, über das sich die Gesamtstichprobe der beteiligten Lehrkräfte in Teilgruppen unterscheiden lässt, ist das der Unterrichtsfächer, die von den Lehrerinnen und Lehrern vertreten werden. Das ist dann auch der Fachgegenstand, für deren Vermittlung die Lehrkräfte in ihren Schulklassen speziell zuständig sind. Ob diese Merkmale als möglicherweise relevant für die in Schulklassen wirksame Wahrnehmung der Lehrkräfte und dann auch für das Erkennen depressionsspezifischer Symptome an den Kindern angesehen werden sollten, lässt sich gestützt auf unterschiedliche und auch widersprüchliche Argumente erörtern. Die Wahrnehmung psychischer Auffälligkeiten an Schulkindern dürfte einerseits eher als durch das Unterrichtsfach einer Lehrkraft von deren persönlicher Sensibilität und ihrem pädagogischen Engagement beeinflusst werden. Andererseits bestehen zwischen den in der Schule gegebenen Unterrichtsfächern durchaus Unterschiede, die den Kontakt zwischen Schülern und Lehrern qualitativ und quantitativ beeinflussen. Ein relevanter Aspekt neben anderen wäre die Tatsache, dass einige Fächer, wie beispielsweise philologische Fächer oder Mathematik, an griechischen Gymnasien etwa viermal pro Woche unterrichtet werden, während anderer Unterricht, so zum Beispiel der in Soziologie oder Gesellschaftskunde, oder auch der in Religion nur jeweils einmal pro Woche stattfindet. In diesen speziellen Fällen ist der mögliche Kontakt

zwischen Schulkindern und Lehrkräften zumindest im Hinblick auf die verfügbare Zeit im Vergleich zu anderen Unterrichtsfächern deutlich eingeschränkt. Ein zweiter Aspekt wäre der freie Raum, den die unterschiedlichen Fächer für Diskussionen über die eigene Person oder die persönliche Weltansicht zur freien Verfügung haben. So fördern von ihrer eigenen Natur her beispielsweise Fächer wie Philologie oder Kunst solche Diskussionen im Vergleich zu Fächern wie Physik oder Chemie in ungleich höherem Maße. Zu berücksichtigen bleibt auch der von Lehrkräften als „objektiv bestehend“ erlebte Zeitdruck aus den mit den Lehrplänen übermittelten Forderungen nach Vermittlung definierter Inhalte in definierten schulischen Zeiträumen.

Nach solchen Überlegungen wird im Folgenden geprüft, ob die Urteile der Lehrkräfte auf der Skala ESDM mit den von den Lehrern und Lehrerinnen vertretenen Unterrichtsfächern variieren.

Als erstes wird gefragt, ob die über das Verfahren ESDM gegebenen Bewertungen der an dieser Untersuchung beteiligten Lehrkräfte in Abhängigkeit von ihrem jeweiligen Unterrichtsfach stehen. Die statistische Überprüfung dieser Frage erfolgt mit einem dazu geeigneten nonparametrischen Verfahren, dem Kruskal-Wallis Test. Diese Nachprüfung lässt sich nur für die Stichprobe der Lehrkräfte an den Gymnasien durchführen, da nur die jeweils für spezielle Unterrichtsfächer ausgebildet sind. Die Lehrkräfte in den griechischen Grundschulen dürfen alle Fächer unterrichten, sie sind nicht für spezielle Fächer ausgebildet. Die ESDM Urteile der Grundschullehrer sind folglich bei der nachfolgenden Analyse nicht mit einbezogen worden. Die aus der Durchführung des Kruskal-Wallis Tests entstehenden Ergebnisse sind in Tabelle 12.37 dargestellt.

Tabelle 12.37: Nach Studienfächer der Lehrkräfte unterschiedene ESDM Urteile für n = 902 Jugendliche an griechischen Gymnasien (Kruskal Wallis Test)(Gymnasiasten)

Studienfach	N	Mean Rank	Test Statistics	
Philologie	560	481,39	Chi-Square	34,188
Mathematik	74	408,26	Df	7
Soziologie	44	379,81	Asymp. Sig.	,000
Chemie	40	413,86		
Religion	21	467,38		
Fremde Sprache	88	356,50		
Geographie	29	550,84		
Kunst	46	370,39		

Wie aus Tabelle 12.37 zu ersehen ist, treten bei den Bewertungen der Gymnasiasten über ESDM Urteile durch ihre Lehrkräfte statistisch signifikante Unterschiede für Variationen mit dem Studienfach der Lehrkräfte auf. Die Variationen sollten nicht als zufällig auftretend gewertet werden. Die Signifikanz gilt für die Gesamttabelle. Numerisch auffallende Unterschiede zwischen den Studienfächern der Lehrkräfte lassen sich vor allem wegen der niedrigen Fallzahlen in einigen Studienfächern nicht statistisch sichern.

Danach ist die folgende Aussage aufzuführen.

12.60 AS: Die über die Skala ESDM abgegebenen Urteile der Lehrkräfte an Gymnasien werden in nicht näher bekannt gewordener Ausprägung auch vom Studienfach der Lehrerinnen und Lehrer beeinflusst.

Die im vorangegangenen Abschnitt mitgeteilte Aussage regt dazu an, noch nicht genutzte Verfahrensweisen einzusetzen, um doch Aufschlüsse darüber zu erlangen, in welcher Weise über die Studienfächer der Lehrkräfte der nachgewiesene Einfluss auf die über die Skala ESDM mitgeteilten Urteile ausgeübt wird. Bekannt wurde, dass die höchste Bewertung von Philologen gegeben wurde und die niedrigsten von Kunstlehrern. Für diese numerischen Unterschiede waren aber keine Signifikanzen zu sichern. Weitere Informationen wurden unter anderem auch über die Durchführung einer univariaten Varianzanalyse (ANOVA) zu erlangen versucht. Der Einsatz dieser Rechnung missachtet die Verfahrensregel über den Nachweis des Vorliegens einer Normalverteilung, was dann nachfolgend, als empfohlener Ausgleich für die Verletzung der genannten Voraussetzungen, zur Anwendung des Wahrscheinlichkeitsniveaus von $\alpha = 0,01$ zwang. Das Ergebnis verweist auf statistisch signifikante Mittelwertsunterschiede im ESDM zwischen den unterschiedenen Fächern der Gymnasiallehrer [$F(7, 894) = 4,593$; $p = 0,000$; Eta squared = 3,5%]. Trotz Nutzung unterschiedlicher Post-Hoc-Tests (Scheffe, Bonferroni, Tukey HSD) zur Überprüfung von signifikanten Unterschieden zwischen den 8 Fachgruppen ergaben sich aber wiederum keine signifikanten Ergebnisse.

Im Weiteren wird nun versucht, die nach Studienfächern der Lehrkräfte unterschiedenen Korrelationen zwischen ESDM und CDI in die Nachforschungen einzubeziehen. In Tabelle 12.38 sind die aus dieser Überprüfung bekannt gewordenen Spearman'schen Korrelationen zwischen den Gesamtpunktwerten in der Skala ESDM und CDI, Selbsteinschätzungen der Jugendlichen, aufgeführt. Verglichen wird das Maß der Übereinstimmung zwischen den Urteilen der Lehrkräfte zum Auftreten depressiver Symptome bei den Schulkindern und der Gesamtpunktwert aus der Selbsteinschätzung der Schulkinder. Der statistische Vergleich folgt der Frage, welche Kinder durch depressive Symptome mehr belastet sind als welche anderen.

Tabelle 12.38: Spearman's Korrelationen zwischen ESDM und CDI für die nach Studienfächer getrennten Gruppen von Lehrkräften an den Gymnasien der Untersuchung 2034GR

Spearman's Korrelation	Fachgebiet							
	1	2	3	4	5	6	7	8
Rho	,279 ***	,278 *	,321 *	,085	,430	,371 ***	,019	-,169
N	557	75	42	40	21	88	29	46

1 = Philologie, 2 = Mathematik, 3 = Soziologie, 4 = Chemie, 5 = Religion, 6 = Fremde Sprache, 7 = Geographie, 8 = Kunst

Die rho-Werte variieren numerisch zwischen -0,169 und 0,43, aber ein Teil der Korrelationen setzt sich durch das Ausbleiben der Signifikanzen nicht von einer Grundgesamtheit mit der Korrelation 0 ab. Die Lehrkräfte mit den Studienfächern Chemie, Religion, Geographie und Kunst bilden gemeinsam diese Gruppe. Von denen unterscheiden sich in ihren Korrelationen die Lehrkräfte mit den Studienfächern Philologie, Mathematik, Soziologie und Fremde Sprachen. Nach diesem Sachverhalt wäre die Feststellung eines Unterschiedes zwischen den zwei genannten Gruppen von Lehrkräften formal gerechtfertigt. Werden dann aber die Zahlen der von diesen beiden Gruppen von Lehrkräften beurteilten Schulkinder verglichen, findet sich ein Verhältnis von $n=762$ zu $n=136$. Die bemerkten Unterschiede sind demnach ungeschützt gegenüber der

Möglichkeit, dass der mit dem Studienfach wechselnde Einfluss auf die ESDM Urteile vor allem ein Effekt der von einer Gleichverteilung signifikant abweichenden unterschiedlichen Schülerzahlen in den Stichproben sei. Je kleiner die Zahl der beurteilten Schulkinder ist, desto kleiner ist auch die Zahl der urteilenden Lehrkräfte. Da zuvor nachgewiesen wurde, dass die unterschiedlichen Urteilsleistungen der Lehrkräfte in das ESDM Urteil eingehen, besteht für kleine Stichproben von Lehrkräften die Möglichkeit, dass sie aus einseitig besetzten Extremgruppen von Lehrkräften bestehen könnten. Die Zahl der Schulkinder verbirgt die viel kleinere Zahl der für sie zuständigen Lehrkräfte. Unterstellte man eine Klassengröße von etwa 20 Kindern oder Jugendlichen, stünde die Zahl 40 zwar für diese Anzahl abgegebener Urteile, aber nur für 2 Lehrkräfte. Das spricht dann bei dem gegebenen Ermittlungsstand dafür, die zuvor getroffene Aussage über mögliche Einflüsse des Studienfaches der Lehrkräfte auf das ESDM Urteil nicht zu erweitern und stattdessen für nachgehende empirische Untersuchungen in anderen Zusammenhängen.

12.6.3 Geschlecht der Kinder

Im Folgenden wird auch der Möglichkeit einer auf Seiten der Schulkinder bestehenden geschlechtsspezifischen Urteilsvergabe durch die Lehrkräfte über das Verfahren ESDM nachgegangen. Die Tabellen 12.39 und 12.40 zeigen die mit den daraufhin angesetzten Analysen erarbeiteten Ergebnisse. Verwendet wurde der Mann-Whitney Test und dies für die Stichprobe der Grundschul Kinder und die der Gymnasiasten.

Tabelle 12.39: Nach Geschlecht der Grundschul Kinder unterschiedene Gesamtpunktwerte (N= 923) im ESDM (Mann-Whitney Test)

Ranks				Test Statistics	
Geschlecht	N	Mean Rank	Sum of Ranks	Mann-Whitney U	95430,000
Männer	491	483,64	237468,00	Wilcoxon W	188958,000
Frauen	432	437,40	188958,00	Z	-2,641
				Asymp. Sig. (2-tailed)	,008

Tabelle 12.40: Nach Geschlecht der Gymnasial Kinder unterschiedene Gesamtpunktwerte (N= 902) im ESDM (Mann-Whitney Test)

Ranks				Test Statistics	
Geschlecht	N	Mean Rank	Sum of Ranks	Mann-Whitney U	88438,000
Männer	491	478,92	227009,00	Wilcoxon W	180244,000
Frauen	432	421,13	180244,00	Z	-3,334
				Asymp. Sig. (2-tailed)	,001

Die in den Tabellen 12.39 und 12.40 mitgeteilten Ergebnisse führen zu den nachstehenden Aussagen:

12.61 AS: Die durch die Lehrkräfte vollzogene Beurteilung des Auftretens von depressiven Symptomen bei den von ihnen unterrichteten Kindern variieren signifikant mit dem Geschlecht der Kinder. Sowohl für die Gruppe der Grundschul Kinder als auch für die der Gymnasiasten

berichten die Lehrkräfte über eine signifikant höhere Anzahl depressive Belastung bei den Jungen als bei den Mädchen.

12.62 AS: Die geschlechtsspezifisch unterschiedliche Beurteilung der an Jungen und Mädchen durch die Lehrkräfte bemerkten depressiven Symptome im Sinne der Skala ESDM kann zumindest auf zwei Ursachen zurückgehen. Zum einen könnten diese Symptome, wie von den Lehrkräften angegeben, in dieser Stichprobe bei Jungen in höherer Zahl als bei den Mädchen auftreten. Zum anderen wäre möglich, dass die Lehrkräfte nicht valide urteilen. Hypothetisch nahe liegend wäre die Annahme, diese Urteile folgten dem aus der Literatur bekannten Phänomen der nach Urteilen von Lehrkräften bestehenden schulischen Überlegenheit der Mädchen.

12.63 AS: Vermerkt werden soll zunächst, dass die hier zuvor geführten Analysen zum Entstehen von ESDM Urteilen der beteiligten Lehrkräfte den determinierenden Einfluss zweier spezieller Bedingungen vermuten lassen. Zum einen liegen Ergebnisse vor, die auf die Abhängigkeit der Urteile der Lehrkräfte von schulisch gepflegten und an schulischen Kriterien orientierten Beurteilungen hinweisen. Zum anderen erscheinen die ESDM Urteile abhängig von der im Hinblick auf das Erkennen psychischer Besonderheiten an Kindern sehr unterschiedlichen Urteilskompetenz der Lehrkräfte. Diese beiden Ergebnisse lassen die ESDM Urteile der Lehrkräfte eher durch schulische Gewohnheiten des Urteilens geprägt erscheinen, als durch empathisches Erfassen psychischer Zustände von Schulkindern. Die Zuschreibungen höherer Zahlen für beobachtetes Auftreten von depressionsspezifischen Symptomen an die Jungen unter den Schulkindern und nicht an die Mädchen, steht im offensichtlichen Widerspruch zur gesamten internationalen Literatur über die geschlechtsspezifische Verteilung depressiver Störungen. Die seitens der Lehrkräfte vorgenommene Zuschreibung steht aber in guter Übereinstimmung mit der in der internationalen pädagogischen Literatur verbreiteten Auffassung, dass Gruppen von Jungen den vergleichbaren Gruppen von Mädchen sowohl in den schulischen Leistungen als auch im schulischen Verhalten häufig unterlegen sind und sie lediglich als Problemträger gravierend überragen. Die in dieser Untersuchung zu beobachtende Zuschreibung höherer depressionsspezifischer Werte an die Jungen ist in diesem Sinne ein drittes Indiz für die Abhängigkeit der ESDM Urteile der Lehrkräfte von schulinternen Kriterien und Gewohnheiten. Hier wird der Verdacht geäußert, dass diese Zuschreibung ein Hofeffekt der allgemein schlechteren schulischen Position der Jungen gegenüber den Mädchen ist.

Der hier zuvor dargestellte Analyseschritt bringt aussagekräftige Ergebnisse, aber die sind in ihrer Beweiskraft wegen der Verwendung der Gesamtpunktwerte durch die dazu vollzogene Summenbildung belastet. Denkbar wäre immerhin, die beteiligten Lehrkräfte hätten bei vielen Items der Skala ESDM die Mädchen als stärker belastet dargestellt, dann aber bei anderen Items die Jungen überproportional berücksichtigt. Unter speziellen Umständen könnten Gesamtsummen die geschlechtsspezifischen Unterschiede verdecken und auch verzerren. Ob solche Bedenken im Falle der untersuchten Stichprobe berechtigt sind, kann ein weiterer Analyseschritt klären. Der müsste den geschlechtsspezifischen Vergleich auf die 16 Items der Skala ESDM ausdehnen. Tabelle 12.41 zeigt, getrennt nach Grundschulkindern und Gymnasiasten, die von den Lehrkräften über das Verfahren ESDM gegebenen Urteile zum beobachteten Auftreten depressiver Symptome bei Schüler und Schülerinnen.

In der Betrachtung der nachstehenden Tabelle 12.41 ist zu beachten, dass die mit dem Sternchen >*< gekennzeichneten Items gegensinnig gepolt sind. Der höhere Testwert steht für eine Verneinung der Aussage, der niedrigere Testwert bestätigt die Aussage.

Für Item 5 wurde verbalisiert „versteht den Unterricht.“ Die Gruppe der Jungen in der Grundschule hat für dieses Item einen signifikant höheren Wert. Demnach gilt die Aussage: „Als Gruppe betrachtet, verstehen die Jungen den Unterricht signifikant schlechter als die Mädchen.“

Für Item 6 * gilt die Feststellung: „Als Gruppe betrachtet, werden die Jungen der Grundschule in der Klasse weniger akzeptiert als die Mädchen.“

Für Item 8 * gilt die Feststellung: „Als Gruppe betrachtet, haben die Jungen in der Grundschule weniger Spaß als die Mädchen.“

Für Item 14* ist die nachstehende Aussage zutreffend: „Als Gruppe betrachtet, zeigen die Jungen in der Grundschule weniger Mitarbeit als die Mädchen.“

Für Item 15 * ist festzustellen: „Als Gruppe betrachtet, bringen die Jungen weniger zufriedenstellende Leistungen als die Mädchen.“

Für Item 16 * wäre, falls Signifikanz vorläge, die nachstehende Feststellung zu treffen: „Als Gruppe betrachtet, spielen die XX weniger mit anderen als die YY.“

Bei der Teilstichprobe der Grundschul Kinder treten geschlechtsspezifisch signifikante Unterschiede bei 6 von 16 ESDM-Items auf. Die Lehrer berichten bei den Jungen ihrer Klassen über signifikant mehr Aufmerksamkeitsprobleme, weniger Spaß in der Schule, mehr Streit und über häufiger auftretende schulische Minderleistungen. Weiterhin sind sie in den Klasse weniger akzeptiert.

Die Items mit sehr signifikanten Unterschieden sind die nachstehend aufgeführten: 2, 6, 8, 10 und 15. Die Einbeziehung von als tendenziell signifikant zu betrachtenden Items wurde unterlassen. Die Stichprobengröße legt das nahe.

Bei der Stichprobe der Gymnasiasten weisen 8 Items des ESDM Abweichungen in Abhängigkeit vom Geschlecht der Kinder oder Jugendlichen auf. Aufmerksamkeitsprobleme, Verständnisprobleme im Unterricht, niedriges Selbstwertgefühl, Streitsucht, Defizite in der schulischen Mitarbeit und Leistungsdefizite belasten aus der Perspektive der Lehrkräfte in ihren Klassen die Jungen stärker als die Mädchen. Bei der Stichprobe der Gymnasiasten gehen in 2 ESDM-Aussagen die auftretenden signifikanten Unterschiede zu Lasten der Mädchen. Die Lehrkräfte beurteilen die Mädchen im Vergleich zu ihren Mitschülern als trauriger. Sie berichten auch, dass die Schülerinnen häufiger als die Schüler weinen.

Tabelle 12.41: Für 16 ESDM-Items gegebene Urteile der Lehrkräfte zum Auftreten depressiver Symptome bei Jungen und Mädchen an Grundschulen und Gymnasien (Mann-Whitney Test)

ESDM-Item	G	N	Grundschule			Gymnasium			
			Mean Rank	Z	p	N	Mean Rank	Z	p
1 „fühlt sich traurig“	J	499	463,18	-1,363	,173	518	470,53	-4,036	,000
	M	446	483,98			484	534,65		
2 „fällt schwer, aufmerksam zu sein“	J	499	495,78	-3,020	,003	519	561,31	-7,304	,000
	M	446	447,52			483	437,24		
3 „ist in der Pause allein“	J	499	475,97	-,567	,571	495	480,34	-,829	,407
	M	446	469,68			455	470,23		
4 „weint“	J	499	469,57	-,566	,572	502	449,42	-5,658	,000
	M	446	476,84			458	514,56		
5 * „versteh den Unterricht“ ¹	J	499	486,30	-1,918	,055	515	530,29	-3,987	,000
	M	446	458,12			481	464,46		
6 * „wird in der Klasse akzeptiert“ ¹	J	499	484,77	-2,154	,031	501	498,55	-1,679	,093
	M	446	459,84			472	474,74		
7 „sieht müde aus“	J	499	475,91	-,421	,674	517	512,10	-1,842	,065
	M	446	469,74			478	482,74		
8 * „hat Spaß in der Schule“ ¹	J	499	490,27	-3,154	,002	490	471,33	-,502	,616
	M	446	453,68			445	464,33		
9 „fühlt sich weniger Wert“	J	491	466,40	-,649	,516	490	491,59	-2,495	,013
	M	432	457,00			455	452,98		
10 „widerspricht und streitet“	J	499	518,54	-6,578	,000	517	541,01	-5,787	,000
	M	446	422,05			483	457,13		
11 „neigt zu Selbstvorwürfen“	J	499	484,08	-2,156	,031	483	459,45	-,883	,377
	M	446	460,60			445	469,99		
12 „kann sich nicht entscheiden“	J	499	465,05	-1,089	,276	497	487,67	-,408	,684
	M	446	481,90			471	481,16		
13 „schwankende Stimmungslage“	J	499	472,26	-,105	,917	502	470,27	-1,938	,053
	M	446	473,83			467	500,84		
14 * „aktive Mitarbeit in der Schule“ ¹	J	499	489,19	-2,278	,023	516	537,06	-4,674	,000
	M	446	454,89			481	458,17		
15 * „zufrieden stellende Leistung“ ¹	J	499	494,69	-2,987	,003	516	542,06	-5,177	,000
	M	446	448,74			482	453,94		
16 * „spielt mit den anderen“ ¹	J	499	478,43	-1,220	,223	487	458,89	-1,492	,136
	M	446	466,92			449	478,92		

Zeichenerklärung: G = Geschlecht der Kinder, J = Jungen, M = Mädchen * = kennzeichnet die umzupolenden Antworten, der höhere Punktwert kennzeichnet die Verneinung der Aussage, der niedrigere für deren Bestätigung.

Die mit diesem Analyseschritt erfassten Verhältnisse geben eine vereinfachte Form der eben betrachteten Tabelle als Übersicht wieder. Sie zeigt, bei welcher Frage der Skala ESDM die Teilgruppen der Jungen und bei welchen Items die Mädchen von den Lehrkräften als signifikant häufiger belastet genannt worden sind (siehe Tabelle 12.42).

12.63 AS: Die geschlechtsspezifische Tendenz der von den Lehrkräften über ESDN gegebenen Urteile über das Auftreten von depressiven Symptomen entsprechen auf der Itemebene den bereits aus der Analyse der Gesamtpunktwerte (ESDM) bekannten Befunde.

12.64 AS: Von insgesamt 32 Urteilen an Grundschulen und Gymnasien werden die Jungen in 13 Fällen (13 Items) als häufiger belastet beschrieben als die nur in zwei Fällen (2 Items) dominierenden Mädchen.

Tabelle 12.42: Unterschiede in den nach 16 ESDM-Items gegebenen Urteile der Lehrkräfte zum Auftreten depressiver Symptome bei Jungen und Mädchen an Grundschulen und Gymnasien (Mann-Whitney Test)

ESDM-Item	G	Grundschule	Gymnasium
		geschlechtsspezifischer Vergleich	geschlechtsspezifischer Vergleich
1 „fühlt sich traurig“	J M	Statistischer Gleichstand	Mädchen hochsignifikant stärker betroffen
2 „fällt schwer, aufmerksam zu sein“	J M	Jungen hochsignifikant stärker betroffen	Jungen hochsignifikant stärker betroffen
3 „ist in der Pause allein“	J M	Statistischer Gleichstand	Statistischer Gleichstand
4 „weint“	J M	Statistischer Gleichstand	Mädchen hochsignifikant stärker betroffen
5* „versteh den Unterricht“ ¹	J M	Statistischer Gleichstand	Jungen hochsignifikant stärker betroffen
6* „wird in der Klasse akzeptiert“ ¹	J M	Jungen hochsignifikant stärker betroffen	Statistischer Gleichstand
7 „sieht müde aus“	J M	Statistischer Gleichstand	Statistischer Gleichstand
8 * „hat Spaß in der Schule“ ¹	J M	Jungen hochsignifikant stärker betroffen	Statistischer Gleichstand
9 „fühlt sich weniger Wert“	J M	Statistischer Gleichstand	Jungen signifikant stärker betroffen
10 „widerspricht und streitet“	J M	Jungen hochsignifikant stärker betroffen	Jungen hochsignifikant stärker betroffen
11 „neigt zu Selbstvorwürfen“	J M	Jungen stärker belastet	Statistischer Gleichstand
12 „kann sich nicht entscheiden“	J M	Statistischer Gleichstand	Statistischer Gleichstand
13 „schwankende Stimmungslage	J M	Statistischer Gleichstand	Statistischer Gleichstand
14 * „aktive Mitarbeit in der Schule“ ¹	J M	Jungen signifikant stärker betroffen	Jungen hochsignifikant stärker betroffen
15 * „zufrieden stellende Leistung“ ¹	J M	Jungen hochsignifikant stärker betroffen	Jungen hochsignifikant stärker betroffen
16 * „spielt mit den anderen“ ¹	J M	Statistischer Gleichstand	Statistischer Gleichstand

Zeichenerklärung: G = Geschlecht der Kinder, J = Jungen, M = Mädchen * = hohe Testwerte stehen für die Negation der Aussagen (z.B. „keine zufrieden stellende Leistungen“)

Eine Variante der Problematik geschlechtsspezifisch unterschiedlicher Urteile von Lehrkräften wird über die Frage zugänglich, ob die Urteile von Lehrern einerseits und Lehrerinnen

andererseits mit dem Geschlecht der beurteilten Kinder signifikant variieren. So könnten beispielsweise Lehrer die Jungen und Mädchen ihrer Schulklassen in identischer oder aber auch in statistisch nicht übereinstimmender Weise beurteilen. Indem über die Nachprüfung der als aussagekräftig angesehenen Daten sich das eine oder das andere erweisen sollte, wäre aber noch nicht entschieden, was sich damit außer der Tatsache selbst beweist.

Tabelle 12.43: ESDM Urteile der Lehrkräfte in doppelter geschlechtsspezifischer Perspektive: Lehrer, Jungen und Mädchen, Lehrerinnen, Jungen und Mädchen (Gesamtpunktwerte, Mann-Whitney Test)

Geschlecht der Lehrkräfte	Geschlecht der Kinder	N	Grundschule			Gymnasium			
			Mean Rank	Z	p	N	Mean Rank	Z	p
Lehrer	J	167	168,20	-2,466	,014	157	164,29	-2,894	,004
	M	145	143,03			143	135,36		
Lehrerinnen	J	324	316,30	-1,542	,123	317	315,48	-2,085	,037
	M	287	294,37			285	285,95		

Zeichenerklärung: N = Anzahl beurteilter Schulkinder; Mean Rank = durchschnittlicher Rang als Kennwert im Mann-Whitney Test; Z = Z-Wert; p = Irrtumswahrscheinlichkeit

Die in der Tabelle aufgeführten Ergebnisse aus dem durchgeführten Mann-Whitney Test führen zu den nachstehenden Aussagen.

12.65 AS: Die in Grundschulen arbeitenden männlichen Lehrkräfte schätzen die depressive Belastung der Jungen signifikant höher ein als die der Mädchen.

12.66 AS: Bei der durch Lehrerinnen abgegebenen Urteilung über die Belastung der Schüler und Schülerinnen treten keine signifikanten Unterschiede auf.

12.67 AS: Die in Gymnasien arbeitenden männlichen Lehrkräfte berichten von einer höheren depressionsspezifischen Belastung der Jungen ihrer Klassen.

12.68 AS: Die Urteile der weiblichen Lehrkräfte an Gymnasien variieren nicht mit dem Geschlecht der Kinder oder Jugendlichen.

12.69 AS: Insgesamt ergibt sich, dass das Geschlecht der Kinder stärker die Urteile der männlichen Lehrkräfte beeinflusst als die der weiblichen.

12.6.4 Alter oder schulische Klassenstufe der Kinder

Die bisher durchgeführten Analysen haben für die ESDM Urteile der Lehrkräfte deren überzufällig ausgeprägte Variation mit dem Geschlecht der Lehrkräfte nachgewiesen. Tabelle 12.44 zeigt nun Ergebnisse aus der Überprüfung der Frage, ob geschlechtsspezifisch unterschiedene Lehrerurteile mit dem Alter der beurteilten Kinder variieren. Die Ergebnisse gingen aus der Durchführung des Kruskal-Wallis Tests an den Daten (Gesamtpunktwerte nach ESDM) für die Kinder aus den Grundschulen hervor. Als Äquivalent für das Alter der Kinder wurden die unterschiedlichen Klassenstufen der Grundschule benutzt. Das erwies sich als notwendig, weil die von einer Lehrkraft beurteilten Kinder einerseits alle einer identischen Schulklasse angehören, andererseits in ihrer Klasse aber leichte Altersvariationen zeigen. Schließlich wären dann auch noch in anderen Klassen einzelne Kinder gleichen Alters zu finden. Entschieden wurde, die Teilgruppen nach den schulischen Klassenstufen zu bilden, womit die Klassenstufe auch als Bezugsgröße genannt werden muss. Die Klassenstufen sind insoweit auch

Alterstufen, als die jeweils höhere Klassenstufe gegenüber der darunter liegenden mehrheitlich ältere Kinder umfasst.

Tabelle 12.44: Gesamtbeurteilung der Grundschul Kinder (N= 923) im ESDM nach Klassenstufen

Grundschule	<u>Männliche Lehrkräfte</u>				<u>Weibliche Lehrkräfte</u>			
	N	Mean Rank	Chi-Square	Asymp. Sig	N	Mean Rank	Chi-Square	Asymp. Sig
3e Klasse	18	92,14	21,747	,000	182	312,59	67,249	,000
4e Klasse	29	117,22			185	375,82		
5e Klasse	140	176,28			125	211,46		
6e Klasse	125	152,72			119	286,68		

12.70 AS: Die von den Lehrkräften angegebene Häufigkeit des Auftretens depressiver Symptome nach ESDM Items variiert mit der Klassenstufe und insoweit auch mit dem Alter von Grundschulkindern sowohl bei weiblichen als auch bei männlichen Lehrkräften.

12.71 AS: Die Bewertungen der männlichen Lehrkräfte in Grundschulen werden mit zunehmendem Alter (Klassenstufe) ihrer Schüler signifikant höher(= höhere Belastung).

12.72 AS: Die Bewertungen der weiblichen Lehrkräfte für die psychische Befindlichkeit der Grundschul Kinder zeigt keine lineare Variation.

12.73 AS: Die Bedeutung der Variationen von ESDM Urteilen mit den Klassenstufen von Grundschulkindern ist insoweit mehrdeutig, als auch Selbstzuweisungen von Symptomen und diagnostische Befunde allgemein Altersvariationen aufweisen.

Der Wiedergabe der Ergebnisse für die Grundschul Kinder der Stichprobe folgt die für die Kinder an Gymnasien.

Tabelle 12.45: Gesamtbeurteilung der Gymnasiakinder (N= 923) im ESDM durch Lehrer und durch Lehrerinnen auf schulische Klassenstufen bezogen (N=902)

Gymnasium	<u>Männliche Lehrkräfte</u>				<u>Weibliche Lehrkräfte</u>			
	N	Mean Rank	Chi-Square	Asymp. Sig	N	Mean Rank	Chi-Square	Asymp. Sig
1e Klasse	112	157,39	1,466	,481	215	261,17	19,611	,000
2e Klasse	92	142,68			177	311,94		
3e Klasse	96	149,95			210	333,99		

12.74 AS: Die über ESDM nach Gesamtpunktwerten abgegebenen Urteile von Lehrerinnen an Gymnasien variieren signifikant mit der Klassenstufe.

12.75 AS: Die Gesamtpunktwerte aus den ESDM Urteile der weiblichen Lehrkräften an Gymnasien steigen mit der Klassenstufe der Schulkinder signifikant an.

12.76 AS: Die Gesamtpunktwerte aus den ESDM Urteile der männlichen Lehrkräfte an Gymnasien variieren statistisch nicht mit der Klassenstufe der Schulkinder.

12.77 AS: Auch für die mit der Klassenstufe signifikant variierenden Gesamtpunktwerte aus den ESDM Urteile der weiblichen Lehrkräften an Gymnasien ist nicht zu entscheiden, worauf die Unterschiede zurückgehen.

12.6.5 ESDM Urteile der Lehrkräfte und die von ihnen abgegebenen Leistungsbeurteilungen für die Kinder in ihrem Unterricht

Die Lehrkräfte der untersuchten griechischen Stichprobe waren nach der Durchführung der Skala ESDM gebeten worden, für jeden einzelnen Schüler und für jede Schülerin aus der von ihnen unterrichteten Schulklasse zwei zusätzliche Fragen zu beantworten.

Die erste Frage bezog sich auf die Leistung des jeweiligen Kindes in dem durch die befragte Lehrkraft dort unterrichteten Schulfach. Die Lehrkräfte sollten ankreuzen, ob im Vergleich zum Klassendurchschnitt die jeweiligen Leistung als „sehr viel niedriger“, „niedriger“, als „im Durchschnitt der Klasse liegend“ oder als „überdurchschnittlich“ einzustufen seien. Diese Frage schafft die Möglichkeit zu prüfen, ob das für das jeweilige Kind gegebene ESDM Urteil der Lehrkraft möglicherweise mit der Leistungseinstufung durch die Lehrkraft variiert. Sollte sich das erweisen, bliebe im Weiteren fraglich, ob der Zusammenhang möglicherweise auf die Beeinflussbarkeit des Urteils der Lehrkraft, oder aber auf reale Leistungsunterschiede zurückgeführt werden sollte. Dies dann allerdings auch nur, falls nicht noch weitere Abhängigkeiten in Betracht zu ziehen wären.

Im Folgenden wird geprüft, ob die das Auftreten depressiver Symptome an den Kindern einschätzenden ESDM Urteile der Lehrkräfte in Beziehung zu den Urteilen über die ebenfalls von der jeweils unterrichtenden Lehrkraft gegebenen Leistungsbeurteilungen stehen.

Tabelle 12.46: ESDM Urteile und Leistungseinschätzungen durch Lehrer und Lehrerinnen für Grundschul Kinder und Gymnasiasten (Kruskal-Wallis Tests)

Geschlecht Lehrkräfte	Leistung	<u>Grundschule</u>				<u>Gymnasium</u>			
		N	Mean Rank	Chi-Square	Asym. Sig.	N	Mean Rank	Chi-Square	Asym. Sig.
Männer	sehr niedrig	18	275,28	130,898	,000	39	258,97	184,154	,000
	niedrig	39	243,49			55	217,45		
	durchschnittlich	95	185,54			87	161,52		
	überdurchschnittlich	160	104,69			119	75,95		
Frauen	sehr niedrig	27	534,24	293,253	,000	71	441,13	277,987	,000
	niedrig	77	474,66			142	411,61		
	durchschnittlich	175	377,79			161	307,65		
	überdurchschnittlich	317	191,86			210	150,10		

Die in Tabelle 12.46 dargestellten Ergebnisse aus der Durchführung des Kruskal-Wallis Tests verweisen auf deutliche Beziehungen zwischen den einerseits von den Lehrkräften gegebenen ESDM Urteilen und den andererseits durch sie mitgeteilten Leistungsbeurteilungen für die von ihnen unterrichteten Kinder. Im Kruskal-Wallis Test wird geprüft, ob für die vier unterschiedenen Leistungsgruppen der Kinder signifikant unterschiedliche „Mean Ranks“ festgestellt werden. Je

höher solch Mean Rank liegt, umso mehr depressive Symptome haben die Lehrkräfte für die jeweils gemeinte Teilgruppe der Kinder beobachtet. Das Testergebnis zeigt eine durchgehende Parallelität der Abfolge der Leistungsgruppen mit den vom Test nach den gefundenen Mean Ranks unterschiedenen Teilgruppen unterschiedlicher ESDM Urteile. Mit steigender Leistungsstufe sinkt der über Mean Ranks angezeigte Stand der nach ESDM Urteilen auftretenden depressiven Symptome. Die nachstehenden Aussagen geben das wieder.

12.78 AS: Die über ESDM durch Lehrkräfte abgegebenen Urteile zum beobachteten Auftreten von depressiven Symptomen bei den von ihnen unterrichteten Schulkindern zeigen signifikante Unterschiede, deren Rangreihe reziprok zur Einschätzung der Leistungen dieser Kinder verläuft. Hohe depressive Belastung und niedriger Leistungsstand fallen zusammen.

12.79 AS: Die ESDM Urteile der männlichen und der weiblichen Lehrkräfte variieren sowohl in der Grundschule als auch am Gymnasium signifikant mit der abgegebenen Leistungsbeurteilung.

12.80 AS: Für alle vier an Grundschule und Gymnasium nach Leistungsbeurteilungen seitens der Lehrkräfte unterschiedenen Teilstichproben ergibt sich, dass die Hinweise der Lehrkräfte durch ESDM Urteile auf Belastungen ihrer Schüler und Schülerinnen durch depressive Symptome in einer linear verlaufenden negativen Beziehung zu den für diese Kinder gegebenen Leistungsbeurteilungen stehen.

12.81 AS: Das Ergebnis dieser Überprüfung wirft die Frage auf, ob die beiden von allen Lehrkräften abgegebenen Urteile über das Auftreten depressiver Symptome einerseits und die Leistungen der Kinder andererseits miteinander in Abhängigkeit stehen, oder ob der Leistungsstand der Kinder mit dem Auftreten depressiver Symptome variiert.

12.82 AS: Die mitgeteilten Ergebnisse stützen den über mehrere andere Befunde in dieser Untersuchung entstandenen Verdacht, dass die ESDM Urteile der Lehrkräfte mit der von ihnen im Berufsalltag entwickelten pädagogischen und vornehmlich an schulbezogenem Verhalten und schulischen Leistungen orientierten Sichtweisen und weniger an Beobachtungen depressiver Symptome orientiert sind.

12.6.6 Beziehungen zwischen ESDM Urteilen und anderen von den Lehrkräften gegebenen Informationen über die Kinder in ihrem Unterricht

Im Verlaufe der durchgeführten Untersuchung waren die beteiligten Lehrkräfte gebeten worden, weiterreichende Informationen über die von ihnen unterrichteten Schulkinder mitzuteilen. Dafür vorgegeben worden waren die nachstehenden Kategorien, die insgesamt spezielle Lebensumstände der Kinder betreffen. Gebeten worden war um Informationen über 1. möglicherweise bestehende gesundheitliche Probleme der Kinder, 2. über gesundheitliche Probleme bei den Eltern und weiterhin ob 3. Eltern getrennt lebten, 4. die Kinder aus Flüchtlingsfamilien kämen und auch ob 5. andere familiäre Belastungen bekannt seien, so etwa Drogenprobleme, familiäre Gewalt, oder psychische Probleme. Weiterhin war um Auskunft gebeten worden, ob 6. das jeweilige Kind spezielle Verhaltensprobleme entwickelt hätte, insbesondere dann ob 7. Zurückgezogenheit oder Überempfindlichkeit bekannt geworden seien, 8. Lernprobleme beständen, 9. aus der Sicht der Lehrkraft auf positive Eigenheiten zu verweisen sei, 10. eine Aufmerksamkeitsstörung oder auch Hyperaktivität vorliegen könnte, 11. das Kind von häufigem Schulwechsel betroffen war und ob 12. Todesfälle aus der Familie bekannt geworden seien (Tod eines Elternteils, eines Geschwisterkindes oder der Tod besonders nahe stehender Verwandter). Eine weitere Kategorie erfasste Hinweise auf die Arbeitstätigkeit von Schülern oder Schülerinnen. Angaben, die in diese Kategorie fallen, liegen, korrespondierend mit dem Alter der Schülerinnen und Schüler, nur für Gymnasien hervor.

Diese an die Lehrkräfte gerichteten Bitten um Hinweise auf besondere Lebensumstände folgen thematisch der Life-event Forschung. Ihnen war aber innerhalb der Untersuchung nur eine möglicherweise in Anspruch zu nehmende Stützfunktion zugeordnet worden. Das erschien bereits wegen der zu antizipierten Unterschiede in der Informiertheit der Lehrkräfte über die Schulkinder in ihrem Unterricht als zwangsläufig und wird so auch weiterhin so eingeschätzt. Wenn nun nachfolgend auf diese Informationen eingegangen wird, geschieht das mit den entsprechenden Vorbehalten. Alle etwa zu beachtenden Ergebnisse aus Analysen an diesen Daten hätten folglich nur Hinweischarakter. Sie könnten die Bildung entsprechender Forschungsfragen anregen.

Die nachstehende Tabelle führt die Ergebnisse an, die aus einer zur Suche nach potentiellen Hinweisen auf Zusammenhänge zwischen den von Lehrkräften berichteten Lebensumständen der Kinder und den auf sie bezogenen ESDM Urteile der Lehrer und Lehrerinnen durchgeführten statistischen Analyse hervorgegangen sind.

Tabelle 12.47: Nach durch Lehrkräfte angegebenen speziellen Lebensumständen unterschiedene Teilgruppen an Kindern und deren ESDM Daten (Kruskal Wallis Test)

R	Nach Hinweisen der Lehrkräfte auf Lebensumstände der Kinder gebildete Untergruppen	N	Mean Rank	Nach Hinweisen der Lehrkräfte auf Lebensumstände der Kinder gebildete Untergruppen	N	Mean Rank
	Kinder an Grundschulen			Kinder an Gymnasien		
1	„Lernprobleme“	17	755,9	„Schüler ist sehr introvertiert und überempfindlich“	9	757,8
2	„Gestörte familiäre Situation“ (Drogen, Gewalt, psychische Probleme)	23	737,6	„Lernprobleme“	14	717,0
3	„Aufmerksamkeitsstörung, Hyperaktivität“	6	719,3	„Getrennte Eltern“	13	695,0
4	„Gesundheitliches Problem des Kindes“	8	692,9	„Verhaltensprobleme“ (S. stört die anderen Kinder, lügt usw., ist ein Egoist, ist unreif)	12	688,1
5	„Getrennte Eltern“	26	655,3	„Gesundheitliches Problem des Kindes“	8	685,6
6	Schulwechsel (S. neu in der Schule)	5	629,1	„Aufmerksamkeitsstörung, Hyperaktivität“	6	671,8
7	Tod eines Elternteils/Geschwister	6	614,5	Schulwechsel (S. neu in der Schule)	5	667,7
8	„Schüler ist sehr introvertiert und überempfindlich“	15	577,7	„Gestörte familiäre Situation“ (Drogen, Gewalt, psychische Probleme)	16	664,7
9	„Gesundheitliches Problem der Eltern“	6	576,5	„Flüchtlinge“	79	562,2
10	„Verhaltensprobleme“ (S. stört die anderen Kinder, lügt usw., ist ein Egoist, ist unreif)	15	554,2	Tod eines Elternteils/Geschwister	13	518,1
11	„Flüchtlinge“	70	554,0	„Gesundheitliches Problem der Eltern“	3	480,3
12	Keine Informationen	707	419,4	Das Kind arbeitet	6	461,4
13	Positive Kommentare durch Lehrkräfte („ruhig“, „lächelnd“, „zielstrebig“)	19	374,9	Keine Informationen	664	422,6
14				Positives Kommentar („ruhig“, „lächelnd“, „zielstrebig“)	54	253,3
<u>Kennwerte für die Grundschulen</u> ESDM total: Chi-Square = 109,461; df = 12; Asymp. Sig. = 0,000; Kruskal Wallis Test.				<u>Kennwerte für die Gymnasien</u> ESDMtotal: Chi-Square = 128,345; df = 13; Asymp. Sig. = 0,000; Kruskal Wallis Test		

Die Kategorie "Kind arbeitet" ist von den Lehrkräften an den Grundschulen nicht verwendet worden. Das entspricht den Erwartungen. Arbeitstätigkeit sollte erst mit zunehmender Annäherung an die Altersgruppe der Erwachsenen auftreten. Das ist dann in der Gymnasialstufe gegeben.

Die durchgeführte statistische Analyse mit dem Kruskal Wallis Test führte zu dem Hinweis, dass die durch die Lehrkräfte abgegebenen ESDM Urteile mit den nach Lebensumständen unterschiedenen Gruppen an Schulkindern variieren. Das trifft sowohl für Grundschulen als auch für Gymnasien zu. Damit ist die Vermutung berechtigt, die von den Lehrkräften angegebenen Beobachtungen zum Auftreten von depressiven Symptomen variierten mit den ebenfalls von Lehrkräften angegebenen Informationen über die Lebensumstände der Kinder. Zum einen ist dieser Hinweis als Ergebnis zu vermerken, zum anderen bleiben dann erneut zwei Fragen aufzuwerfen. Die ESDM Urteile könnten zum einen mit den durch die Lehrkräfte selbst abgegebenen Hinweisen auf die Lebensumstände der Kinder variieren. Zum anderen ließe sich vermuten, die ESDM Urteile der Lehrkräfte variierten mit den Lebensumständen. Beide Mutmaßungen sind keineswegs identisch. Zu beachten sind drei unterschiedliche Phänomene, die alle miteinander in Beziehung stehen dürften, dies aber nicht müssen. Das eine Phänomen ist das in seiner jeweiligen Ausprägung bestehende ESDM Urteil (A) einer Lehrkraft, das andere ist die individuelle Antwort (B) auf die Frage nach speziellen Lebensumständen des Kindes, weiterhin ist das Bestehen (C) von Lebensumständen des gemeinten Kindes einzubeziehen und schließlich ist das qualitativ und quantitativ zu unterscheidende Auftreten depressiver Symptome als ein außerhalb der Lehrkraft bestehender Sachverhalt zu betrachten. Eine im gegebenen Zusammenhang relevante Struktur der denkbaren Beziehungen zwischen den Phänomenen A, B, C und D wird durch den jeweils zu treffenden Entscheid hergestellt, ob die jeweils betrachteten Phänomene innerhalb von entsprechenden Toleranzgrenzen in einseitiger oder wechselseitiger Beziehung stehen. Möglich wäre, dass die Lebensumstände C in einer einseitigen oder wechselseitigen Beziehung zu den Ausprägungen des Auftretens von depressiven Symptomen stünde. Die Ergebnisse der Life-event Forschung lassen das annehmen. Keinerlei Aufschlüsse sind innerhalb der hier geführten Untersuchung darüber verfügbar, ob das Urteil B der Lehrkraft über die Lebensumstände der Kinder valide sei. Die hochfrequent auftretende Kategorie „Keine Information über besondere Lebensumstände durch die Lehrkraft“ kann Unterschiedliches bedeuten. Möglicherweise wäre die Lehrkraft nicht an den Lebensumständen der Kinder interessiert, oder aber aus diesen oder jenen Gründen an der Informationsnahme gehindert. Die Lehrkraft könnte auch fehlerhafte Informationen weitergegeben haben. Wenn dann die ESDM Urteile A und die Hinweise auf Lebensumstände B beide valide sein sollten, müssten die urteilenden Lehrkräfte jeweils auch beide Phänomene C und D realistisch wahrgenommen und wiedergegeben haben. Die Zahl der möglichen a anderen Kombinationen ist relativ hoch. Erörtert werden müssen sie an dieser Stelle der Erörterungen aber nicht, weil die Datenlage keine Entscheidung über das Zustandekommen der festgestellten Beziehungen zwischen ESDM Urteilen der Lehrkräfte einerseits und den von ihnen gegebenen Informationen über Lebensumstände der Kinder in ihren Klassen andererseits zulassen.

So bleibt das gewonnene Ergebnis mit der nachstehenden Aussage zu vermerken.

12.83 AS: Die ESDM Urteile der Lehrkräfte variieren in einer gegenüber zufälligen Abweichungen signifikant unterschiedenen Weise mit den nach ihren Angaben zu speziellen Lebensumständen der Kinder hergestellten Teilgruppen.

12.84 AS: Die am Mean Rank (Kruskal Wallis Test) für die ESDM Urteile der Lehrkräfte an Grundschulen orientierte Rangreihe reicht vom Höchstwert für die nach Lehrerangaben mit „Lernproblemen“ belasteten Schülergruppe zu einer durch heraushebende positive

Einschätzungen durch die Lehrkräfte gekennzeichneten Schülergruppe mit dem niedrigsten Mean Rank der ESDM Urteile.

12.85 AS: Die am Mean Rank (Kruskal Wallis Test) für die ESDM Urteile der Lehrkräfte an Gymnasien orientierte Rangreihe reicht vom Höchstwert für die nach Lehrerangaben gebildete Gruppe der Kinder und Jugendlichen, die jeweils als „sehr introvertiert und überempfindlich“ beurteilt worden waren bis hin zu einer durch heraushebende positive Einschätzungen durch die Lehrkräfte gekennzeichneten Schülergruppe mit dem niedrigsten Mean Rank der ESDM Urteile. Die nachstehende Tabelle führt auf, welche Rangplatzdifferenzen für die einzelnen Kategorien an von Lehrkräften unterschiedenen Lebensumständen im Vergleich zwischen den ESDM Urteilen von Lehrkräften an Grundschulen und an Gymnasien auftreten.

Tabelle 12.48: Rangplatzdifferenzen zwischen ESDM Urteilen von Lehrkräften an Grundschulen und Gymnasien für die nach Informationen über Lebensumstände durch die Lehrkräfte unterschiedenen Teilgruppen an Kindern

Nach Hinweisen der Lehrkräfte auf Lebensumstände der Kinder gebildete Untergruppen	RGru	RGym	Differenz RGru - RGym
Kinder an Grundschulen			
„Lernprobleme“	1	2	-1
„Gestörte familiäre Situation“ (Drogen, Gewalt, psychische Probleme)	2	8	-6
„Aufmerksamkeitsstörung, Hyperaktivität“	3	6	-3
„Gesundheitliches Problem des Kindes“	4	5	-1
„Getrennte Eltern“	5	3	+2
Schulwechsel (S. neu in der Schule)	6	7	-1
Tod eines Elternteils/Geschwister	7	10	-3
„Schüler ist sehr introvertiert und überempfindlich“	8	1	+7
„Gesundheitliches Problem der Eltern“	9	11	-2
„Verhaltensprobleme“ (S. stört die anderen Kinder, lügt usw., ist ein Egoist, ist unreif)	10	4	+6
„Flüchtlinge“	11	9	+2
Keine Informationen	12	13	-1
Positive Kommentare durch Lehrkräfte („ruhig“, „lächelnd“, „zielstrebig“)	13	14	-1
Kind arbeitet		12	****

Zeichenerklärung: RGru = Rangplatz in der Rangreihe für die Urteile der Lehrkräfte aus Grundschulen; RGym = Rangplatz in der Rangreihe für die Urteile der Lehrkräfte aus Gymnasien; **** = keine Differenz berechnet, Kategorie wurde nur an Gymnasien genutzt

Die auftretenden Rangplatzdifferenzen entstammen zwar einer von Zufallswerten signifikant abweichenden Verteilung, sind aber dadurch noch nicht auf zufällige Schwankungen der Rangplätze geprüft. Rangplatzdifferenzen unter +/- 4 werden in der nachfolgenden Betrachtung übergangen. Danach ist dann auf Unterschiede aufmerksam zu machen, die zumindest wegen ihrer relativ hohen Rangplatzdifferenzen bemerkenswert werden. Die nachstehenden Aussagen fassen die so festgestellten Ergebnisse zusammen.

12.86 AS: Die Lehrkräfte an Grundschulen und Gymnasien geben über die Skala ESDM mit Teilgruppen von kategorial nach Lebensumständen gebildete Gruppen von Kindern variierende und insgesamt signifikant unterschiedliche Häufigkeiten für das Auftreten depressiver Symptome an.

12.87 AS: Mit hochsignifikant niedrigeren Frequenzen des Auftretens depressiver Symptome wird an beiden Schultypen die Teilgruppe der von den Lehrkräften wegen positiver persönlicher Eigenheiten von anderen Schülerinnen und Schülern abgegrenzten Teilgruppe beurteilt.

12.88 AS: Die Teilgruppe der Schülerinnen und Schüler mit Lernproblemen wird von den Lehrkräften beider Schultypen auf einem der beiden Spitzenplätze der Rangreihe platziert und damit auch durch Höchstwerten an beobachteten depressiven Symptomen gekennzeichnet.

12.89 AS: Auf den vorderen Rangplätzen der Reihen beider Schultypen sind Teilgruppen von Kindern zu finden, deren kategoriale Unterscheidung gegenüber anderen auf persönliche, schulisch nicht förderliche Eigenheiten oder auf Familienprobleme verweist.

12.90 AS: Gegenüber der für Grundschulen geltenden Rangreihe werden an Gymnasien 4 Schülergruppen auf niedrigere Rangplätze positioniert, was mit der Zuweisung von im Vergleich für an Gymnasien höhere Frequenzen an beobachteten depressiven Symptomen als an Grundschulen verbunden ist. Dazu gehört mit einer Rangplatzdifferenz von +7 und in der Rangliste der Gymnasien auf dem 1. Platz liegend die Teilgruppe der als introvertiert und überaus sensibel bezeichneten Gymnasiasten. Weiterhin folgt mit einer Platzdifferenz von +6 die Teilgruppe der mit Verhaltensproblemen belasteten Gymnasiasten. Die beiden genannten Teilgruppen werden also an Gymnasien als stärker von depressiven Symptomen betroffen gesehen als an Grundschulen.

12.91 AS: Stärkere Ausprägungen an auftretenden depressiven Symptomen beobachteten die Lehrkräfte an Grundschulen gegenüber den an Gymnasien für Kinder aus Familien mit Drogenkonsum, Gewalt und psychischen Problemen.

12.92 AS: Die zwischen den Rangreihen an Grundschulen und Gymnasien auftretenden Rangplatzdifferenzen treten in Übereinstimmung mit pädagogischen oder klinisch psychologischen Befunden auf. Sie korrespondieren weitgehend mit den über Altersunterschiede zwischen den Schultypen erwartbaren divergierenden schulischen Problembereichen.

12.6.7 ESDM Urteile der Lehrkräfte und Frequenz der Kinder in den Schulklassen

In die Suche nach potentiellen Determinanten der ESDM Urteile von Lehrkräften wird nachfolgend auch die Frequenz der Mädchen und Jungen in einer Schulklasse einbezogen. Die Öffnung dieser Perspektive folgt der Annahme, dass die Informationen über das, was die Schüler und Schülerinnen denken, von der Zahl der Kinder in der Klasse abhängig sein könnte, weil sich mit unterschiedlichen Frequenzen auch die Bedingungen für die Kommunikation zwischen Kindern und Lehrkräften verändern.

Die nun folgende Tabelle unterscheidet die Schulklassen an Grundschulen und an Gymnasien nach der Zahl der ihnen jeweils zugehörenden Kinder und stellt die statistischen Kennwerte dar, die sich über die statistische Prüfung mit dem Kruskal Wallis Test für die Unterschiede in den ESDM Urteilen ergeben haben.

Tabelle 12.49: ESDM Urteile und Frequenzen in den Schulklassen

Tabellenteil der ESDM Urteile über Grundschulkinder (N= 923)								
Schüler pro Klasse	<u>Männliche Lehrkräfte</u>				<u>Weibliche Lehrkräfte</u>			
	N	Mean Rank	Chi-Square	Asymp. Sig	N	Mean Rank	Chi-Square	Asymp. Sig
5 – 14	107	126,36	20,754	,000	191	314,59	44,688	,000
15 – 22	155	166,64			265	345,55		
23 – 32	50	189,58			155	227,80		

Tabellenteil für die ESDM Urteile über Gymnasiasten (N= 902)								
Schüler pro Klasse	<u>Männliche Lehrkräfte</u>				<u>Weibliche Lehrkräfte</u>			
	N	Mean Rank	Chi-Square	Asymp. Sig	N	Mean Rank	Chi-Square	Asymp. Sig
5 – 14	23	150,61	11,409	,003	191	47	269,91	,000
15 – 22	166	164,84			265	224	245,64	
23 – 32	111	129,03			155	331	343,79	

Die Ergebnisse der durchgeführten Unterschiedsprüfungen verweisen in allen vier Teilbereichen auf signifikante und mit der Schülerzahl variierende Unterschiede in den auf Teilgruppen bezogenen ESDM Urteilen.

Diese Variante der Einflussfaktoren auf ESDM Urteile ist bislang kaum mitgeteilt worden. Sie erweist sich im Übrigen auch als uneinheitlich. Worauf sie zurückgehen bleibt fraglich. Eine alle vier Teilgruppen übergreifende Aussage scheint nicht möglich. Deswegen kann die nachstehende Aussage auch nur den in der Tabelle wiedergegebenen Sachverhalt vermerken.

12.93 AS: Werden die zur untersuchten Stichprobe gehörenden Mädchen und Jungen in Teilgruppen aufgeteilt, die sowohl nach Schülerzahl in den jeweiligen Schulklassen als auch nach dem Geschlecht der Lehrkraft unterscheidet, ergeben sich signifikante Unterschiede in den ESDM Urteilen für diese Gruppen.

In einem nächsten Analyseschritt wird der These gefolgt, dass die ESDM Urteile auch mit den untersuchten einzelnen Schulklassen signifikant variieren könnten.

Tabelle 12.50: ESDM Urteile und Schulklasse

Klasse Nr.	N	Mean Rank	Klasse Nr.	N	Mean Rank	Klasse Nr.	N	Mean Rank	Klasse Nr.	N	Mean Rank
1	16	1409,66	27	8	449,13	53	15	907,10	79	22	1369,41
2	18	718,53	28	10	714,50	54	26	1117,17	80	29	1183,74
3	18	1046,22	29	7	1395,43	55	20	812,88	81	24	514,27
4	21	921,19	30	19	1289,34	56	26	841,04	82	10	672,75
5	13	696,15	31	2	1089,00	57	24	1006,56	83	14	383,11
6	14	697,54	32	19	530,34	58	26	492,96	84	14	1035,21
7	18	565,19	33	10	1111,10	59	21	1133,52	85	18	558,08
8	16	495,16	34	14	1165,29	60	10	1278,35	86	23	1178,72
9	19	1086,08	35	10	540,60	61	15	1536,30	87	28	1096,45
10	27	440,83	36	14	694,96	62	24	1023,67	88	31	1157,08
11	28	570,52	37	25	1470,68	63	21	581,45	89	19	1118,39
12	15	891,30	38	23	1416,89	64	30	1297,53	90	14	326,86
13	13	1241,31	39	21	1020,55	65	9	919,56	91	17	699,85
14	14	560,86	40	21	544,26	66	11	879,86	92	10	593,50
15	20	1213,72	41	24	874,85	67	13	811,54	93	8	702,44
16	29	1060,90	42	17	788,50	68	7	872,71	94	9	987,83
17	24	1276,81	43	15	1130,17	69	18	835,03	95	5	404,50
18	17	1037,85	44	20	660,45	70	21	1290,52	96	13	262,81
19	18	1090,64	45	19	891,95	71	13	649,85	97	8	1209,00
20	24	985,42	46	21	759,69	72	18	1085,53	98	13	1265,77
21	24	1231,23	47	15	1151,37	73	14	731,61	99	13	694,54
22	12	782,21	48	19	813,08	74	17	846,38	100	16	1067,63
23	25	797,94	49	22	1020,20	75	13	1046,65	101	17	751,44
24	18	774,25	50	11	862,59	76	17	572,68	102	21	574,69
25	18	1090,11	51	24	315,94	77	17	951,09	103	22	1178,02
26	9	1339,89	52	24	754,33	78	19	941,63	104	20	763,20

ESDM total: Chi-Square = 552,512; df = 103; Asymp. Sig. = 0,000; Kruskal Wallis Test.

Die Überprüfung der Verteilung der ESDM Urteile auf die untersuchten Schulklassen unter Anwendung des Kruskal Wallis Tests führt zum Nachweis einer signifikanten Varianz der Mean Ranks (Kruskal Wallis Test) und damit auch zum Nachweis einer von Zufallswerten abweichenden Variation in der Gesamtverteilung.

12.94 AS: Die für die untersuchten Schulklassen durch die dort unterrichtenden Lehrkräfte abgegebenen ESDM Urteile als quantifizierte Angaben über das beobachtete Auftreten depressiver Symptome variieren insgesamt mit von Zufallsunterschieden signifikant abweichenden Werten.

12.7 Aspekte der Bewährung des ESDM in der griechischen Untersuchung 2034GR

Für ein Messinstrument wie die Skala ESDM kann in unterschiedlichen Zusammenhängen geprüft werden, ob sie sich bewährt. Dann ist jeweils zu prüfen, ob sich gehegte spezielle Erwartungen erfüllen. Bewährung schließt auch nicht aus, dass manche Hoffnungen unerfüllt blieben, wenn sich ein Objekt wenigstens in der einen oder anderen Hinsicht bewährte. Bewährung kann auch für einen Einzelfall belegt sein, obschon sie in vielen anderen Zusammenhängen ausblieb. Bewährung ist in diesem Sinne eine spezielle Relation zwischen Erwartungen und deren empirischen Korrelaten. Domènech und Mitarbeiter (1985) bezeichneten die Skala ESDM als nützliches und hilfreiches Messinstrument bei der Bewältigung der Aufgabe, große Stichproben spanischer Schulkinder im Hinblick auf das Vorliegen depressiver Störungen zu untersuchen. Auch für die griechische Untersuchung zur Stichprobe 2034GR war die Skala ESDM unter speziellen Erwartungen für den Einsatz als Messinstrument ausgewählt worden. Zum einen war erwartet worden, das Verfahren werde einen Beitrag zur Identifizierung von Schulkindern mit depressiven Störungen leisten. Zum anderen sollte das Verfahren einen Perspektivenwechsel in der Wahrnehmung depressiver Schulkinder von der Selbsteinschätzung zur Fremdeinschätzung ermöglichen und dies nicht zu einer beliebigen, sondern eben zu einer schulisch präsenten. Dies war gewünscht, weil abzusehen ist, dass die präventiven und kurativen Bemühungen um die psychische Gesundheit von Schulkindern auch in der mittelfristig abgeschätzten Zukunft trotz belasteter öffentlicher Haushalte allgemein arbeitsintensiv bleiben werden. Wenn Lehrer und Lehrerinnen innerhalb solcher Bemühungen eine kompetent auszufüllende Funktion übernehmen könnten, nützte das den Bemühungen um die psychische und körperliche Gesundheit von Schulkindern.

Der Einsatz der Skala ESDM in der durchgeführten Untersuchung führte zu einer insgesamt 104 Schulklassen und 2034 Schulkinder unterschiedlicher Altersklassen an Grundschulen und Gymnasien betreffenden und über die ESDM Urteile von 104 Lehrkräften hergestellten Datenbasis.

Die Analyse dieser Datenbasis ist in den vorausgegangenen Abschnitten dieses Kapitels eingehend beschrieben worden. Über diese Darstellungen sind auch die zahlreichen Ergebnisse dieser Analysearbeit mitgeteilt und erläutert worden.

Diese Ergebnisse legitimieren in ihrer Gesamtheit die nachstehend aufgeführten Feststellungen zu Fragen nach der Bewährung des ESDM in dieser Untersuchung:

1. Die von den Lehrkräften abgegebenen ESDM Urteile erreichen in ihrer Gesamtheit nur eine schwache Übereinstimmung mit den über CDI oder DTK zum Auftreten depressiver Symptome erfassten Selbsteinschätzungen der Schulkinder.
2. Eine etwa 20% der Lehrergruppe umfassende Teilgruppe erreichte in signifikanter Absetzung von der Mehrheit der Lehrerinnen und Lehrer eine bis etwa zu einer mit dem Korrelationskoeffizienten $r = 0,70$ zu quantifizierende Annäherung an die Selbsteinschätzungen der Schulkinder.
3. Das Auftreten zweier signifikant unterschiedenen Gruppen von Lehrkräften, von denen die etwa 80% umfassende Mehrheit sich nicht von einem Koeffizienten ± 0 abzusetzen vermochte, weist nach, dass persönliche Unterschiede in der Informiertheit der Lehrkräfte über die Selbstwahrnehmung der von ihnen unterrichteten Schulkinder eine der Determinanten für die Annäherung der Urteile von Lehrkräften an die der Schulkinder ist.
4. Weitere Ergebnisse aus den durchgeführten Analysen legen nahe, die über ihre Berufspraxis in der Wahrnehmung von Schulkindern ausgebildeten Sichtweisen als eine andere Determinante der über die Skala ESDM abgegebenen Lehrerurteile aufzufassen.

5. Da die Selbsteinschätzungen von Schulkindern in Validitätsstudien mittlere bis hohe Korrelationen mit Diagnosen von depressiven Störungen aufweisen, dürften die über die Skala ESDM abgegebenen Urteile der Mehrheit der Lehrkräfte in ihrer Validität als unter der von Selbsteinschätzungen der Schulkinder liegend vermutet werden.
6. Die Ergebnisse aus der Anwendung der Skala ESDM in der griechischen Stichprobe 2034 sprechen gegen die Annahme einer allgemein beachtenswerten Validität der ESDM Daten gegenüber dem Auftreten depressiver Störungen bei Schulkindern.
7. In Verbindung mit der gleichzeitigen Anwendung eines Verfahrens für die Selbsteinschätzung depressiver Symptome durch Schulkinder eignet sich die Skala ESDM als diagnostisches Instrument für die Identifizierung der Informiertheit von Lehrkräften über die depressionsspezifische Selbstwahrnehmung der von ihnen unterrichteten Schulkinder.

Kapitel 13

Der aus den Teilergebnissen der durchgeführten Untersuchung zu generierende Gesamtbefund

Die bis hier dargestellte und erörterte Untersuchung an einer Stichprobe von 2034 griechischen Schulkindern und ihren Klassenlehrerinnen und -lehrern setzte an einem durch die Wahrnehmung von Wissensdefiziten über emotionale Störungen bei griechischen Schulkindern begründeten Forschungsinteresse an.

Das stieß zunächst fast zwangsläufig auf kognitive Widersprüche zwischen dem aktuellen und dem auf griechischen Boden schon historisch vorfindbaren Wissensstand über Depressionen. Die aus dem Rückblick auf die antike hellenische Epoche Griechenlands aufbrechenden Wahrnehmung der ideengeschichtlichen Entwicklung der Psychologie in diesem Teil Europas veranlasste zu einer der eigentlichen empirischen Studie vorangehenden Befassung mit den Bedingungen des Erwerbs und des Erhalts gesellschaftlichen gespeicherten Wissen. Beides erweist sich als einem der zahllosen und miteinander verflochtenen Entwicklungsstränge im Fortgang des durch die Jahrtausende führenden Lebens von Völkerschaften, Nationen und Religionen überantwortet. Die für das Leben der Menschen folgenschweren Brüche und Verwerfungen verschonen auch psychologische Traditionen nicht. Die Divergenz zwischen dem für die griechische Antike nachgewiesenen psychologischen Wissensstand und dem als defizitär erscheinenden aktuellen Wissen über psychosoziale Aspekte des Lebens in Griechenland und dabei dann insbesondere die eher geringen Kenntnisse über die Verbreitung emotionaler Störungen unter griechischen Schulkindern wurde als ein kognitiver Widerspruch empfunden, der nach begründeten Erklärungen suchen ließ. Die Nachforschungen zum aktuellen Stand der Verbreitung emotionaler Störungen unter griechischen Schulkindern wurden auf diese Weise historisch relativiert. Sie erwies sich selbst trotz ihrer intendierten Aktualität als in einen kaum noch überschaubaren historischen und gesellschaftlichen Kontext eingebunden.

Darin verdeutlichte sich zum einen die potentielle Instabilität der Existenz und des Fortbestandes gesellschaftlich verfügbaren Wissens, was zum anderen die Notwendigkeit ständiger Bemühungen um die Sicherung der Wissensbestände innerhalb einer Gesellschaft und dort dann auch in der Psychologie für Kinder und Jugendliche, die, wenn sie entsprechend relevante Wissensbestände erarbeitet hat, der jeweiligen Gesellschaft nutzbares Wissen zur effizienten Gesundheitsvorsorge für die nachwachsende Generation bietet. Ob die es dann zu nutzen weiß, ist von dem Entwicklungsstand gesellschaftlich organisierter und gepflegter Sorge für das Wohlergehen der Mitmenschen abhängig. Im Übrigen ist selbst die Beantwortung der Frage, ob depressive Störungen schon seit historischen Vorzeiten aufgetreten sind, von Bedeutung für die aktuell vertreten Theorien über depressive Störungsbilder. Aufschlüsse darüber, ob depressive Phänomene und entsprechende Störungen in der bisherigen Entwicklung der Menschheit aufgetreten sind oder doch eher unbekannt gewesen seien, haben Informationswert für das dem Umgang mit Depressionen unterliegende Verständnis der Störungen und den mit ihnen entstehenden individuell existentiellen und den daraus erwachsenden gesellschaftlichen Problemen.

Mit den in den Anfangskapiteln dieser Arbeit geführten Erörterungen werden Argumente und Belege für die These gesammelt, dass die jeweils persönliche Betroffenheit durch depressive Phänomene in der Menschheitsgeschichte immer gegenwärtig gewesen ist und sich erwartungsgemäß mit zunehmender Geschichtlichkeit des Menschen in ihrem Auftreten zunehmend eindeutiger nachweisen lässt. Für die Annahme einer die depressiven Phänomene

nach sich ziehenden genetischen Veränderung innerhalb der Entwicklungsgeschichte des Menschen fehlen entsprechende Hinweise und selbst relevante Verdachtsmomente. Zumindest ist beides in der einschlägigen Literatur nicht erörtert worden. Die Hypothese eines vorgeschichtlichen Auftretens depressiver Symptome darf als wohlbegründet gelten. Für die geschichtlichen Zeiten liegen Nachweise für die psychische Repräsentanz der depressiven Störung vor, die in ihrer Beweiskraft den heute geltenden Belegen nicht nachstehen.

Die mit den Krankheiten der Menschen ihrer Umgebung befassten philosophischen und medizinischen Ärzte (Plutarch) der hellenistischen Antike berichten über das Krankheitsbild der Depression mit den zu ihrer Zeit benutzten Begrifflichkeiten. Sie stellen weiterhin um die Zeit der Geburt Christi das Vorliegen von Wissen um Depressionen als historisch verbürgt dar. Weiterhin sind in den ältesten Mythen und Epen, die hier als Spiegelungen des Weltwissens der Menschen vergangener Zeiten verstanden werden, an vielen Stellen spezifische Hinweise auf das Auftreten depressiver Episoden innerhalb der berichteten Geschehnisse. Exemplarisch nachgewiesen wurde in einem der Anfangskapitel der vorgelegten Ausführungen, dass Plutarch, als einer der Philosophen und philosophischen Ärzte zum Ende der hellenischen Antike und wohlvertraut mit römischem Leben, zum Auftreten der Depressionen eine ähnliche Perspektive öffnet. Er sieht sie als in seiner Zeit und in der ihm fassbaren hellenischen Geschichte als weithin verbreitet an. Als beispielhafter Beleg für die von Plutarch vertretene Sichtweise diene in den vorausgegangenen Kapiteln sein Text „Über Euthymie“, der nach wohlgemeinten Umformulierungen in der deutschen Übersetzung zu einem als Verständnishilfe eingesetzten Titel kam, dem „Von der Heiterkeit der Seele“. Dieser Titel steht programmatisch vor einem Text, der die Lebensfreude und Tatkraft zerstörende Wirkung der Melancholie darstellt und verdeutlicht, um die Notwendigkeit einer Befreiung für die von Melancholie betroffenen Menschen und den von Plutarch entdeckten Weg zu diesem Ziel um so deutlicher erkennbar werden zu lassen.

Eine exemplarische Analyse dieses Textes belegt, dass in dieser Zeit bereits sehr spezielles Wissen über die Bedingungen und die Auswirkungen des Auftretens depressiver Episoden bekannt war. Passagenweise ähneln die Ausführungen dem, was in Lehrbüchern der kognitiven Therapie von Depressionen um 2000 geschrieben wird. Beispielsweise hätten A. Beck (1976) oder Hautzinger und Mit. (1989) sich mit ihren Ansätzen zur kognitiven Therapie auch auf Plutarch berufen können. Hinzuweisen ist auf seine in jener Zeit schon Erfolg entsprechende kurative und präventive Interventionen gegenüber dem Auftreten depressiver Symptome und Syndrome. Mit der Kenntnis über die Verfügbarkeit dieses Wissens im klassischen griechischen Altertum erscheint die im Vergleich mit mitteleuropäischen Ländern noch um 1980 allgemein defizitäre wissenschaftliche Situation der griechischen Psychologie als eine nach Auflösung verlangende kognitive Dissonanz. In dieser Perspektive ist der auch noch um das Jahr 2000 defizitäre Stand der psychologischen Forschung zum Auftreten von Depressionen bei griechischen Kindern eine Variante der Folge mehrerer über 1500 Jahre Geschichte zurück zu verfolgenden Reihung kultureller, gesellschaftlicher, ökonomischer und sozialer Brüche. Zugleich erweist sich damit dann auch wie unabweisbar die nun um das Jahr 2000 im Wachstum begriffene griechische Psychologie ihrerseits und jede einzelne von ihr betriebene empirische Studie in den weiteren Verlauf der Geschichte einbezogen ist.

Der Ertrag der in den vorausgegangenen Erörterungen dieser Arbeit eingebetteten exemplarischen historischen Analyse zu dem für Plutarch nachweisbaren Wissen über Depressionen ist nicht nur der erarbeitete Beleg für den vorgefundenen hohen Wissensstand. Dieser Beleg wird zum Argument für den Erklärungswert der in den hier darzustellenden Befund der durchgeführten Untersuchung gehörenden nachfolgenden Feststellung. „Als Ergebnis der in dieser Arbeit erfolgten Befassung mit dem in hellenistischer Zeit verfügbaren Wissen über Depressionen und mit dem nachfolgend durch historische Ereignisse bewirkten allgemeinen und in die griechische Gesellschaft des 20. Jahrhunderts hinein reichenden Verlustes an gesellschaftlich nutzbarem

psychologischen Wissen wird die Abhängigkeit der wissenschaftlichen psychologischen Arbeit von den sie insgesamt hervorbringenden, behindernden oder verhindernden gesellschaftlichen Bedingungen gesehen und in der Zusammenfassung des Gesamtergebnisses dieser Untersuchung hervorgehoben. Aus dem Sachverhalt dieser Bedingtheit der wissenschaftlichen Psychologie wird die Folgerung gezogen, dass, soweit sie zur Sicherung und Verbesserung der Lebensbedingungen von Menschen beizutragen vermag, zwei allgemeine Zielsetzungen für die wissenschaftliche Psychologie zu unterscheiden sind, die Realisierung psychologischer Arbeit zum einen und die Positionierung der psychologischen Arbeit in der Gesellschaft zum anderen. Beiden Zielsetzungen ist notwendigerweise permanent zu folgen. Die Positionierung in der Gesellschaft steht in Wechselwirkung mit der sozialen Nutzbarkeit psychologischer Arbeit und psychologischen Wissens.“

Die Wahrnehmung dieser Zusammenhänge festigte das Arbeitsziel, empirisch begründetes Wissen über das Ausmaß der aktuellen Betroffenheit griechischer Schulkinder durch emotionale Störungen in Erfahrung zu bringen. Das aus diesem allgemeinen Interesse erwachsende Untersuchungsziel war, in einer hinreichend großen und in möglichst vielen Aspekten auch repräsentativen Stichprobe depressionsspezifische Informationen zu sammeln, die möglichst verlässliche Aussagen über die Betroffenheit der Schulkinder dieser Stichprobe vom Auftreten depressiver Symptome und depressiver Störungen erlauben würden und zugleich eine Abschätzung der Verbreitung dieser psychischen Belastung unter griechischen Schulkindern gestatteten. Auf einer noch sehr allgemeinen Betrachtungsebene könnte erwartet werden, dass die Ermittlung von quantifizierenden Daten für das Auftreten depressiver Störungen unter griechischen Schulkindern ein wesentliches das Untersuchungsziel sei. Unter Wahrnehmung der gegebenen Untersuchungsmöglichkeiten ergibt sich über entsprechende Analysen, dass unabwendbare methodische Restriktionen dazu nötigen, das Untersuchungsziel auf das Sammeln von symptom-spezifischen Informationen zu beschränken. Die notwendige Präzisierung des Untersuchungszieles führte in Kapitel 6 der vorangegangenen Erörterungen zu einem Perspektivenwechsel in der Bestimmung des Untersuchungszieles. War zunächst das Untersuchungsziel als determinierende Vorgabe für das Untersuchungsgeschehen zu verstehen, ergab sich im Verfolgen dieser Zielsetzung unter dem Zwang zur Präzisierung und Objektivierung des Untersuchungsansatzes die Notwendigkeit einer das Vorgehen relativierenden Umkehrung der Denkweise. Die mit der anzusetzenden Untersuchung realisierbaren Messungen bestimmen über die mit ihnen gewonnenen Daten, welche empirisch gesicherten Aussagen aus der Durchführung der Untersuchung dem der Arbeit zu Grunde liegenden Untersuchungsinteresse geboten werden können.

Im Kapitel 4 führte dies zu der nachstehend zitierten Setzung. „In einer verkürzten Schreibweise wäre festzustellen, dass mit der Wahl von vier Messverfahren das gesuchte Untersuchungsobjekt bestimmt worden sei.“ Die Wahl der Messinstrumente bestimmt zwangsläufig und unumkehrbar, welche Art von Daten aus der Durchführung der Untersuchung zu erwarten sind. Dementsprechend findet sich in Kapitel 4 auch die im Folgenden zitierte Feststellung. „Nachstehend wird versucht, die Beschreibung des Untersuchungsobjektes zu präzisieren. Es tritt, da es bei allen an der Untersuchung teilnehmenden Mädchen oder Jungen und allen teilnehmenden Lehrkräften gesucht wird, im Plural auf.“ Die angekündigte Beschreibung ist diese: „Untersuchungsobjekte sind die von den teilnehmenden Kindern, Jugendlichen und Lehrkräften auf die Darbietung von depressionsrelevanten Fragen markierten Varianten vorgegebener Antworten.“ Das Produkt der durchgeführten Untersuchung sind die in der Datenmatrix aus allen eingesetzten Verfahren gespeicherten Informationen.

Die in dieser Matrix aufgeführten Daten erlauben, für sich genommen, noch keine dem Forschungsinteresse dienende Aussage. Erst durch die Datenanalyse ergeben sich Aussagemöglichkeiten, die ihrerseits durch die verwendeten Analyseverfahren determiniert sind.

Die in den vorangegangenen Kapiteln ausführlich beschriebenen Analyseschritte führten zur Formulierung einer größeren Menge von Teilergebnissen. Im folgenden ist nun zu versuchen, aus diesen Teilergebnissen einen Gesamtbefund zu generieren, der Antworten auf die Frage nach dem Auftreten depressiver Belastungen unter griechischen Schulkindern gibt, die durch die empirisch gewonnenen Daten zu belegen sind.

13.1 Befunde in der Perspektive der Datenaufnahme als kommunikativer Prozess

Hauptanliegen aller in dieser Arbeit bisher dargestellten Erörterungen war die Auseinandersetzung mit speziellen Aspekten der psychischen Gesundheit griechischer Schulkinder. Die Bemühungen waren auf empirisch fassbare Anzeichen für das Auftreten von Symptomen aus dem Störungsbild der Depressionen bei Kindern und Jugendlichen gerichtet und befassten sich mit wissenschaftlichen Gegenständen in einer zugleich für die griechische Gesellschaft bedeutsamen Problematik. Sollte sich erwiesen haben, dass verlässliche Indizien für das verbreitete Auftreten von depressiven Störungen bei griechischen Schulkindern aufzufinden waren, wäre das eine Aufforderung an die Eltern, die Schulen, das Gesundheitswesen und die griechische Gesellschaft insgesamt, davon Kenntnis zu nehmen und sich mit der Bedeutung des nachgewiesenen Sachverhaltes zu befassen.

Die mit einer Stichprobe von 2034 griechischen Schulkindern realisierte empirische Arbeit führte über die zuvor eingehend beschriebene Datenaufnahme und die daraufhin vorzunehmenden Analysen zu insgesamt mehreren Hunderten von Aussagen. Der nun folgende abschließende Arbeitsschritt ist dem Ziel gewidmet, die aufgefundenen zahlreichen Einzelergebnisse zu den primären Forschungsfragen dieser Arbeit in Beziehung zu setzen und darzustellen, welche empirisch begründeten Antworten zu geben sind. Das sollte dann erlauben, die potentielle Nutzbarkeit des Untersuchungsbefundes und der mit ihm vermittelten Aufschlüsse zu erörtern.

Der nach dieser Zielsetzung durchzuführende Arbeitsschritt ist, abgesehen von seiner Position in diesen Erörterungen, formal wie alle anderen zu behandeln, was nach der Festlegung des Zieles eine Entscheidung darüber verlangt, wie das Arbeitsziel erreicht werden soll. Verlangt ist, die vorliegende Vielzahl an Aussagen in eine nützliche Ordnung zu bringen. Nützlich ist hier, was, wie angestrebt, die Beantwortung der primären Forschungsfragen dadurch erleichtert, dass der Zusammenhang zwischen den gewonnenen Ergebnissen und der durchgeführten Tätigkeit wiederhergestellt wird. Die Ergebnisse sind aus dieser Tätigkeit hervorgegangen, und die erklärt sie über die Bedingungen ihres Entstehens. Das ist für die Bewertung der Ergebnisse wesentlich, weil die durch die Untersuchungstätigkeit bedingt sind. Sie stehen in Wirkverbindung. Der Begriff wird statt des anscheinend näher liegenden Terminus einer kausalen Beziehung verwendet, weil kausale Verbindungen eine Kette von fortlaufend auftretenden Ursachen und deren Wirkungen annehmen lassen könnten. Solche Vorstellungen entsprechen nicht dem realisierten Ablauf der Untersuchung. An dessen Beginn stehen psychische Repräsentationen und Prozesse in einer Person, vielleicht auch in mehreren Personen einer mit der Leiterin der Untersuchung beratenden Gruppe. Die dabei ablaufenden innerpsychischen und kommunikativen Prozesse stellen noch keine wirkliche Verbindung mit den späterhin 2034 teilnehmenden Schulkindern her, sie sind intra- und interpersonale psychische Abläufe in einem sozialen Subsystem. Durch die Kontaktaufnahme mit den auf den sozialen Verbindungen zu den Schulkindern anzusprechenden Persönlichkeiten setzt die Untersuchungsleiterin, die auch die einzige in der Untersuchung tätige Person auf deren universitären Seite ist, Kommunikationsprozesse in Gang, die sich von Persönlichkeit zu Persönlichkeit in anderen

sozialen Subsystemen mit ihrem weiteren Verlauf den Schulkindern nähern, die Mitglieder der Stichprobe sein sollen und dies wahrscheinlich auch werden. Eine kausale Kette im engeren Sinne stellt sich dabei nicht ein, kausale Ereignisse im engeren Sinne wirken nur auf den Wegen zwischen den jeweils eigene Subsysteme darstellenden Persönlichkeiten. Von Persönlichkeit zu Persönlichkeit entscheiden die Beteiligten nach intra- und interpersonalen Bedingungen, ob überhaupt und wenn, was und wie das im eigenen Subsystem kommuniziert wird. Die Untersuchungsleiterin hat diesen kommunikativen Prozess zielgerichtet angestoßen. Wie er dann verläuft ist zwar nicht völlig, aber doch weitgehend unbekannt. Am Ende der zahlreichen kommunikativen Stränge werden mehr als Schulkindern von den kommunikativen Ereignissen erreicht. In dem Tag, in dem die Untersuchungsleiterin den Schulkindern einer Schulklasse gegenüber steht, kommt dann auch die Stunde, in dem der kommunikative Prozess abläuft, über den die Daten entstehen, die als Datenbasis für die Suche nach den Untersuchungsergebnissen dienen werden. Dieser Daten generierende kommunikative Prozess hat seine eigenen und im jeweiligen Untersuchungsgang weitgehend unbekannt Bedingungen. Die Untersuchenden bringen sich selbst in die Situation, wissen zu sollen, was in der stummen Kommunikation zwischen Schulkind und Fragebogen die Antworten bestimmt. Um die Objektivität ihrer Datenerhebung belegen zu können, sollten sie die Antworten der Schulkinder als ein Resultat der stummen Kommunikation zwischen dem einzelnen Kind und dem Fragebogen darstellen können. Die Antwort wäre ein Produkt der im jeweils einzelnen Kind zum Zeitpunkt des Antwortens ablaufenden innerpsychischen Prozesse. Die Antworten werden durch die Fragen, fast wörtlich zu nehmen, „hervor – gerufen“. Irgendwo im Psychischen entstehen Antworten, die stimuliert von der Prozedur eines vorgelegten Fragebogens, das Abgeben einer schriftlichen Mitteilung fordern. Von welchen Bedingungen diese Mitteilung wirklich determiniert wurde, ist unbekannt. Die kommunikativen Prozesse im schulischen, familiären oder örtlichen Umfeld können Anteil am Zustandekommen der abgegebenen Mitteilung haben. Sicher ist lediglich, dass die spezielle Mitteilung gegeben wurde, aber nicht warum.

Der dargestellte Zusammenhang erklärt die jeweils einzelne Antwort als eine durch die Untersuchungsprozedur hervorgerufene Mitteilung.

Die Annahme einer durch die Untersuchenden in ihren Bedingungen kontrollierten Datenaufnahme trifft nur zu, wenn die in Betracht gezogene psychische Wirklichkeit auf abstrahierende Konstrukte reduziert wird.

Unter der Voraussetzung, dass die Existenz und die Bedeutung der von den Konstrukten überlagerten Mikrostrukturen eines jeden Untersuchungsteiles mitgedacht wird, können oberbegrifflich drei Phasen des Untersuchungsablaufes unterschieden werden. Das ist zum einen die Phase der zielgeleiteten spezifischen Vorbereitung der Datenaufnahme, der dann zweitens die der Datenaufnahme folgt, nach der in der dritten Phase die am Untersuchungsziel orientierte Datenanalyse vollzogen wird. Die eingehende Darstellung der in jeder dieser drei Phasen hierarchisch-sequentiell zu planenden und zu vollziehenden speziellen Arbeitsschritte ist der Inhalt der bis an diese Textstelle hier geführten Erörterungen.

Noch nicht dargestellt wurde, was dann insgesamt mit der Realisierung der Untersuchungsarbeit und dem damit geschaffenen und Hunderte von Einzelergebnissen umfassenden Arbeitsertrag erreicht worden ist. Das wäre identisch mit der Frage, was jetzt mehr als zuvor gewusst wird, nachdem die Einzelergebnisse erarbeitet und vorgelegt worden sind.

Wenn die beiden ersten Phasen des Untersuchungsganges, so wie hier eben zuvor ausgeführt, insgesamt als ein auf die Abgabe von schriftlichen Antworten auf schriftlich vorgelegte Fragen abzielender kommunikativer Prozess aufzufassen sind, läge nahe, den durch die Untersuchung

insgesamt erreichten, auf Hunderte von Einzelergebnissen verteilt und deshalb nun zu strukturierenden Wissenszuwachs auch kommunikationstheoretisch zu ordnen.

Auch von einer solchen Ordnung müsste erwartet werden, dass sie zum Erreichen des für diesen Arbeitsteil gesetzten Zieles beiträgt.

Ungeachtet der noch nicht mitgeteilten Entscheidung, auf welche Kommunikationstheorie die Zusammenfassung und Bewertung der erarbeiteten Einzelergebnisse der gesamten Arbeit bezogen werden könnte, ist eine erste und Konsequenzen nach sich ziehende Aussage bereits vorab durch die Mitteilung der beabsichtigten Bezugnahme gegeben worden. Konsequenzen folgen ihr, weil die kommunikationstheoretische Zuordnung bereits ein erster und die Herstellung einer speziellen Ordnung einleitender Arbeitsschritt ist. Der reduziert die zuvor noch bestehenden Freiheitsgrade bei der Wahl einer anderen, nicht auf Kommunikation bezogenen Ordnung der angestrebten Befundbildung. Die Konsequenzen dieser Restriktion lassen sich leicht übersehen und übergehen, obschon sie für die mit der Untersuchung insgesamt angestrebten Einsicht und für alle aus ihr abzuleitenden Folgerungen die Koordinaten des zu aktivierenden Potentials an Aussagemöglichkeiten setzt. Die Kategorisierung aller mit dieser Untersuchung erhobenen Daten, als Produkte kognitiver Prozesse deklariert sie in ihrem Ursprung ausnahmslos als subjektiv. Alle auf sie bezogenen Zuweisungen und Zuordnungen stehen, soweit nicht nur ihr kommunikatives Auftreten gemeint wäre, unter dem Vorbehalt einer fraglichen Validität. Die müsste jeweils gesondert nachgewiesen werden. Insoweit als die Ergebnisse der Untersuchung als aus kommunikativen Prozessen hervorgegangen bezeichnet werden, unterscheiden sie sich in ihrer Qualität von denen mit anderem Ursprung.

Nach Wittgenstein (siehe dazu Hörmann, 1967) ist für das Herstellen jeglicher Ordnung von einer Vielzahl möglicher Ordnungen auszugehen, von denen jeweils jede durch ihre eigene Struktur mit einem bestimmten Zweck verbunden werden kann. Was Wittgenstein als philosophische Einsicht mitteilte, wäre in der Handlungspsychologie als eine durch die psychischen Bedingungen der Handlungsregulation zwangsläufig entstehende zielgerichtete Struktur zu beschreiben. Die Handlungsregulation erfasst als erstes das antizipierte Ziel und ordnet den dann auszuführenden Handlungsverlauf so zu, dass dieses antizipierte Ziel erreicht werden kann. Die Ordnung der Handlung dient dem Zielerreichen. Eine bestimmte Ordnung ist zum Erreichen eines speziellen Zieles nutzbar.

Die der durchgeführten empirischen Arbeit unterliegende Zielsetzung ist von der diese Arbeit führenden Untersucherin antizipiert worden. Die hat, als allein damit beschäftigte Arbeiterin die Planung gerichtet und mit der Ausführung des Planes den gewünschten kommunikativen Prozess organisiert, stimuliert und den Untersuchungsablauf aktiviert. Aber alle an dem Untersuchungsgeschehen beteiligten Schüler, Schülerinnen und Lehrer wie Lehrerinnen sind ebenfalls tätig geworden und hatten ihre jeweilige Tätigkeit als Teilhaber an dem Gesamtgeschehen regulativ organisiert und dies, wie die Untersuchungsleiterin, ebenfalls von antizipierten Zielen ausgehend, um eine dem Erreichen dieses eigenen Zieles dienende Ordnung des eigenen Tuns zu finden. Wie die insgesamt etwa 2200 teilnehmenden Personen nach welchen Zielen kommunikativ tätig geworden sind, wird von niemandem der Realität entsprechend gewusst. Die Feststellung, eine solche Untersuchung sei in ihren Bedingungen kontrolliert gewesen, ist nur haltbar, wenn von der Mehrheit der individuellen Bedingungen abgesehen wird. Sicher ist nur, dass über die darstellbare Planung und deren mehr oder weniger gelungener Umsetzung in Untersuchungsgeschehen die vorzeigbaren Produkte, die Antworten der Befragten, entstanden sind.

Die Annahme ist zu rechtfertigen, dass unter Bezugnahme auf eine Kommunikationstheorie die bislang nicht hinlänglich erklärte Struktur des als kommunikatives Geschehen bezeichneten Untersuchungsablaufes und damit dann auch die Entstehung der daraus hervor gegangenen kommunikativen Produkte mittelbar aufgeschlossen werden könnte.

Die Einbeziehung eines kommunikationstheoretischen Ansatzes in die Erörterung bringt die Nötigung mit sich, die bislang als Produkte bezeichneten Antworten der teilnehmenden Kinder, Jugendlichen und Lehrkräfte kommunikationstheoretisch einzuordnen. Damit ist mehr als die Wahl eines anderen Wortes verbunden. Die Benennung der jeweils gegebenen Antwort als Produkt wird durch den kommunikationstheoretischen Begriff der Nachricht ersetzt. Andere, aber ähnliche Begriffe wären als Alternativen wählbar gewesen. Der Begriff Produkt bezeichnet das gemeinte Objekt als Ergebnis einer auf das Hervorbringen von Produkten gerichtete Tätigkeit. Unter Nachricht ist eine intendierte und an Empfänger gerichtete Übermittlung von Informationen zu verstehen.

Zunächst ist dann mitzuteilen, dass kommunikationstheoretisch bei Watzlawick und in seinem wissenschaftlichen Umfeld die erforderliche Unterstützung gesucht wird (Watzlawick und Mit., 1969; Schulz von Thun, 2005, Sonderausgabe). Thematisiert werden sollen in der hier weiter zu führenden Erörterung die Eigenheiten von Nachrichten, wodurch Aufschlüsse über psychische Aspekte des Abgebens von Nachrichten erfasst werden könnten. Diese thematische Akzentuierung korrespondierte mit der Nachfrage, was die in der Untersuchung befragten Teilnehmer mit ihren Antworten möglicherweise mitzuteilen beabsichtigten. Dazu ließe sich an dem auch durch Schulz von Thun benutzten psychologischen Konzept ansetzen, das als „Quadrat“ einer Nachricht“ bezeichnet worden ist. Dieses Quadrat könnte innerhalb der vielen kommunikativen Teilprozesse in dieser Untersuchung aus mehreren Perspektiven betrachtet werden. Durch die Untersuchungsprozedur wird ein Ausschnitt aus dem kommunikativen Gesamtprozess herausgelöst und, wie in einer mikroskopischen Untersuchung, fixiert. Die Bearbeitung der Fragebögen ist jeweils der fixierte kommunikative Ausschnitt. Dann sollten nach Schulz von Thun (2005) die als Daten fixierten Mitteilungen der Schulkinder vier kommunizierte Inhalte haben, die durch die Auswahl einer Antwort im jeweiligen Fragebogen als Nachricht weitergegeben worden sind. Falls dieser Sachverhalt zuträfe, wäre er in der die Ergebnisse der durchgeführten Untersuchung speichernden Datenmatrix nicht enthalten, weil immer nur ein Merkmal kodiert wurde, eben die zu einem Item ausgewählte und vorformulierte Antwort.

Die zur eingehenderen Analyse der im Verlaufe der Untersuchung vollzogenen Datenaufnahme zu Rate gezogene Kommunikationstheorie unterstellt, die über die Fragebögen aufgenommenen schriftlichen Nachrichten der Befragten, hätten nicht nur eine kommunikative Seite, sondern vier. Nach Schulz von Thun (2005) sind dies die nachfolgend genannten.

Der *Sachinhalt* einer Botschaft ist der erste der darzustellenden kommunikativen Inhalte, der Seiten einer Nachricht. In der Untersuchung 2034GR sind fast 200 000 Sachinhalte zur Basis der Befundbildung geworden. Danach repräsentieren alle in den voran gegangenen Teilen erarbeiteten und mitgeteilten Einzelergebnisse die mit den Sachinhalten eingegangenen Informationen. Dem Sachinhalten wird in dem Bericht über die Untersuchung der größte Darstellungsraum zugestanden.

Der zweite Aspekt, die zweite Seite jeder Nachricht, ist die *Selbstoffenbarung*. Sie beinhaltet gewollte und unfreiwillige Informationen über den Sender, über seine Art, Gedanken zu entwickeln, bestimmte Dinge wichtig zu finden (Schulz von Thun, 2005, S. 26-27).

Der in der Datenmatrix fixierte Ausschnitt aus dem mit der Untersuchung insgesamt aktivierten kommunikativen Geschehen der durch die Untersuchung insgesamt provozierten Kommunikation enthält anscheinend aber sowohl bei den Schulkindern als auch bei den Lehrkräften ausschließlich Sachinformationen, die über die Nachrichten der Befragten gesammelt worden waren. Die

Schulkinder sagen beispielsweise aus, ob sie sich als vorwiegend heiter oder traurig erleben. So gesehen, ist jede einzelne Frage die Aufforderung zur Selbstoffenbarung: Für die teilnehmenden Lehrkräfte gilt dies mit einer spezifischen Akzentverschiebung ebenso. Sie bekunden mit ihren Äußerungen über die Schulkinder, ob sie im Sinne der jeweiligen Frage gegenüber den Kindern ihrer Schulklasse sensibel sind und Beobachtungen erinnern und zuordnen können, oder ob das nicht gelingt. Unter diesem kommunikativen Aspekt ergibt sich für die Lehrkräfte, wie an anderen Stellen der Arbeit ausführlich erörtert, eine zu vermerkende Inhomogenität dieser Teilnehmergruppe. Eine durchaus umfangreich zu bezeichnende Teilgruppe der Lehrkräfte, weigert sich, die ESDM Bögen auszufüllen. Vielleicht meinen sie, den dazu erforderlichen Zeitaufwand in der Untersuchungssituation nicht auch noch aufwenden zu können, weil sie ohnehin schon überfordert seien. Ungeachtet aller erdenkbaren persönlichen Gründe für die Weigerung ist die auch eine Selbstoffenbarung. In einer weiteren Perspektive wäre auch das Untersuchungsvorhaben selbst, nur für sich genommen, eine Selbstoffenbarung, obschon diese Sichtweise nicht zu den allgemeinen Gepflogenheiten zu rechnen wäre. Solche hier gemeinte Selbstoffenbarung wird, den Regeln für wissenschaftliches Arbeiten folgend, mehr oder weniger entpersonalisiert als Untersuchungsinteresse bezeichnet. Die Selbstoffenbarung wird, vermutlich wegen ihrer Subjektivität, umgangen, obschon das Bekunden eines Interesses zumindest auf eine Persönlichkeit zu beziehen wäre, die eine entsprechende Selbstoffenbarung äußerte. Die offenbarte, dass sich jemand um die potentiellen Folgen des in Griechenland nachweisbaren Unwissens über die Verbreitung emotionaler Störungen unter Schulkindern sorgte. Das lässt sich auch anders gefasst niederschreiben, etwa so wie folgt: "Die vorliegende Arbeit ist aus dem allgemeinen Forschungsinteresse an der Verbreitung emotionaler Störungen unter griechischen Schulkindern entstanden." Vermutlich trifft die Behauptung eines allgemeinen Interesses nicht zu, eher wäre realitätskonform von dem selbstoffenbarten Interesse einer Person auszugehen, eben der, die diese Studie betreibt.

Eine dritte Seite der Nachricht ist der von ihr in die Kommunikation gebrachte Beziehungsaspekt. Die Umstände des betrachteten kommunikativen Handelns geben, wie die Inhalte der übermittelten Nachricht auch, Hinweise auf die Beziehungen zwischen den kommunizierenden Personen.

Weiterhin übermittelt die Nachricht nach Schulz von Thun einen Appell. Unter diesen Begriff fallen alle an den Empfänger der Nachricht übermittelten Forderungen und Aufforderungen.

Eine auf den vierten Aspekt in kommunikativen Prozessen, auf den Appellcharakter von Mitteilungen, zu beziehende allgemeine Feststellung greift der Entwicklung der an dieser Stelle zu führenden speziellen Erörterung vor. Aber allgemein zutreffend ist, dass empirische Untersuchungen mit Kindern oder Jugendlichen die Möglichkeit eines mit den erhobenen Daten verbundenen Auftretens von aus dem Klientel kommenden Appellen fast ausnahmslos übergeht. Die jeweilige Untersuchung veranlasst beispielsweise im Falle von Untersuchungen zum Auftreten depressiver Störungen die beteiligten Kinder und Jugendlichen sich im Hinblick auf das Auftreten von depressionsspezifischen Erlebenszuständen zu äußern und zu offenbaren. Jede solche Mitteilung mit Symptomcharakter wird dann durch die Untersuchungsmethodik verwandelt, etwa zu einem für den Gesamtscore mit anderen aufzuaddierenden Itempunkt. Für das dem Fragebogen antwortende Kind ist die abgegebene Mitteilung möglicherweise mit dem Erleben von Phänomenen verbunden, die bei gegebenem Symptomcharakter, das alltägliche Leben belasten und das emotionale Erleben verdüstern. In Übereinstimmung mit der herangezogenen Referenztheorie ist mit hoher Wahrscheinlichkeit zu erwarten, dass ein unbekannter Anteil von Antworten der teilnehmenden Schulkinder mit dem Appell um Hilfe verbunden ist. Dass Untersuchungsdesign ist nicht darauf eingestellt, solche Appelle zu identifizieren und aufzunehmen. Demnach ist für den Gesamtbefund, unabhängig von weiteren

und eventuell noch nachzureichenden Feststellungen, die folgende Aussage zu treffen: „ Auch die durchgeführte Untersuchung 2034GR hat die mit hoher Wahrscheinlichkeit von einer unbekanntem Zahl von Kindern oder Jugendlichen implizit geäußerten Appelle um Hilfeleistung nicht wahrgenommen. Die Untersuchung hat im wissenschaftlichen Eigeninteresse die Offenbarungen der Schulkinder über das Auftreten von depressiven Symptomen genutzt, ohne die Offenbarung von Hilfsbedürftigkeit, den jeweils gegebenen Appell, zur Kenntnis zu nehmen. Im Hinblick auf Folgeuntersuchungen bleibt zu überdenken, wie, etwa durch eine kleine Zahl von Zusatzfragen für jedes Kind, ein individueller Indikator für selbsterlebte Hilfsbedürftigkeit erfasst werden kann. Durch entsprechend einzuführende Vereinbarungen zur allgemein zugesicherten Anonymität ließe sich eine Betreuung für Kinder mit hohem Indikator für selbsterlebte Hilfsbedürftigkeit vermitteln.“

In diesem Zusammenhang wird hier die These geäußert, eine Psychologie für Kinder und Jugendliche, die fortlaufend in epidemiologischen Untersuchungen die potentiell auftretenden impliziten Appelle um Hilfeleistungen betroffener Kinder und Jugendlicher übergeht, verstoße gegen die ethischen Grundlagen ihrer psychologischen Arbeit.

Wird Schulz von Thun gefolgt, trägt also auch jede einzelne der von den Schulkindern mit den Testbögen übermittelte Nachricht spezifische Informationen zu jeweils vier kommunikativen Aspekten. Für die Analysearbeit an dem Datenmaterial war durch den exemplarischen Bezug auf eine Kommunikationstheorie versucht worden, Strukturhinweise aufzufinden, die bei der Beschreibung des Vorgangs der Abgabe einer Antwort durch das jeweils zu betrachtende Schulkind hilfreich sein könnten. Unter Hinweis auf Schulz von Thun und in der Perspektive kommunikativer Handlungen ließe sich die nachstehende Beschreibung vertreten.

Die projektierte Untersuchung bringt mit ihrer Durchführung die teilnehmenden Schulkinder in ein spezielles und die Kinder aktivierendes und zur Teilnahme motivierendes interaktionales Geschehen. An dem beteiligen sich die Kinder – wie auch ihre Lehrerinnen und Lehrer - durch eigenes Tun. Die Untersuchungsarbeit wendet sich einem definierten kommunikativen Anteil des Geschehens zu und nimmt das dort durch die Tätigkeit der Teilnehmer entstehende Produkt als Objekt ihrer Analysearbeit auf. Ein solches Produkt ist jede Einzelantwort eines teilnehmenden Schulkindes oder einer beteiligten Lehrkraft. Das Gesamtprodukt ist das Gesamtaufkommen an Antworten. Die Regulation der individuellen kommunikativen Tätigkeit beim Abfassen der jeweiligen Antwort, setzt bei dem subjektiv antizipierten Ziel an und führt zur Auswahl der aus dem vorgegebenen Repertoire eines Antwortbogens ausgewählten Antwort, die nun als eigene persönliche Mitteilung – wenn Schulz von Thun gefolgt wird – eine viergliedrige Struktur hat, die den Sachinhalt, die Selbstoffenbarung, den Beziehungsaspekt und den aktuellen Appell trägt. Die vom Sender gemeinte Bedeutung der Nachricht hat die beschriebene viergliedrige Struktur. Die gemeinten Bedeutungen vermögen nur bis zum Empfänger vorzudringen, wenn sie in der Botschaft symbolisiert sind. Der jeweilige Empfänger entschlüsselt die nach überindividuellen regulativen Normen strukturierte und auf ihn überkommene individuelle Nachricht korrekt, wenn er die gesamte viergliedrige Information zu erfassen vermag.

Obschon die viergliedrige Struktur der Bedeutung einer Nachricht durch ein überindividuell wirksames psychisches Regulationsprogramm von allen Beteiligten hergestellt wird, ist die Entschlüsselung der viergliedrigen Bedeutung nur über die vom Empfänger der Mitteilung zu leistende persönliche Interpretationsarbeit zu vollziehen. Bereits dieser Sachverhalt verweist auf die in den nicht zu umgehenden Prozess der Entschlüsselung funktional implementierte Wahrscheinlichkeit für das Auftreten von fehlerhaften Interpretationen der Bedeutung von Mitteilungen. Diese Wahrscheinlichkeit wird durch die in jeder Nachricht unterschiedliche Verteilung der symbolisierten Informationsträger auf die vier Seiten der Nachricht erhöht. In der Untersuchung 2034GR hat der Sachinhalt der jeweils zu betrachtenden individuellen Nachricht,

der Antwort pro Item, einen hoch dominierenden Informationsgehalt. Die Interpretation der Selbstoffenbarung, des Beziehungsaspektes und des Appells in einer Einzelantwort ist hoch spekulativ und führt somit zu hypothetischen Annahmen über den jeweiligen interpretierten Bedeutungsaspekt.

Die in den vorangegangenen Kapiteln dargestellte Analysearbeit steht ebenso in Übereinstimmung mit dieser Sichtweise wie die nun folgende Bildung eines über die Vielzahl der mitgeteilten Einzelergebnisse zu generierende Gesamtbefund. Bevor die Arbeit daran fortgesetzt wird, folgen einige auf kommunikative Aspekte der Untersuchung bezogene Anmerkungen.

Die mit den Antworten von Kindern, Jugendlichen und Lehrkräften aufgenommenen Sachinhalte erlauben qualitative und quantifizierende Aussagen über die Wahrscheinlichkeit der Belastung der untersuchten Schulkinder durch depressive Symptome. Neben den depressionsspezifischen Sachinhalten der aufgenommenen Antworten gestatten die auch die Bildung von Hypothesen zu den Selbstoffenbarungen, den Beziehungsaspekten und dem potentiellen Auftreten von Appellen.

Die mit jeder Mitteilung abgegebenen Informationen, die als Beziehungsaspekt der Botschaft (Schulz von Thun) bezeichnet werden, scheinen nach der hier vertretenen Einschätzung schwieriger erschließbar zu sein als die zum Aspekt der Selbstoffenbarung. Sicher ist, dass die überwiegende Mehrheit der Schulkinder durch die für sie geltenden und schulisch wirksamen individuellen Beziehungsstrukturen nicht daran gehindert war, die Fragebögen zu beantworten. Immerhin könnte die Qualität der Antworten durch Beziehungsaspekte beeinflusst sein. Die persönlichen Erfahrungen der Versuchsleiterin mit den Kindern und Jugendlichen in den griechischen Schulen lassen auf ein starkes Bedürfnis nach Kommunikation über das eigene schulische Erleben schließen. Nahezu in jeder Schule ergaben sich Erlebnisse, die auf eine jeweils durch die Untersuchung aktuell stimulierte Bedürftigkeit zur Aussprache über schulische Beziehungsaspekte. Diese Erfahrungen berechtigen zu nicht mehr als zu der nachstehenden Hypothese, aber die verdient beachtet und geprüft zu werden: „Unter den griechischen Schulkindern der untersuchten Stichprobe besteht ein auffallend starkes Bedürfnis mit fachkundigen aufgeschlossenen Erwachsenen über die emotionalen Aspekte des Schulalltags zu kommunizieren.“ Diese Hypothese lässt sich begründet verallgemeinern und relativ leicht überprüfen. Falls sie sich bestätigen ließe, hätte dies für die Psychologie eine eher bescheidene Bedeutung, nicht aber für die pädagogische Praxis in griechischen Schulen. Ein solches Ergebnis sollte als Nachweis für ein von den Schulkindern erlebtes Defizit verstanden werden. Das Erkennen dieses Defizits könnte zum Anlass genommen werden, mit entsprechenden pädagogischen Interventionen zu dessen Reduktion zu erarbeiten.

Eine eindeutigere Botschaft zum Beziehungsaspekt der Kommunikation zwischen Versuchsleiterin und an der Untersuchung teilnehmenden Personen übermittelte eine Teilgruppe der zur Abgabe von Einschätzungen einiger depressionsspezifischer Verhaltensmerkmale an den von ihnen unterrichteten Schulkindern aufgeforderten Lehrkräfte. Sie füllten die zur Bearbeitung ausgeteilten ESDM-Bögen nicht aus. Dadurch teilten sie implizit mit, sich in der gegebenen Situation nicht entschließen zu können, den Aufwand einer über relativ wenige Items reichenden Einschätzung beobachtbarer Verhaltenseigenheiten der von ihnen unterrichteten Kinder zu übernehmen. Die Gründe für die Weigerung sind nicht bekannt, aber der Sachverhalt der Verweigerung ist eindeutig. So bleibt dies auch als Teil des Gesamtbefundes mitzuteilen. „Eine Teilgruppe der an der Untersuchung teilnehmenden Lehrkräfte beteiligte sich nicht an der im Untersuchungsablauf erbetenen Einschätzung emotionaler Eigenheiten der von ihnen unterrichteten Kinder.“ Die Gründe dafür mögen interpersonal unterschiedlich sein. Gemeinsam ist allen Ablehnungen, dass die sie vertretenden Lehrkräfte nicht von der

pädagogischen Bedeutung der emotionalen Befindlichkeit von Schulkindern überzeugt zu sein scheinen, der dass sie zur Aufgabe ihres Widerstandes gegenüber einer an sie gerichteten spezifischen Aktivitätsforderung genötigt hätte. Der mitgeteilte Sachverhalt ist insoweit bedeutsam, als er der These gegenüber steht, die Sorge um die psychosoziale Befindlichkeit und die Gesundheit der Schul Kinder sei eine auch die Schule und die Schulpädagogik verpflichtende soziale und gesellschaftliche Aufgabe. Die durch den mitgeteilten Sachverhalt erkennbar gewordenen Problematik wird hier als hinreichender Anlass interpretiert, der empirisch begründeten Hypothese vom Widerstand einer beachtlichen Teilgruppe aller (griechischen) Lehrkräfte gegen die Verpflichtung, sich pädagogisch zielgerichtet um die emotionale Befindlichkeit der von ihnen unterrichteten Schul Kinder bemühen zu sollen. Eine solche Hypothese ist mit spezifischen empirischen Untersuchungen prüfbar. Den für das Schulwesen und die Lehrerbildung verantwortlichen Personen und Institutionen wäre anzuraten der dargestellten Problematik mit den ihnen zur Verfügung stehenden Mitteln nachzugehen. Die pädagogische und schulpraktische Befassung mit dieser Problematik erscheint notwendig. Sie darf aus zeitlichen Gründen zunächst nicht vom Vorliegen weiterer psychologischer Untersuchungsergebnisse abhängig gemacht werden. Die schulbezogene Befassung mit der Thematik wäre auch überaus nützlich, wenn der vermutete Anteil von widerständigen Lehrkräften deutlich geringer sein sollte als in der durchgeführten Untersuchung. Die Thematisierung dieser Problematik würde in jedem Fall eine Sensibilisierung der Lehrkräfte gegenüber dem Problem der emotionalen Belastungen und Störungen von Schülerinnen und Schülern erhoffen lassen.

Die in diesem Kapitel geäußerte Erwartung, mit dem Rückgriff auf eine Kommunikationstheorie analytisch zu nutzende regulative Strukturen für die von den teilnehmenden Schulkindern und Lehrkräften abgegebenen Einzelantworten gewinnen zu können, hat sich, wie darzustellen war, erfüllt:

1. Die Einzelantworten sind individuelle Nachrichten, die einem viergliedrig strukturierten kommunikativen Ziel dienen.
2. Die Bedeutung der als Nachricht zu verstehenden Einzelantwort muss in der Analyse, wie von jedem Empfänger einer Nachricht, interpretiert werden.
3. Eine vollständige Interpretation erfasst die vom Absender mitgedachte Bedeutung aller der nach Schulz von Thun kommunizierten vier Aspekte einer Nachricht.
4. Jede Interpretation einer Nachricht stößt unumgänglich auf die Wahrscheinlichkeit des Auftretens von Interpretationsfehlern.
5. In der Untersuchung 2034GR besteht die relativ geringste Wahrscheinlichkeit für das Auftreten von Fehlinterpretationen für den kommunikativen Aspekt des Sachinhalts einer Nachricht.
6. Die Interpretation von Selbstoffenbarungen, Beziehungsaspekten und Appellen sollte in dieser Untersuchung nur zur Bildung von hypothetischen Aussagen genutzt werden.
7. Hypothetische Aussagen bieten die Möglichkeit ihrer Überprüfung und die damit erreichbare Veränderung des Erkenntnisstandes an.

Die vorstehenden sieben Aussagen nennen Grundsätze, an denen die Datenanalyse in dieser Untersuchung orientiert ist. Spätestens hier ist dann daran zu erinnern, dass zu Beginn der Untersuchung keine einsetzbaren griechischen Verfahren zur Verfügung standen. Von vier dann eingesetzten Messverfahren sind drei für diese Studie entwickelt worden. Ob sich alle eingesetzten Verfahren in der Untersuchung bewährt hatten, wird nun vor der Abfassung eines Gesamtbefundes erfragt.

13.2 Befunde über die Bewährung griechischer Versionen depressionsspezifischer Messinstrumente

Als Arbeitsmittel zur Gewinnung der für die Beantwortung der aufgeworfenen Forschungsfrage notwendigen empirischen Daten waren insgesamt vier, aus der psychologischen Literatur bekannte Messinstrumente eingesetzt worden. Die drei depressionsspezifischen Verfahren DTK, CDI (Selbsteinschätzung der Schulkinder) und ESDM (Fremdeinschätzung der Schulkinder durch Lehrkräfte) fanden in dieser Studie ihre erste Anwendung im griechischen Sprachraum. Die verwendeten griechischen Versionen des DTK und des CDI entstanden nach der Methode der „font-back translation“. Da die Adaption eines fremdsprachigen Verfahrens mehr als eine Übersetzung seiner Items in die Zielsprache erfordert, wurde ihre Anwendung in der vorliegenden Arbeit mit der für den weiteren Arbeitsverlauf bedeutungsvollen Frage verbunden, ob deren Messqualitäten den Werten für die jeweils ursprüngliche Fassung entspräche und den Einsatz in einer griechischen Untersuchung erlaubte. Das vierte eingesetzte Verfahren zur Erfassung allgemeiner psychischer Belastung und von prosozialem Verhaltens, das SDQ, war bereits ins Griechische übersetzt und dabei auch einer Prüfung seiner Messqualitäten unterzogen worden (Adaption: Bibou-Nakou, I., Stogiannidou, A., Kioseoglou, G. & Papageorgiou, V.; Bibou-Nakou, Kioseoglou & Stogiannidou, 2001). Seine Anwendung in unserer Studie war in erster Linie mit einer erneuten Prüfung der psychometrischen Eigenschaften seiner griechischen Version verbunden, nachdem die sich in zwei früheren griechischen Studien (Bibou-Nakou et al., 2001; Kypraiou, 2002) als nicht besonders befriedigend erwiesen hatten. In dem zu erwünschenden Fall, dass bei der Anwendung des griechischen SDQ in unserer im Vergleich durchaus größeren und repräsentativen Stichprobe die Messqualitäten den messtheoretischen Standards genügen würden, könnten aus dem Verfahren relevante Hinweise auf das Bestehen komorbider Störungsbilder in der untersuchten Stichprobe erwartet werden.

Die einzelnen Ergebnisse aus den Analysen zur Überprüfung der eingesetzten Messverfahren sind im Kapitel 6 ausführlich dargestellt worden. Im Folgenden sollen sie zusammenfassend gewürdigt werden. Das beginnt mit dem nachstehenden ersten Befund als erstem der speziellen Teile des Gesamtbefundes.

1. Befund zu den Messqualitäten der griechischen Version des DTK

Der DTK hat bei seiner griechischen Anwendung innerhalb der in dieser Arbeit beschriebenen Untersuchung seine aus deutschsprachigen Publikationen bekannt gewordenen Messqualitäten ungemindert erhalten. Das belegen die Ergebnisse aus der Itemanalyse (Trennschärfe und Schwierigkeitsgrad), der Reliabilitätskontrolle und der Validitätsprüfung (faktorielle und konvergente Validität).

Die einzige Abweichung zwischen den messtechnischen Werten der griechischen und deutschen Fassung des DTK ist in den Ergebnissen aus der Faktorenanalyse zu finden und betrifft Items der dritten Skala „Müdigkeit/ autonome Reaktionen“, die nach Schlafgewohnheiten fragen. Diese Fragen wurden von griechischen Kindern anders beantwortet als von deutschsprachigen. Aufgrund der ansonsten hohen Übereinstimmung der Ergebnisse aus der Faktorenanalyse und im Hinblick auf die Sachinhalte der Antworten auf abweichende Items (Schlafgewohnheiten) scheint die Annahme einer interkulturell begründeten Unterschiedlichkeit höheren Erklärung als andere Annahmen zu bieten.

Als Gesamtbefund ist zu vermerken, dass die griechische Version des DTK sich als ein einsetzbares Messinstrument mit guten Messqualitäten erwiesen hat. Die Qualitäten der griechischen Version stehen denen der deutschen Originalfassung nicht nach.

2. Befund zu den Messqualitäten der griechischen Version des CDI

Die innerhalb dieser Arbeit gefundenen Messqualitäten der griechischen Version des CDI erreichen die Standards der klassischen Messtheorie bezüglich der Gütekriterien eines Verfahrens zu Forschungszwecken. Darüber hinaus replizieren sie weitgehend die der originalen amerikanischen Fassung.

Im Hinblick auf die vieldiskutierte Faktorenstruktur des Verfahrens ist zu erwähnen, dass, wie bei der großen Mehrheit der internationalen Studien, auch aus den Antworten der untersuchten griechischen Kinder eine eigene – im übrigen an sich logische- Variante der Faktorenstrukturen ergeben hat, die von der im CDI-Manual (Kovacs, 2004) dargestellten Verfahrensstruktur abweicht.

Als Gesamtergebnis zur Bewährung des CDI in der griechischen Version ist die Einstellung der Messqualitäten der englischsprachigen Originalfassung hervorzuheben. Die Messqualitäten der griechischen Version stellen die Standards für Forschungsinstrumente ein und sind entsprechend einsetzbar.

3. Befund zu den Messqualitäten der griechischen Version des SDQ

Die innerhalb dieser Arbeit gefundenen Ergebnisse aus den Analysen zu den Messqualitäten einer von Bibou-Nakou ins Griechisch übersetzten Version des Verfahrens SDQ erweisen sich für die 5 Subskalen als unbefriedigend. Aus messtechnischer Sicht dürfte nur die Gesamtskala, die Hinweise auf das selbstberichtete Erleben internalisierter und externalisierter Schwierigkeiten zu bieten vermag, als reliabel bezeichnet werden. Die hier gefundenen Reliabilitätswerte stimmen mit den entsprechenden Ergebnissen aus griechischen Studien (Bibou-Nakou et al., 2001; Kypraiou, 2002) und den in anderen europäischen Ländern durchgeführten Anwendungen (ausführliche Literatur dazu in Abschnitten 6.5 und 11) überein. Die im Abschnitt 11 dargestellten Analysen zu den verfahrensternen korrelativen Beziehungen zwischen den SDQ-Items entsprechen den Ergebnissen aus Studien außerhalb des griechischen Raumes. Sie bestätigen so die These, dass neben der geringen Itemzahl jeder Skala auch Faktoren wie die Formulierung der Aufgaben (positive oder negative) und deren diskutierbare inhaltliche Relevanz zum jeweiligen zu erfassenden Konstrukt als mögliche Erklärungen für die niedrige Reliabilitätswerte der SDQ-Subskalen heran gezogen werden können. Diese Übereinstimmung ist insofern von Bedeutung, als sie empirisch darauf hinweist, dass die auch innerhalb dieser Arbeit gefundenen niedrigen Reliabilitätswerte eher in Verbindung zu strukturellen Besonderheiten der originalen Fassung des SDQ zu sehen sind und hochwahrscheinlich nicht in Zusammenhang mit der vorliegenden griechischen Übersetzung des SDQ stehen.

Für die griechische Version des SDQ ergibt sich nur eine auf die Gesamtskala bezogene und auch dann eingeschränkt bleibende Anwendbarkeit.

4. Befund zu den Messqualitäten der griechischen Version des ESDM

Die innerhalb dieser Arbeit gefundenen statistischen Werte zur Reliabilität der verwendeten, ins Griechische übersetzten Version des spanischen ESDM sprechen für eine zufrieden stellende Reliabilität dieser Skala und stimmen mit entsprechenden Ergebnissen aus anderen ESDM-Studien (u.a. Casanova, 2003; Domènech & Polaino, 1990; Ezpeleta, 1987) überein. Den Messqualitäten des Verfahrens ist zu vertrauen.

Fraglich scheint allerdings, was die Skala misst. Ihre Validität wird hier auf Grund der durchgeführten Analysen problematisiert. Die Selbsteinschätzungen der Schulkinder zu ihrer Betroffenheit vom Auftreten depressiver Symptome korreliert nur schwach mit den

Fremdeinschätzungen durch die Lehrkräfte. Einige Analyseergebnisse legen die Annahme nahe, die Lehrkräfte urteilten über Schulkinder eher innerhalb eines schulischen Bezugssystems und dessen schulbezogenen Kategorien und Dimensionen statt in der durch das Verfahren ESDM verlangten empathischen und auf subjektive Befindlichkeiten gerichteten Perspektive. Sie verfügen anscheinend nur über geringe Kenntnisse zur persönlichen Befindlichkeit der Kinder ihrer Klasse. Sie antworten nicht auf die durch die Skala ESDM gestellte Fragen nach dem Zustand, in dem sich die Kinder anscheinend befinden, sondern eher auf die nicht gestellte Frage nach den „guten und den schlechten Schülern und Schülerinnen“.

Empirisch begründbar ist die Annahme, dass Urteile aus Selbsteinschätzungen einem klinisch diagnostischen Urteil näher stehen als Instrumente zur Aufnahme von Fremdeinschätzungen. Diese Zuordnung könnte insoweit als inhaltswalide gelten, als Diagnosen von depressiven Störungen ebenfalls vornehmlich auf Selbsteinschätzungen beruhen. Die Skala ESDM erscheint als ein Verfahren, das eher zur Identifizierung des Grades der Distanz von Lehrkräften zu den emotionalen Selbstwahrnehmungen von Schulkindern als für einen Beitrag zur „Information über emotionale Zustände von Schulkindern“ geeignet ist. Diese Feststellung sollte nicht als Polemik verkannt werden, die Einschätzung einer solchen Eignung dieses Messverfahrens wird hier als empirisch nutzbar gesehen.

5. Gesamtbefund zur messtechnischen Bewährung der eingesetzten Verfahren

Die Messqualitäten der griechischen Versionen von DTK, CDI und ESDM erweisen sich in der durchgeführten Untersuchung als gut. Den mit diesen Verfahren gewonnenen Daten ist insoweit zu vertrauen. Für das SDQ gelten demgegenüber deutliche Einschränkungen seiner Nutzbarkeit. Insgesamt sprechen die Ergebnisse für eine interkulturelle Homogenität der Messqualitäten aller eingesetzten Verfahren. Insoweit sind auch die dann eher doch unbefriedigenden Analyseergebnisse für das SDQ wenigstens in dieser Perspektive auch positiv zu bewerten.

13.3 Hinweise zum Auftreten depressiver Symptome und depressiver Störungen in der untersuchten Stichprobe: Hinweise über Daten aus einem Messverfahren

Die Datenbasis der durchgeführten Untersuchung ist in mehreren Arbeitsschritten eingehend analysiert worden. Diese Analysen haben Sachverhalte erschlossen, die in einigen Hundert von Teilbefunden festgehalten worden sind. Diese Feststellungen werden nun als empirische Grundlage für die Beantwortung der Forschungsfrage und den aus ihr ableitbaren Nachfragen genutzt. Aus dem zum Anfang der Erörterungen dargelegten Forschungsinteresse sind im Verlaufe der Ausführungen Forschungsfragen abgeleitet worden, die dann im Kapitel 4.1 nach der präzisierten Bestimmung des Untersuchungsgegenstandes auf den eingestellt worden sind. Diese Formulierung der Forschungsfrage, die so gehalten worden war, dass weitere Fragen aus ihr herleitbar wurden, findet sich nachfolgend wiederholt: „Welche Hinweise auf die Belastung der untersuchten griechischen Schulkinder durch depressive Symptome und depressive Störungen enthalten die mit dieser Untersuchung aufgenommenen Daten und welche Schlüsse ergeben sich daraus für die griechischen Schulkinder insgesamt?“

Die erarbeiteten Teilergebnisse sind, der eingesetzten Forschungsfrage folgend, zunächst auf Hinweise abzusuchen, die das Auftreten von depressiven Symptomen oder depressiven Störungen bei den Schulkindern der untersuchten Stichprobe annehmen lassen. Der Begriff „Hinweise“ steht in seiner allgemeinen Bedeutung für aufzufindende Sachverhalte, die auf die Existenz von anderen Gegebenheiten verweisen, ohne die bereits nachweisen zu können. Damit ist, wenn Gewissheit über die zu vermutenden Gegebenheiten erlangt werden sollte, Hinweisen mit der

Absicht nachzugehen, deren Aussagewert unter dem höchsten erreichbaren Grad an Gewissheit zu prüfen. Das danach abzugebende Urteil befindet sich nach dem gegebenen Erkenntnisstand darüber, ob die fraglichen Gegebenheiten als existent angenommen werden und dann vorerst auch als gegebener Erkenntnisstand gelten dürfen. Hinweise sind in dem damit dargestellten Erkenntnisprozess die Ergebnisse einer vorausgegangenen Informationsnahme, die Annahmen konstituieren lässt, deren Aussagewert zu prüfen bliebe. Auch die Hinweise selbst ließen sich, den vorliegenden Umständen entsprechend, auf den Grad des für sie ausgeprägten Hinweischarakters überprüfen.

Der Hinweischarakter ist identisch mit der mutmaßlichen Beweiskraft des Hinweises. Diese Wahrscheinlichkeit lässt sich als Verhältnis von p zu q ausdrücken, wobei p der Anteil von 1 ist, mit dem 1 als Symbol für die uneingeschränkte Bestätigung des über den Hinweis vermuteten Sachverhaltes steht. Die Abschätzung der Wahrscheinlichkeit, mit der Hinweise sich als zutreffend erweisen werden, ist für den Entschluss zur weiteren Verfolgung der Hinweise bedeutsam. Niedrigere Wahrscheinlichkeiten für die Bestätigung der Hinweise werden in der Mehrzahl der Fälle auch zu niedrigeren Bestätigungen führen als die Mehrzahl der Hinweise mit höherer Wahrscheinlichkeit des Erreichens einer Bestätigung. Aber jedem p -Wert unter 1 steht der q -Wert für das Verfehlen der Bestätigung des Hinweises gegenüber. Sichere Hinweise schließen die Begriffsbestimmung aus und setzen die mit an Sicherheit grenzende Wahrscheinlichkeit einer Bestätigung als höchsten erreichbaren Grenzwert ein.

Die Informationsnahme, über die hier zu berichten war, ist die durchgeführte Untersuchung an einer Stichprobe von 2034 griechischen Schulkindern. Die sich daraus ergebenden Hinweise zum Auftreten von depressiven Symptomen und depressiven Störungen unter den Schulkindern der Stichprobe sind in den Teilergebnissen niedergelegt. Der Grad der Wahrscheinlichkeit mit dem diese Hinweise sich als zutreffend bestätigen sollten ist unterschiedlich. Darüber war im Einzelnen bereits bei der Abfassung der Teilergebnisse zu entscheiden gewesen.

Die Forschungsfrage verlangt ausdrücklich Antworten, die das Erfassen von Hinweisen betreffen, die zum einen Aussagen über die Wahrscheinlichkeit des Auftretens depressiver Phänomene in der untersuchten Stichprobe griechischer Schulkinder und zum anderen auch solche über griechische Schulkinder allgemein gestatten. Sowohl das Auftreten einer relativ großen Zahl von im weiteren dann auch unterschiedlichen Hinweisen als auch die notwendig werdende Unterscheidung von Aussagen zur Stichprobe einerseits und zur Population griechischer Schulkinder andererseits verlangen die strukturierte Gliederung der Antworten auf die Forschungsfrage.

Das Verlangen, auftretende Hinweise zu nennen, setzt voraus, dass überhaupt solche Hinweise aufgetreten sind. Im Hinblick auf die Ergebnisse der internationalen Forschung zum Auftreten depressiver Störungen bei Kindern und Jugendlichen nach ICD und DSM und auch unter Kenntnisnahme einiger griechischer Untersuchungen war das Auftreten von Hinweisen zu erwarten. Da Erwartungen nie Gewissheiten sind und Gegenerwartungen gegen sich haben, wird ein spezifischer erster Befund ausdrücklich und trotz der als nahezu sicher geltenden Wahrscheinlichkeit dafür, auf das Auftreten von Hinweisen bezogen.

1. Befund (ST= die Stichprobe betreffend): Die durchgeführte Untersuchung zur Erfassen von Hinweisen zum Auftreten depressiver Symptome und depressiver Störungen in einer Stichprobe griechischer Schulkinder führte zur Aufnahme einer so großen Zahl aussagekräftiger spezifischer Hinweise, dass mit an Sicherheit grenzender Wahrscheinlichkeit vom Auftreten depressiver Belastungen von Schulkindern der untersuchten Stichprobe in einem über Einzelfälle hinausgehenden und epidemiologisch relevanten Umfang auszugehen ist.

Der mitgeteilte erste Befund scheint eine Selbstverständlichkeit einzustellen. Das durchaus nicht selbstverständliche und auch gesellschaftliche Gewicht des ersten Befundes wird deutlich, wenn nach den daraus zu ziehenden Folgerungen gefragt wird. Die werden hier wie folgt gesehen. Für den internationalen Forschungsstand ist die erste Aussage eine Bestätigung der allgemein vertretenen Annahme, dass depressive Symptome auch unter griechischen Schulkindern auftreten. Das dürfte eine Reihe von absehbaren Nachfragen auslösen, auf die dann bereits die nachfolgend mitgeteilten Befunde eingehen werden. Für die Familien der betroffenen Kinder, für die zuständigen Schulen und Gemeinden besagt der Befund, dass unter den von ihnen betreuten Schulkinder eine beachtenswerte Teilgruppe lebt, deren Angehörige in ihrer Entwicklung durch depressive Phänomene bedroht sind. Damit ist, zunächst noch an einigen Orten Griechenlands die Frage zu stellen und zu beantworten, was aktuell, kurzfristig und mittelfristig für diese Kinder und ihr soziales Umfeld getan werden kann, getan wird und getan werden soll.

Der erste, eben formulierte Befund zum mit an Sicherheit grenzender Wahrscheinlichkeit Bestehen von depressiven Belastungen unter den 2034 untersuchten Schulkindern beantwortet zusammengefasst die Frage, ob sich die theoriegeleitete Erwartung der Auffindbarkeit von Hinweisen zum Auftreten von depressiven Symptomen, Episoden und depressiven Störungen in der untersuchten Stichprobe von griechischen Kindern bestätigen lassen oder nicht. Die durch zahlreiche Einzelergebnisse gestützte Bestätigung führt an dieser Stelle zu der nun notwendigen Nachfrage, welcher Art die hier gemeinte Hinweise sind und worauf sie inhaltlich hinweisen.

Eine erste Gruppe von Hinweisen bezieht sich auf das selbstberichtete Erleben von depressiven Symptomen. Dass die über DTK- und CDI-Fragenlisten aufgenommenen Erlebnisinhalte jeweils auch depressionsrelevante Phänomene betreffen, darf unter Hinweis auf den theoretischen Entwicklungshintergrund beider Verfahren angenommen werden. Beide Messinstrumente wurden inhaltlich auf die DSM- und ICD-Kriterien für depressive Störungen bei Kindern und Jugendlichen eingestellt. Die Befunde aller international bekannt gewordenen einschlägigen Untersuchungen sprechen dafür, dass depressive Beschwerden, zunächst noch auf der Symptomebene und unter Vernachlässigung des Auftretens als psychische Störung betrachtet, bereits bei Kindern und dann vor allem unter Jugendlichen auftreten. Das wurde auch durch diese Arbeit bestätigt, indem aus den Analysen der DTK- und CDI-Daten hervorging, dass alle vom DTK und CDI mit den Fragelisten repräsentierten Phänomene unter den beteiligten griechischen Schulkindern bekannt waren.

2. Befund: Die über CDI und DTK erfassten Daten aus einer Stichprobe von griechischen Schulkindern sprechen für das Auftreten eines jeden durch die Messinstrumente verbal präsentierten depressiven Symptoms. Jedes dieser Symptome ist in der untersuchten Stichprobe bekannt.

Der 2. Befund stellt fest, dass alle durch die Messinstrumente angesprochenen depressiven Phänomene unter den Schulkindern der Stichprobe bekannt waren. Das leitet zu der Nachfrage über, wie viele der Kinder und Jugendlichen sich von welchen Symptomen betroffen fühlten. Aus den Analysen des Antwortverhaltens der Schulkinder bei Vorlage der DTK- und CDI-Fragenlisten hat sich ergeben, dass in den Fragenlisten des CDI und DTK einige Phänomene als Items auftreten, die für die Mehrheit der beteiligten Kinder zum Zeitpunkt der Untersuchung persönlich relevant waren. Die Angst, etwas falsch zu machen, die Sorgen über die eigene Zukunft, Alpträume, Schwierigkeiten beim Einschlafen und Aufstehen (Items 33, 50, 8, 45 und 46 im DTK) wie auch Unentschlossenheit, Entwertung der eigenen Person nach dem Vergleich mit Gleichjährigen und eine pessimistische oder hoffnungslose Erwartungshaltung (Items 13, 24 und

2 im CDI) stellen Aussagen dar, die mehr als die Hälfte der untersuchten Kinder betreffen. Der Grad der Betroffenheit variiert. Für die genannten DTK-Items bekannten sich jeweils, von Item zu Item schwankend, zwischen 60,0% und 75,0% der Angehörigen der Stichprobe als von dem jeweiligen Symptom belastet. Für die CDI-Items variierten die entsprechenden Werte zwischen 62,3% und 74,6%. Der 3. Befund fasst diese Feststellungen zusammen.

3. Befund: Einzelne depressive Symptome, die in dieser Untersuchung durch 9 Items symbolisiert waren, sind unter den Schulkindern der Stichprobe so verbreitet, dass sich von denen, mit den Items variierend, zwischen 60 und 75% als betroffen erklärten.

Indem die gegebenen Antworten von den Kindern selbst ausgewählt wurden, sind sie auch insoweit uneingeschränkt inhaltsvalide. In diesem Sinne sprechen sie zunächst auch nur für sich selbst. Sie sind Selbstzuschreibungen von emotional getönten Erlebenszuständen.

Wird nach deren Bedeutung gefragt, sind die aufzufindenden Antworten von dem jeweils gewählten Bezugssystem abhängig. Bedeutung haben Sachverhalte für oder im Hinblick auf etwas. Indem die befragten Schulkinder aus der mit CDI und DTK angebotenen Itemmenge die auswählen, die für von ihnen bereits erlebte Zustände stehen, erhalten die Inhalte der Items eine das Erleben der befragten Kinder repräsentierende Bedeutung. Sie geben, unabhängig von jeder speziellen Testtheorie bekannt, was die befragten Kinder nach ihren Selbstbekundungen, neben anderem, auch erleben. Folgte man Regelungen der klassischen Testtheorie, müssten, was methodisch sinnvoll ist, die hochfrequent selbstberichteten und somit aus dem alltäglichen Leben mitgeteilten Schwierigkeiten statistischen Analysen unterzogen werden. Das ermöglichte das Erkennen von Aussagetrends, die nicht als zufällig variierend zu verstehen wären. Die unbearbeiteten Aussagen für sich sind pure Bekundungen des Eigenerlebens und lassen in dieser Form weder die Bildung unterschiedlicher Risikogruppen noch diagnostische Aussagen zu. Die behauptete Inhaltsvalidität und die hochfrequente Nennung der aufgeführten Items beweist vor jeder weiteren Aussage die Normalität des Auftretens einiger depressionsspezifischer Erlebenszustände, indem eine Zweidrittel-Mehrheit der Schulkinder angab, sie erlebt zu haben. Was die Mehrheit erlebt, ist als Normalität zu verstehen. Der 3. Befund steht so genommen als Beleg für die Normalität des Auftretens depressiver Symptome auch unter griechischen Schulkindern. Als Grundeinsicht gilt, dass subjektiv störende oder belastende emotionale Zustände, so beispielsweise Angst, unter Kindern und Jugendlichen immer wieder und in situativen Zusammenhängen auftreten. So sind dann auch, wie der genannte Befund nachweist, depressive Symptome oder Episoden bei Schulkindern alltäglich zu beobachtende Phänomene. Dem Sprachgebrauch im Deutschen folgend, besagt dies nicht, dass sie täglich in einem gemeinten Individuum auftreten. Wenn das Auftreten von Angst nach vielen Angsttheorien im Zusammenhang mit der Wahrnehmung von Gefahren steht, liegt nahe, auch nach entsprechenden mutmaßlichen Zusammenhängen für das Auftreten depressiver Symptome zu fragen. Verfolgenswert scheint die hier geäußerte Mutmaßung, dass deren psychoregulative Bedeutung im motivationalen Bereich liegt und von dort durch das Erleben von Unzufriedenheit die Widerständigkeit des eigenen Verhaltens fühlbar stimuliert und reguliert. Dann wäre das Entstehen einer psychischen Störung durch den progressiven Verlust der Widerständigkeit gegenüber schwer ertragbaren Umständen gekennzeichnet. Diese psychoregulative Interpretation stünde dem Konzept der erlernten Hilflosigkeit nach Seligman sehr nahe, wäre aber auch in Übereinstimmung mit biologischen Verursachungstheorien. In einer solchen Sichtweise wird die depressionsspezifische Bedeutung von Interventionen jeder Art akzentuiert, die Widerständigkeit anregen, entwickeln und stärken.

Die Tatsache, dass die zuvor beschriebene Häufigkeit des Auftretens einzelner depressiver Symptome zugleich auch die Vertrautheit der Mehrzahl aller untersuchten Kinder mit diesen Schwierigkeiten beweist, belegt die Normalität des Auftretens dieser Phänomene. Mit anderen Worten ist dies die in Gruppen von Kindern und Jugendlichen aufweisbare Alltäglichkeit depressiver Symptome und Episoden. Diese Fakten verbieten jede Form der Pathologisierung des Auftretens einzelner depressiver Symptome. Im Zusammenhang damit wäre die Durchdringung des beschriebenen Phänomens von einigem wissenschaftlichen Interesse. Möglicherweise entsteht die berichtete Häufigkeit des Auftretens unter Einwirkung von altersspezifischen interpersonellen Sprachreglungen. In entsprechender sozialer Umgebung behaupten Jungen überzufällig häufig, nicht von Angst betroffen zu sein. Anhänger von Sportmannschaften setzen häufiger als deren sportliche Kontrahenden auf den Sieg ihrer Mannschaft, oder hoffen, bei zu hohen Leistungsunterschieden, auf Wunder im Ablauf von Wettkämpfen und erträumen sie sich. Möglicherweise entwickeln sich unter Schulkindern auch sozial geprägte Gepflogenheiten, die dazu führen, dass hohe interindividuelle Übereinstimmung zwischen individuellen Selbstberichten entsteht. Weiterhin wäre zu wissen von Bedeutung, ob die depressionsspezifischen Selbstzuschreibungen kurzfristig variieren oder mittelfristig, falls sie sich nicht doch als dauerhafter erweisen sollten. Das inhaltlich Gemeinsame an den als hochfrequent auftretenden Symptomen ist, dass sie mit Befürchtungen oder Folgen von Bedrohungen in Zusammenhang stehen. Insofern könnte das Auftreten der genannten Symptome bei einer 2/3 Mehrheit der befragten Schulkinder zu einer Aussage wie der nachstehenden veranlassen: „Im Erleben dieser Mehrheit ist die Wahrnehmung einer anscheinend dauerhaften Bedrohung repräsentiert.“ Die Frage drängt sich auf, welche Lebensereignisse oder welche Institutionen des alltäglichen Lebens das Potential haben, überdauernde Bedrohungen erleben zu lassen. Unter anderem stünde die Schule im Verdacht, vielleicht eben auch diese Wirkung, diese Nebenwirkung zu verursachen. Um dieser Bedrohung entgegen wirken zu können, bedarf es anscheinend dauerhafter Abwehr. Dann könnte eine darauf bezogene Hypothese sein, dass die Ergebnisse aus der Durchführung eines CDI oder des DTK während der Schulzeit einerseits und in den Sommerferien andererseits für identische Stichproben signifikante Unterschiede aufweisen werden. Befürchtungen verweisen im Allgemeinen auch bei Schulkindern auf Lebensumstände, die sich nach Meinung der Betroffenen verändern sollten. Für solche Veränderungen stehen entsprechende Ziele, die sich möglicherweise auch durch eigene erfolgreiche Handlungen erreichen ließen, was dann gleichbedeutend mit der Möglichkeit wäre, einige Befürchtungen aufgeben zu können. Dieser denkbare Zusammenhang wäre dann vermutlich mit einer Präventionsstrategie zu stützen, die darauf abzielte, die auf Problemreduktion gerichtete Tatkraft von Schulkindern und ihren Optimismus in der Einschätzung der eigenen Möglichkeiten zur Überwindung von Problemen zu stimulieren, zu entwickeln und erproben zu helfen. Solche Überlegungen führen, ausgehend von dem Sachverhalt des hochfrequenten Auftretens einzelner depressionsspezifischer Schwierigkeiten anscheinend folgerichtig zu Seligman und seiner Aussage, Kinder bräuchten Optimismus. Solch Optimismus wäre als Vertrauen auf die Wirksamkeit des eigenen Tuns, der persönlichen Tatkraft zu verstehen. Die Prävalenzraten für depressive Störungen von Schulkindern quantifizierten vielleicht diejenigen, die, aus welchen Gründen auch immer, die Versuche zur Bewältigung der täglichen Forderungen und daraus erwachsenden Bedrohungen reduziert oder aufgegeben haben. Durch solche Überlegungen ist nicht nachgewiesen, dass Schule ursächlich an solchen Entwicklungen beteiligt ist. Da aber die untersuchten Kinder und Jugendlichen ungeachtet der Altersunterschiede alle Schulgänger sind und Schule einen pädagogischen Auftrag hat, ist sie zweifellos auch an der als notwendig erachteten Vermittlung von optimistischer Handlungsbereitschaft zu beteiligen. Das ist einerseits eine zutiefst pädagogische und auch schulpädagogische Aufgabe, deren präventiver Wert zudem relativ hoch anzusetzen ist. Das ergibt für die Schule zwei gute Gründe, dem mit ihrer Praxis bewusst zu

folgen. Zur psychologisch eingehenderen Durchdringung der Bedingungen dafür, dass Schulkinder dann tatsächlich zum Handeln und zum Handlungserfolg kommen, ist mehr als nur die mitgeteilte Zielsetzung zu bedenken, aber viele Bedingungen dafür sind pädagogisch beeinflussbar.

Die hier zuvor geöffnete Perspektive schaut nicht nach den von depressiven Störungen betroffenen Kindern und Jugendlichen aus, sondern achtet auf die Frequenzen und Inhalte der alltäglich in ihrem Auftreten und Abziehen beobachtbaren depressiven Symptomen und Episoden. Wissenschaftlich favorisiert ist die Beachtung des Auftretens depressiver Störungen. Das in der Perspektive auf das alltägliche Betroffensein von Symptomen und Episoden erfassbare emotionale Geschehen ist aber zu wichtig für die Entwicklung individueller Handlungskompetenz, als dass es zu einem Nebenaspekt verkommen dürfte. Die absolute Mehrheit aller Kinder und Jugendlichen wäre insoweit in ein wissenschaftliches Ghetto gedrängt. Der an dieser Stelle gegebene Verweis auf den genannten Sachverhalt wird als ein wichtiger und entsprechend hervor zu hebender und an Konsequenzen reicher Teilbefund dieser Untersuchung gewertet.

Aus dem Datenaufkommen über die Anwendung von DTK und CDI wurden bis hier einige der Aspekte an den von den teilnehmenden Kindern und Jugendlichen erbetenen Selbstzuschreibungen von Itemvarianten. Die eingenommene Perspektive erfasste die über DTK und CDI eingeführten Symptome im Hinblick auf deren Auftretenshäufigkeit in der untersuchten Gesamtstichprobe. Wie deren Frequenzen zu anderen Messzeitpunkten gewesen wären oder sein würden, ist nicht bekannt. Betrachtet man die Messergebnisse aus einer zweiten Perspektive mit Blick auf die von jedem Kind berichtete Anzahl an Symptomen, erweist sich zunächst - wie zu erwarten - dass sowohl die über den DTK als auch jene über den CDI bekannt werdenden individuellen Mengen, als Gesamtpunktwerte bezeichnet, in einer für diese Verfahren typischen Weise variieren. Die gefundenen linksasymmetrischen Verteilungen weisen darauf hin, dass die von den zwei Verfahren erfassten Symptome nicht alle Kinder der Stichprobe in ähnlicher Weise betreffen. Die vom Nullpunkt der Verteilung her zunehmenden höheren Selbstzuschreibungen an Symptomen verringern sich zum Verteilungsende hin in der Anzahl der jeweils als betroffen bekannt werdenden Schulkinder kontinuierlich. So stellt sich - in Übereinstimmung mit Ergebnissen aus ähnlichen internationalen Studien - auch für die griechische Stichprobe der höhere Gesamtpunktwert im CDI und im DTK als ein innerhalb der untersuchten Gesamtheit von der Norm abweichendes Ereignis dar. Dieser in Untersuchungen mit CDI oder DTK immer wieder bestätigte Sachverhalt ergibt sich messtheoretisch nicht zwangsläufig, sondern nur unter bestimmten Bedingungen. Die sind als Untersuchungsbefund zu vermerken.

4. Befund: Mit der Ausnahme einer Minderheit von 3,5% der untersuchten Kinder und Jugendlichen, die zum Zeitpunkt der Untersuchung keine oder höchstens eine Aussage (0 oder 1 GPW) als sie selbst treffend angekreuzt hatten, berichteten die anderen Kinder unterschiedliche Mengen an depressiven Symptomen, wie sie vom DTK und CDI erfasst werden, als für sie zutreffend. Die Zuweisung eines jeden Schulkindes zu einer von drei statistischen Klassen an Gesamtpunktwerten in jedem der beiden benutzten Messverfahren bietet die Möglichkeit, die Relationen zwischen Teilnehmern oder Teilnehmerinnen mit eher niedrigen, mit mittleren und mit höheren Gesamtpunktwerten zu verdeutlichen. Die drei statistischen Klassen, hier als Gruppen bezeichnet, waren durch die nachfolgend angegebenen Mengen an Selbstzuweisungen von Symptomen definiert. Gruppe 1 umfasste niedrige Mengen, bis zu 8 GPW; Gruppe 2, mittlere Menge, von 9 bis zu 18 GPW; Gruppe 3, alle Mengen ab 19 GPW. Die nach den über den GPW

in Skala 1+2 des DTK gebildeten drei Gruppen zeigten ein prozentuales Verhältnis von 34,1 zu 47,6 zu 18,3. Das Verhältnis der Gruppierungen nach Ergebnissen aus dem CDI erwies sich als 41 zu 44,3 zu 14,7.

Aus statistischer Sicht dürfte dieser Befund als ein Indikator für die vermutete klinische Relevanz der für die Konstruktion der Verfahren benutzten Konstrukte verstanden werden. Zugleich bestätigt sich die Nutzbarkeit des von der Theorie her behaupteten Messprinzips. Die Mehrheit der Schulkinder sollte nur von einzelnen Symptomen betroffen sein, aber eine von der damit angenommenen Norm abweichende Minderheit würde über deutliche Symptommhäufungen berichten. Dieser Sachverhalt ist für die untersuchte Stichprobe eingetreten und entsprechend nachgewiesen.

In der untersuchten Stichprobe weicht eine Minderheit von Schulkindern durch die von ihnen gegebenen Selbstzuschreibungen an depressiven Symptomen von der ihnen gegenüber stehenden Mehrheit ab. In der aktuellen Literatur zum Auftreten von Depressionen bei Kindern und Jugendlichen wird übereinstimmend die Auffassung vertreten, die von Persönlichkeiten jeweils mitgeteilte Anzahl sie selbst betreffenden Symptome habe Informationswert. Sowohl von Kovacs (2004) als auch von Rossmann (2005) wird übereinstimmend angenommen, dieser Informationsgehalt vermittele Aufschlüsse über die Wahrscheinlichkeit mit der die jeweilige Persönlichkeit von einer depressiven Störung betroffen sein könnte. Die persönlich erreichte Gesamtpunktzahl aus der Durchführung des DTK oder des CDI wird als in Zusammenhang mit den der jeweiligen Messung zuzuerkennenden Kennwerten für deren Validität stehend verstanden. Der Kurvenlauf für den angenommenen Zusammenhang zwischen der Ausprägung des Gesamtpunktwertes und der Wahrscheinlichkeit des Bestehens einer depressiven Störung ist ebenso unbekannt wie die Antwort auf die Frage, ob ein verfahrensspezifischer Kurvenverlauf angenommen werden könnte, oder ob die Validität des Verfahrens mit den Anwendungen in unterschiedlichen Stichproben variere. Immerhin gelten einige Annahmen als hochwahrscheinlich zutreffend. Das sind die folgenden:

1. Zwischen der Zahlen selbst zugeschriebenen Symptomen bei der Durchführung von DTK oder CDI und den Wahrscheinlichkeiten für das Bestehen einer depressiven Störung bestehen Zusammenhänge
2. Jede über diese Relation auffindbare Wahrscheinlichkeit für das Bestehen einer depressiven Störung ist als Verhältnis von Wahrscheinlichkeit (p) und Gegenwahrscheinlichkeit (q) auszudrücken
3. Die Wahrscheinlichkeit steigt mit der Menge selbst zugeschriebener Symptome, also mit der Zahl der Gesamtpunktwerte
4. Das mit $p = 1$ und $q = 0$ ausgedrückte Verhältnis wird als nicht erreichbar angesehen.

Für die untersuchte griechische⁴ Stichprobe mit 2034 Schulkindern sind sowohl die individuellen Gesamtpunktwerte (GPW) aus DTK und CDI als auch deren Verteilung bekannt. Wenn allgemein zutrifft, dass die die GPW aus beiden Messverfahren immer auch spezifische Informationen über die Wahrscheinlichkeit des Auftretens depressiver Störungen tragen, sollte versucht werden, die für die untersuchten 2034 Schulkinder zu entschlüsseln. Von Entschlüsselung der Informationen muss gesprochen werden, weil jeder GPW nur die individuell gewählte Menge an Symptomen mitteilt. Welche Wahrscheinlichkeit des Auftretens einer depressiven Störung für die jeweils betrachtete Persönlichkeit besteht, würde erst bekannt werden, wenn zu erfahren wäre, über welchen Algorithmus GPW in valide individuelle oder gruppenspezifische Wahrscheinlichkeiten für das Bestehen depressiver Störungen umzurechnen wären.

Obschon der um 2005 aktuelle Stand der Forschung zu dieser Problematik widersprüchlich ist – er ermutigt einerseits, an der Lösung der erkannten Aufgaben zu arbeiten, die andererseits aber grundsätzlich nicht lösbar erscheinen – wird innerhalb der hier geführten Erörterungen der Versuch unternommen, empirisch fundierte Abschätzungen der Wahrscheinlichkeiten für das mutmaßliche Vorliegen depressiver Störungen im klinischen Sinne innerhalb der untersuchten Stichprobe abzugeben. Der damit für eine Aussage über das Auftreten depressiver Störungen in der Stichprobe erreichbare Status an empirisch begründeter Gewissheit, wird durch die Verwendung des Begriffes Indizien gekennzeichnet. Die liegen in der dazu denkbaren Skala unterhalb von Indikatoren.

Dass die verwendeten Instrumente für die Erstellung klinischer Diagnosen nicht geeignet sind, ergibt sich bereits aus dem Vergleich der einerseits - so nach DSM und ICD - zu fordernden unterschiedlichen Informationen und den andererseits durch DTK und CDI maximal erreichbaren Informationen. Die Verfahren sind defizitär gegenüber den für Diagnosen einzusetzenden empirischen Standards. Die für die griechische Stichprobe 2034 vorliegenden empirischen Daten könnten aber gestatten, eine begründete Einschätzung des Anteiles von Schulkindern zu erarbeiten, die möglicherweise als von depressiven Störungen im klinischen Sinne betroffen gedacht werden sollten.

Eine solche Einschätzung der Zahl von Kindern oder Jugendlichen, die mit großer Wahrscheinlichkeit eine depressionsspezifische Risikogruppe bilden, basierte auf der inzwischen aus internationalen Studien empirisch fundierten Annahme, höhere Gesamtpunktwerte in depressionsspezifischen Skalen seien Indikatoren für unterschiedliche Grade der Belastung durch depressive Symptome. An die Stelle der Bezeichnung von GPW als Indikatoren wird, an entsprechende Überlegungen anknüpfend, erneut vorgezogen, sie als Indizien zu betrachten. Indikatoren sollten einen hohen Grad an Gewissheit über die Validität der mit ihnen gegebenen Aussage vermitteln, so etwa durch eine 95%ige Wahrscheinlichkeit für deren Erklärungswert.

Die Suche nach empirisch begründbaren Abschätzungen für das Ausmaß des mutmaßlichen Auftretens depressiver Störungen in der untersuchten Stichprobe griechischer Schulkinder wird, den vorausgegangenen Ausführungen folgend, an der Feststellung angesetzt, dass die Menge der bei Durchführung von DTK oder CDI nach subjektiven Kriterien selbst zugeschriebenen depressiven Symptome spezifische Informationen über die Wahrscheinlichkeit der individuellen Betroffenheit durch depressive Störungen trägt. Die Menge selbst zugeschriebener Symptome steht in Relation zum Ausmaß des Auftretens depressiver Symptome. Der jeweils erreichte Gesamtpunktwert wird als Indiz für die Ausprägung eines individuellen Störungsbildes betrachtet. Diese Sichtweise wird so unter anderem auch von Canals, Domènech-Llaberia, Fernández-Ballart und Marti-Henneberg (2002) und von Autoren vertreten, die Kovacs (2004, S. 65-66) speziell für den CDI mit entsprechenden Literaturverweisen aufgeführt hat. Für den DTK begründet Rossmann (2005) eine mit Kovacs in den Grundannahmen übereinstimmende Position und führt dazu ebenfalls relevante Studien an. Im Verlaufe der hier geführten Erörterungen wurden sowohl in Kap. 9.1 als auch in 11.1 Studien genannt und gewürdigt, die den diskriminanten Leistungen des DTK und des CDI empirisch verlässlich nachgegangen sind. Der Gesamtpunktwert gilt im übrigen nicht nur als ein mit seiner Stärke variierendes Indiz für den aktuellen Stand des Auftretens depressiver Symptome, sondern auch als ein mit unterschiedlichem Gewicht auftretender Risikofaktor für die zukünftige Entwicklung einer depressiver Störung (siehe dazu u.a. Canals und Mit., 2002; DuBois und Mit., 1995; Mattison und Mit., 1990). Weiterhin ist für die Höhe der GPW ein starker Zusammenhang mit einer Reihe von klinisch relevanten Beeinträchtigungen auf unterschiedlichen psychosozialen Ebenen nachgewiesen worden, wie aus den dazu aufgeführten Literaturhinweisen in Kap. 11.1 zu entnehmen ist.

Indem 1. der DTK und der CDI sich mit unterschiedlichen Trefferquoten bei Versuchen zur Identifizierung depressiv gestörter Kinder als sensitive Screeningverfahren erwiesen haben, muss am Erklärungswert der Aussage über den Zusammenhang zwischen der Variation von Gesamtpunktwerten und dem Ausprägungsgrad der persönlichen Belastung durch depressive Symptome nicht grundsätzlich gezweifelt werden. Fraglich und umstritten bleibt, welche Trefferquoten allgemein und dann auch in der jeweils betrachteten Untersuchung unter Rückgriff auf diese in ihrer Tendenz durchaus gut bestätigten Beziehung zu erreichen sind.

In Wiederholung einer weiter oben getroffenen Feststellung bleibt fraglich, welcher Algorithmus bei Vorhersagen vom Gesamtpunkt看wert im DTK oder CDI auf die Wahrscheinlichkeit des Auftretens von Belastungen durch depressive Symptome zu verwenden ist. Alle Untersuchungen, die zur Stützung der eben getroffenen Aussagen heranzuziehen sind, wurden zudem außerhalb Griechenlands durchgeführt. Wenn dann jedoch 2. die Messqualitäten der in unserer griechischen Untersuchung mit 2034 Schulkindern eingesetzten griechischen Versionen der Messinstrumente durchgehend Werte erreichen, die den in Untersuchungen mit originalen Versionen der Verfahren gleichen – und dies bis hin in die Faktorenstrukturen – ist ein Analogieschluss gerechtfertigt. Solange nicht gegenteilige und zwingend zu beachtende neue Sachverhalte auftreten, darf den eingesetzten griechischen Versionen der Messinstrumente eine mit den Originalversionen übereinstimmende Qualität als Screeningverfahren zugesprochen werden.

Daraus folgert 3. die Berechtigung zu der nachfolgenden Aussage. Ausgegangen wird nun davon, dass die über den Einsatz der griechischen Versionen des DTK und des CDI gewonnenen Daten in gleicher Weise spezifische Informationen zum Auftreten von Belastungen durch depressive Symptome tragen wie die Originalversionen. Somit sind auch Hinweise zum mutmaßlichen Bestehen depressiver Störungen in ähnlicher Qualität wie bei den Ursprungsverfahren zu erwarten.

Die Erörterungen der mit der durchgeführten Untersuchung erarbeiteten Möglichkeiten zu Aussagen über das mutmaßliche Auftreten depressiver Störungen in der untersuchten griechischen Stichprobe führen zunächst zu der Feststellung, die mit den griechischen Messinstrumenten gewonnene Datenbasis sei der mit der Anwendung deutschsprachiger Verfahren erreichbaren Basis gleichwertig. Den griechischen Versionen der Messinstrumente wird in eben dem Maße vertraut, wie Vertrauen zu den deutschsprachigen Versionen gerechtfertigt ist. In identischer Weise bleibt dann für die griechischen Messdaten aber auch das Problem der aus den Gesamtpunktwerten vorzunehmenden Ableitung von validen Wahrscheinlichkeiten für das Auftreten von Belastungen durch depressive Symptome erhalten.

Zur Begründung der weiteren Arbeitsschritte bei der Bestimmung empirisch begründeter Ableitungen von Wahrscheinlichkeiten aus den individuell erreichten Gesamtpunktwerten in DTK und CDI wird eine Umformulierung der selbst eingeführten Zielsetzung vorgenommen. Sie dient der Senkung des Schwierigkeitsgrades der Aufgabe. Da die Bestimmung eines aus dem individuellen Gesamtpunkt看wert in DTK oder CDI abgeleiteten Wahrscheinlichkeit für das Auftreten unterschiedlicher Grade von Belastungen durch depressive Symptome im Sinne der Zuweisung eines individuellen Rangplatzes schwieriger erscheint als die Zuweisung von Individuen zu Klassen von Wahrscheinlichkeiten, wird dies als ein anzustrebendes neues Teilziel eingesetzt. Eine darauf bezogene erste Feststellung besagt allgemein, dass die gesuchte Wahrscheinlichkeit mit steigendem Gesamtpunkt看wert ansteigt. Obschon der dann darauf zu beziehende Kurvenverlauf des Anstiegs nicht bekannt ist, ergibt sich dennoch eine erste als mit hoher Wahrscheinlichkeit zutreffende Folgerung. Eine nach Verfahren getrennte dreistufige Klassenbildung der gemessenen Gesamtpunktwerte schafft, wenn sie nach aufsteigenden Punktwerten organisiert wird, eine Trennung der Mitglieder der Gesamtstichprobe in drei Gruppen mit unterschiedlichen mittleren Wahrscheinlichkeiten für das Bestehen von

unterschiedlichen Ausprägungen an Belastungen durch depressive Symptome. Zwar bliebe unbekannt, welche Klassen von Ausprägungsgraden damit zu unterscheiden waren, aber die Relationen wären begründet abzuschätzen. Die Gruppe mit den niedrigsten Durchschnittswerten an GPW dürfte als die mit dem geringsten Belastungsgrad gelten. Die Gruppe mit dem höheren durchschnittlichen Gesamtpunktwert sollte auch als die dann höher belastete gelten. Die dritte Gruppe, die mit dem höchsten durchschnittlichen Gesamtpunktwert wäre auch die mit der vermutlich höchsten Ausprägung an Belastungen durch depressive Symptome. Nonparametrische Verfahren könnten prüfen, ob die Verteilung der Mitglieder der Stichprobe auf die drei Klassen durch zufällige Variation der Mengen anzunehmen wäre. Unbekannt bliebe dennoch, ob der Ausprägungsgrad der Belastungen durch depressive Symptome zwischen den drei gebildeten statistischen Klassen wirklich signifikante Unterschiede aufzeigte und ob eine der Gruppierungen das Niveau klinisch beachtenswerter Belastungen erreichte. Immerhin vermittelte eine solche Bildung von drei wahrscheinlich unterschiedlich belasteten Gruppen von Schulkindern eine praxisrelevante Orientierung. Sie zeigte an, welche Schulkinder im Vergleich zu anderen eine schnellere und intensivere Betreuung bräuchten. Jede dieser drei Risikogruppen würde sowohl einen Anteil von dieser Risikogruppe zu Recht zugewiesener als auch einen Anteil irrtümlich zugeordneter Kinder und Jugendlicher umfassen. Die Fehlerhaftigkeit einer solchen Zuweisungspraxis sollte nach Möglichkeit durch validere Vorgehensweisen ersetzt werden. Die zu entwickeln wäre möglich, wenn sich in Erfahrung bringen ließe, welche Gesamtpunktwerte welchen Ausprägungsgraden an Belastungen durch depressive Symptome entsprächen. Dann ließen sich die angestrebten Klassenbildungen von Risikogruppen unter Kenntnis dieser Beziehungen vornehmen. Die Grenzen zwischen den nach Gesamtpunktwerten zu bildenden statistischen Klassen könnten zwischen Punktwerten gezogen werden, die einem definierbaren Unterschied im Ausprägungsgrad der valide identifizierten Belastung durch depressive Symptome entsprächen. Dann wären in den drei nach Gesamtpunktwerten gebildeten unterschiedlichen Klassen jeweils Schulkinder, die auch im Ausprägungsgrad ihrer eigenen Belastung unterschiedlichen statistischen Klassen zugehörten.

Auf das in diesem Zusammenhang zur Kenntnis zu nehmende Dilemma ist in den hier geführten Erörterungen zuvor schon mehrfach verwiesen worden. Eine valide Skala der Belastungen durch depressive Symptome ist nicht bekannt. Eben deshalb bleibt auch der Algorithmus für die Umrechnung von Gesamtpunktwerten in Ausprägungsgrade für Belastungen durch depressive Symptome unbekannt. Einziger Fixpunkt für die Suche nach Auswegen aus diesem Dilemma wäre die klinische Diagnose depressiver Störungen, falls die als fehlerfrei gelten dürfte, was sie nachweislich nicht ist.

Die skizzierte Problematik ist in der Fachliteratur zur Depression, auf welches Verfahren auch immer bezogen, die der Bestimmung von Trennwerten zur Unterscheidung von Gruppen mit unterschiedlichen Ausprägungsgraden psychischer Störungen. Je nach Zielsetzung sind jeweils die nicht betroffenen von den betroffenen Personen zu trennen, die unbelasteten von belasteten, die ohne Diagnose von denen mit klinisch relevanter Diagnose.

Versucht wird hier nun, das zuvor eingeführte Teilziel der empirisch begründeten Zuweisung der Schulkinder dieser Untersuchung zu depressionsspezifischen Belastungs- oder Risikogruppen über das Eingehen auf die Diskussion zur Praxis der Verwendung von Trennwerten oder Cut-offs für DTK und CDI Gesamtpunktwerte zu erreichen.

Zu einer an Wahrscheinlichkeiten orientierten Einschätzung der Verbreitung von nach Ausprägungsgraden klassifizierten Belastungen der 2034 Schulkinder durch depressive Symptome wurden auf Basis der vorliegenden Daten und unter Berücksichtigung der gefundenen

Messqualitäten der verwendeten Instrumente (DTK, CDI und ESDM) zwei, auch in vorausgegangenen Studien anderer Autoren verwendete Möglichkeiten als angesehen.

Die erste einsetzbare Vorgehensweise ist, die Zuweisung der Schulkinder zu Belastungs- oder Risikogruppen getrennt nach den Ergebnissen aus den jeweils eingesetzten Verfahren zu vollziehen. Auf eine in diesem Zusammenhang denkbare Einbeziehung der Ergebnisse aus dem ESDM wird zunächst verzichtet, weil die für das Vorhaben zu fordernde Spezifität der Ergebnisse aus diesem Verfahren, wie schon im Kapitel über die Skala ESDM ausführlich dargestellt, begründet anzuzweifeln ist. Die nach den Ergebnissen aus einzelnen Messverfahren zu realisierende Zuweisung der Schulkinder zu den unter Einsatz von definierten Trennwerten, Cutoffs, gebildeten Belastungs- oder Risikogruppen würde zur Kenntnis bringen, welche Gruppe von Kindern nach Hinweisen aus welchem Messverfahren mit welcher Wahrscheinlichkeit als von klinisch ausgeprägten depressiven Störungen betroffen angesehen werden sollte. Die Beschreibung dieser Belastungs- oder Risikogruppe kann in unterschiedlicher Weise erfolgen. Zum einen wären alle die Schulkinder aufzuführen, die einer solchen Gruppe zugewiesen werden mussten. Zum anderen wäre der Umfang dieser Gruppe als prozentualer Anteil an der Gesamtzahl der Angehörigen dieser Stichprobe aus Schulkindern darzustellen.

Eine zweite Möglichkeit, die als Strategie des Einsatzes einer „multi-instrument-batterie“ zu bezeichnen wäre, ergibt sich aus der Anwendung unterschiedlicher Messverfahren in einer Stichprobe. Das soll nach Ergebnisse aus einschlägigen klinischen Studien (Almqvist, 1999; Carey et al., 1987; Doerfler et al., 1988; Fundudis et al., 1991; Kazdin, Colbus und Rogers, 1986; Kresanov et al., 1998; Mattisson et al., 1990; Wendel et al., 1988) zu einer präziseren Abschätzung einer Zahl der mutmaßlich mit klinisch relevanten Ausprägungen depressiv gestörten Kinder verhelfen. Auf die Möglichkeit einer solchen Strukturierung der Analysearbeit war in den vorausgegangenen Erörterungen nur beiläufig eingegangen worden, obschon die Auswahl mehrerer zum Einsatz vorgesehener und dann auch benutzter Messinstrumente mit der Intention einer multiplen Ergebnisbildung aufgenommen worden war. Zwischenzeitlich vorgenommene Vergleiche der Daten aus den vier eingesetzten Verfahren ließen jedoch Zweifel an der Möglichkeit des Erreichens einer gegenüber den Ergebnissen aus einzelnen Verfahren gesteigerten Validität aufkommen, die nicht aufzuheben warfen. Lediglich die Kombination der Ergebnisse aus zwei Verfahren schien weiterhin bedenkenswert, nämlich die der Messungen aus CDI und DTK.

Bei der auf Daten von einzelnen Verfahren zu beziehenden Suche nach definierten und empirisch geprüften Wahrscheinlichkeiten für das Vorliegen depressiver Störungen bei Angehörigen der untersuchten Stichprobe wurden als Indikatoren für die kritische Menge der selbst zugeschriebenen einzelnen Symptome zwei unterschiedliche Bestimmungsweisen für Trennwerte, Cutoffs, genutzt. Das Kriterium der ersten Art ist die Bewährung von Trennwerten aus der Varianz der jeweiligen Gesamtpunktwerte in Untersuchungen zur differentiellen Validität des jeweiligen Messverfahrens. In Untersuchungen mit externen Stichproben erwiesen sich die von dort mitgeteilten Trennwerte, Cutoff-Werte, als prognostisch mit begrenzter Wahrscheinlichkeit und dazugehöriger Gegenwahrscheinlichkeit valide. Das zeigte sich, indem die Kinder der Gruppe oberhalb des definierten Trennwertes mit empirisch über die Trefferquoten bekannt gewordener Wahrscheinlichkeit zutreffend als belastet durch klinisch ausgeprägte depressive Störungen identifiziert werden konnten. Als Kriterium für das Bestehen einer Störung war das Vorliegen einer aktuell gültigen klinischen Diagnose eingesetzt worden. Mit dieser Vorgehensweise wird einerseits eine gegenüber der Zuweisung zu Risikogruppen allein nach der Höhe vorliegender Gesamtpunktwerte ohne Kenntnisnahme von empirisch qualifizierten Trennwerten hochwahrscheinlich eine Annäherung an valide Trennungen von Kindern mit und ohne klinischen Ausprägungen von depressiven Störungen erreicht. Andererseits muss dabei ein in seiner

Ausprägung in der Regel nicht erkennbares Risiko für Zuweisungsfehler zumindest aus zwei Gründen hingenommen werden. Zum einen entsteht dieses Risiko, weil der empirische Standard der herangezogenen Diagnose zumeist unbekannt bleibt. Zum anderen ist nicht in Erfahrung gebracht worden, welche Stichprobenvarianzen für solche Trennwerte bestehen. Solange aber keine verlässlicheren Verfahrensweisen zur Bestimmung von Belastungs- oder Risikogruppen bekannt und anwendbar sind, darf die Verwendung von Trennwerten, die sich in anderen Untersuchungen empirisch bewährten, als Mittel der Wahl gelten.

Für die Trennung der mutmaßlich in klinisch relevantem Ausmaß durch depressive Symptome belasteten griechischen Schulkinder der untersuchten Stichprobe von den wahrscheinlich in geringerem Maße betroffenen Kindern und Jugendlichen werden für die Gesamtpunktwerte aus dem DTK die von Pötz (2003) und für Werte aus der Anwendung des CDI die von Kovacs (2004) empfohlenen Grenzpunktwerte Verwendung finden. Diese Wahl war so zu vollziehen, weil zum Zeitpunkt der Entscheidung keine erfolgreichereren Trennwerte zur Auswahl standen.

Das Kriterium der zweiten Art, nach dem in der zweiten alternativen Vorgehensweise der jeweils benötigte Trennwert bestimmt wurde, ist verteilungsabhängig bestimmt. Zu beobachten ist, dass die Anzahl der Schulkinder pro Gesamtpunkt看wert vom ersten Wert in der Verteilung, dem $GPW=0$ an zunächst ansteigt und dann wieder zurückgeht. Der $GPW=0$ ist, wie die Frequenzen nachweisen, ein relativ seltenes Ereignis. Mit dem ihm nachfolgenden Werten entsteht eine Verteilung, die an einem höheren Gesamtpunkt看wert bis in die unmittelbare Nähe der Frequenz für $GPW=0$ absinkt oder sie einstellt. Der dann folgende GPW hat bereits eine Frequenz die unter der für $GPW=0$ liegt. Dieser GPW ist demnach ein seltenes Ereignis als der Punkt看wert am Anfang der Verteilung. Die für Screeningverfahren zum Auftreten depressiver Symptome zu beobachtenden Verteilungen haben alle die Eigenheit, bei den höheren Gesamtpunkt看werten in der Frequenz unter der vom Kurvenanfang zu fallen. Das zweite Kriterium folgt dieser Beobachtung. Als kritischer Wert und als Trennwert wird verteilungsabhängig der GPW bestimmt, für den die Frequenz erstmals unter jener für $GPW=0$ sinkt. Dieser Wert trennt den Teil der Verteilung ab, der dem Augenschein folgend außerhalb der Grenzen für eine Symmetrieforderung für die Verteilung liegt und damit innerhalb der gesamten Verteilung deutlich von den Parametern einer Normalverteilung abweicht. Wenn diese Bestimmung als nach dem Augenschein vorgenommen bezeichnet wird, steht diese Einordnung für die Tatsache, dass nicht, wie statistisch zu fordern wäre, eine Prüfung auf die Möglichkeit des Bestehens einer zufälligen Randposition des als Trennwert bestimmten Gesamtpunktwertes vorgenommen wird. Diese Verletzung statistischer Regeln wird als in einer klinischen Perspektive nützlich betrachtet. Unterstellt wird, dass die Betroffenheit von depressiven Symptomen in einer untersuchten Stichprobe zwar auf spezifische Verursachungen zurück geht, die selbst aber in ihrer Konfiguration anscheinend zufällig über den von den Angehörigen der Stichprobe belegten Bereich verteilt sind. Unterstellt wird, es sei normal, dass eine solche Variation depressiver Symptome, die sich in GPW ausdrückt, im alltäglichen Leben auf trete. Das Niveau dieser normalerweise auftretenden Verteilung variiert von Stichprobe zu Stichprobe. Jener Teil der Verteilung, der durch einen Trennwert abgeschnitten wurde, der nach dem Kriterium einer für ihn beobachteten Besetzung unterhalb der Frequenz für den $GPW=0$ bestimmt worden war, steht im Verdacht, nicht durch die zufällig verteilten Bedingungen determiniert worden zu sein, sondern durch solche, die Ausprägungen von Gesamtpunktwerten verursachen, die mit höherer Wahrscheinlichkeit depressive Störungen anzeigen als die GPW der hochfrequenten Mehrheit der Teilnehmer in dem Hauptteil der Verteilung. Da die Schulkinder erkannt werden sollen, für die diese Wahrscheinlichkeit besteht, wird der verteilungsabhängige Trennwert mit der gewählten Bestimmung vorsorglich am Rand der Mehrheit der verteilten Werte positioniert. Die klinisch bedeutsamen Einzelfälle sind in der Ausprägung der Indizien immer randständig und dies durch Annäherung an das kritische

Maximum von Indizien. Der vermutete Nutzen einer verteilungsabhängigen Bestimmung der Trennwerte muss sich, wie bei jeder anderen Bestimmungsweise, an dem Grad der so zu erreichenden Validität beweisen. In den vorausgegangenen Kapiteln der hier geführten Erörterungen ist dieser verteilungsabhängige Trennwert als „Frequenzunterschreitung von $GPW=0$ “ in der jeweils betrachteten Verteilung der Gesamtpunktwerte bezeichnet worden.

Die Verwendung dieses Kriteriums basiert, wie dargestellt, auf der Feststellung, dass ab dem definierten Trennwert der sich von GPW zu GPW fortsetzende Verlauf der Verteilung eine andere Gangart als zuvor aufnimmt. Innerhalb der untersuchten Verteilung findet sich ab dem definierten Trennwert eine beobachtbare kritische Abweichung vom zuvor wahrzunehmenden Verlauf der Verteilung der individuell erbrachten Gesamtpunktwerte aus DTK und CDI.

Die Ergebnisse aus diesen zwei Vorgehensweisen, die ausführlich an entsprechenden Stellen dieser Arbeit dargestellt und diskutiert wurden (Kap. 9 und 11) führen zu den nachstehenden Befunden:

5. Befund: Unter Anwendung eines von Pötz (2003) erarbeiteten und mitgeteilten Trennwertes für die DTK-Skala 1+2 ($GPW \leq 17$) ergab sich mit den Daten aus der dargestellten Untersuchung einer Stichprobe von 2034 griechischen Schulkindern für die Bildung einer depressionsspezifischen Risikogruppe ein prozentualer Anteil von 25,9% an der Stichprobengröße. Für die Mitglieder dieser Gruppe, deren GPW in Skala 1+2 des DTK 17 oder mehr Punkte erreichte, ist eine höhere Wahrscheinlichkeit für das Bestehen einer depressiven oder einer komorbiden depressiven Störung erkannt worden.
6. Befund: Unter Anwendung eines von Kovacs (2005) mitgeteilten Trennwertes für den CDI ($GPW \leq 19$) ergab sich ein Prozentsatz von 14,7% der beteiligten Kinder, deren GPW im CDI auf eine erhöhte Wahrscheinlichkeit für das Erleben einer depressiven Störung verweist.
7. Befund: Den aus der Untersuchung 2034GR bekannt werdenden Prozentsätze für die Anteile der wahrscheinlich von depressiven oder komorbiden Störungen betroffenen Schulkinder kommt nach den auf der Basis vorliegender empirischer Daten vorgenommenen Abschätzungen für die korrekte Zuordnung der Schulkinder zu Risikogruppen eine Wahrscheinlichkeit zwischen 75% und 80% zu. Dementsprechend ist von einer Fehlerquote von 20 bis 25% auszugehen.

Die vorher dargestellte Befunde zu den über die Anwendung des DTK und CDI in einer Stichprobe von 2034 griechischen Schulkindern gefundenen Prozentsätzen zum Anteil von mutmaßlich in einem klinischen Ausmaß psychisch belasteten Kindern geben Anlass zu der Nachfrage, ob inzwischen durchgeführte griechische Untersuchungen zu ähnlichen oder abweichenden Ergebnissen gekommen sind. Zugleich wären vorsorglich Erklärung für die schon in der Studie 2034GR stark von einander abweichende Ergebnisse in der Abschätzung der Anzahl der durch depressive Störungen betroffenen Kinder und Jugendlichen zu suchen und zu erörtern.

Relevant dafür könnten neben anderen die theoretischen Positionen von Reicher (1998) und Rothenberg und Woerner (2004) sein. Die letztgenannten Autoren vertreten die Meinung, dass „... interpretation of (such) cross-cultural comparisons of score levels requires careful methodological considerations, because (1) genuine differences in child psychopathology between countries may be differently reflected in questionnaire scores and clinical prevalence rates, (2) variable standards of reference...., and (3) even a perfect translation may still contain small variation in the wording of a questionnaire item or slightly different shades of meaning within the same semantic field (e.g with respect to severity or frequency)“ (Rothenberger und Woerner, 2004, S. II/2). Reicher andererseits nennt als Faktoren, die die Angaben zur Häufigkeit depressiver Störungen beeinflussen können, die folgenden 1. Merkmale der Stichprobe, 2. diagnostische Aspekte bzw.

die Anwendung verschiedener Methoden und 3. die unterschiedliche Definition eines Falles (Reicher, 1998, S. 52). Zu den Stichproben als Fehlerquelle äußert sich Reicher unter anderem wie folgt: „...Der Anteil depressiver Störungen kann hier beispielsweise durch Selektionsfehler und hohe Komorbiditätsraten überschätzt werden“ (s.o., S. 52).

Daten zu depressiven Störungen bei griechischen Kindern und Jugendlichen bieten mit dem Stand von Ende 2007 insgesamt 6 Studien, wenn auch nicht alle ausdrücklich auf das Erfassen von Depressionen gerichtet waren. Im einzelnen sind dies die Untersuchungen von Bibou-Nakou & Kioseoglou (2001a,b), Kleftaras und Didaskalou (2007), Madianos, Gefou-Madianou und Stefanis (1993) und Trilivas, Vassilaki und Chimienti (1998).

In der Untersuchung von Bibou-Nakou und Kioseoglou (2001a, b) wurden unter Einsatz des Fragebogens „Moods and Feelings Questionnaire“ (MFQ, Angold, Costello & Pickles, 1987) Informationen über das Auftreten von Depressionen in einer Stichprobe von 610 Kindern und Jugendlichen aus Thessaloniki im Alter von 9 bis 18 Jahren erfasst. Die Autoren fanden Hinweise auf das Bestehen einer Prävalenzrate von etwa 11,5%. Trilivas und Mitarbeiter (1996) fanden nach Anwendung des „Beck Depression Inventory“ (BDI, Beck, 1978) in einer Stichprobe von 800 Jugendlichen im Alter von 17 und 18 Jahren aus Kreta, dass 41,5% der untersuchten Jugendlichen von depressiven Symptomen „nicht oder kaum“ belastet seien, aber 25,6% der Teilnehmer „mild depressiv“, 11,4% „mild-to-moderate depressiv“, 16,4% „moderate-to-severe-depressiv“ und 5,1% „severe depressiv“ wären. Die unterschiedenen Belastungsgrade ergaben sich nach den für das eingesetzte Verfahren empfohlenen Trennwerten. Aus der Studie von Madianos und Mit. – übrigens die einzige griechische Depressionsstudie, die Kinder aus mehr als einem Wohnort untersucht hat – ergab sich unter Anwendung des Verfahrens CES-D in einer Stichprobe von 1325 Jugendlichen zwischen 12 und 17 Jahren ein Prozentsatz von 20,3% für wahrscheinlich depressiv belastete Teilnehmer und Teilnehmerinnen.

Aufgrund starker methodologischen Unterschiede (s.o. Reicher) können die Ergebnisse aus den Einzelstudien nicht hinreichend aussagekräftig miteinander verglichen werden.

Der nachstehende 8. Befund fasst den unter Kenntnisnahme der heranziehbaren weiteren griechischen Untersuchungen zum Auftreten depressiver Störungen entstehenden Erkenntnisstand zusammen.

8. Befund: Die aus der Untersuchung 2034GR bekannt werdenden Prozentsätze an Kindern und Jugendlichen, die als Teilgruppe ihrer Stichprobe mit hoher Wahrscheinlichkeit durch klinisch relevante Ausprägungen von depressiven Symptomen belastet gelten, zeigen allgemein Übereinstimmungen zu den Ergebnissen aus 6 weiteren griechischen Untersuchungen. Da die zu vergleichenden Studien erhebliche methodische Unterschiede aufweisen, ist vor allem zur Kenntnis zu nehmen, dass alle diese Untersuchungen relevante Hinweise auf das Bestehen depressiver Störungen mit einem prozentualen Anteil von mindestens etwa 5% und einem Maximum von etwas über 20% geben. Die mit dem Verfahren DTK in der Untersuchung 2034GR gewonnenen Prozentsätze liegen im oberen Bereich der zu vergleichenden Werte. verteilt über alle Bereiche heranziehbaren Ergebnisse. Die aus der Durchführung des CDI bekannt werdenden Prozentsätze sind im Mittelfeld der Vergleichsdaten platziert. Die unter Anwendung des verteilungsabhängigen Trennwertes in der Durchführung von CDI und DTK ermittelten Prozentsätze gehören zu den Einschätzungen, die den unteren Bereich der gesammelten Ergebnisse bilden. Dem Forschungsstand folgend, ist vom Nachweis des Auftreten depressiver Symptome, Episoden und klinisch relevanten Störungen unter griechischen Schulkindern mit an Sicherheit grenzender Wahrscheinlichkeit auszugehen. Die aus den verschiedenen Studien mitgeteilten unterschiedlichen Abschätzungen des Anteils betroffener Kinder und

Jugendlicher an der jeweiligen Gesamtheit einer untersuchten Stichprobe ist zum einen unter Hinweis auf die gegebenen Fehlermöglichkeiten zu relativieren. Zum anderen ist auch zu erwarten, dass die unterschiedlichen Prozentsätze mit nach Ausprägungsgraden unterschiedlich zusammengesetzten Teilgruppen korrelieren. Hier wird die Hypothese gesetzt, der Anteil von Kindern mit klinisch relevanter Ausprägung einer depressiven Störung nehme von Schätzwerten um 5% ausgehend mit steigenden Prozentsätzen kontinuierlich ab.

Der Vergleich der Schätzungen aus den verschiedenen methodisch sehr unterschiedlichen Untersuchungen kann gegenwärtig vor allem der Orientierung über die Forschungsaktivitäten dienen. Aussagen über die psychische Situation der untersuchten Kinder und Jugendlichen gewinnen ihren eigenen Erklärungswert mit höherer Wahrscheinlichkeit durch dessen Fundierung über die Herkunftsstudie. Der gegenwärtig in Griechenland bestehende Mangel an Studien zur klinischen Validität psychologischer Messverfahren behindert sowohl die angestrebte wissenschaftliche Entwicklung des Forschungsbereiches als auch die Entwicklung der mit emotionalen Störungen bei Kindern und Jugendlichen in Zusammenhang stehenden klinischen, soziokulturellen und pädagogischen Praxisbereiche in Griechenland.

Der nach dem 8. Befund zu betrachtende Erkenntnisstand erlaubt die Aussage, dass am Auftreten depressiver Episoden und Störungen unter griechischen Schulkindern keine begründbaren Zweifel bestehen. Unsicher bleibt, wie hoch der Anteil betroffener Kinder und Jugendlicher an allen Schülerinnen und Schülern ist. Die Wahrscheinlichkeit nach der dieser Anteil zwischen 5 und 20% liegt ist sehr hoch.

Die Ergebnisse aus der Untersuchung 2034GR stärken den Erklärungswert dieser Feststellung.

Der über den CDI gefundene Prozentsatz von 14,7% für den Anteil depressiv hoch belasteter Mädchen und Jungen an der Gesamtstichprobe steht im Allgemeinen auch in guter Übereinstimmung mit den Daten aus internationalen Studien. Die Mehrheit der Studien mit dem CDI im deutschsprachigen und amerikanischen Raum berichtet einen Prozentsatz von etwa 10% der untersuchten Kindern (u.a. Doerfler et al., 1988; DuBois, Felner, Bartels et al., 1995; Lobert, 1989b; Nevermann, 1992). Almquist und Mit. (1999) fanden in einer Stichprobe von 6017 achtjährigen Kindern aus Finnland einen Prozentsatz von 6,9%, wobei in unserer Studie der Prozentsatz der Teilstichprobe der achtjährigen mit 8,2% ermittelt wurde (N= 171). Frigerio, Pesenti et al. (2001) fanden unter 284 achtjährigen aus Norditalien einen Prozentsatz von 10,6. Chen, Rubin und Li (1995) berichten aus ihrer Studie mit 210 chinesischen Kindern im Alter von 10 Jahren einen Prozentsatz von 10% (Cut-off= 20 GPW im CDI). Bahls (2002) teilt als Befund aus seiner Studie mit dem CDI und ebenfalls nach Anwendung des empfohlenen Cutoffwertes von 19 GPW einen Prozentsatz von 20,3% mit (N= 463 Kinder zwischen 10 und 17 Jahren aus Brasil). In unserer Stichprobe 2034GR haben 16,5% der Teilnehmer ab 10 Jahren einen Gesamtpunktwert von 19 oder höher erreicht. Chan (1997) berichtet aus seiner Untersuchung an 612 Kinder (12-17 Jahre alt) aus Hong Kong über einen Prozentsatz von 20,24 (Cutoff = 20). In der griechischen Stichprobe 2034GR lag der Prozentsatz für die 12jährigen oder älterer Kinder bei 19,1.

Im Gegensatz zu der vergleichsweise hohen Übereinstimmung zwischen den CDI Ergebnissen aus unserer Studie 2034GR und Werten aus internationalen Untersuchungen weicht der über den DTK ermittelte Prozentsatz von manchen Vergleichswerten deutlich ab. Solch Unterschied tritt auch im VBergrleich zwischen CDI und DTK innerhalb der Studie 2034GR auf. Bereits zuvor war an mehreren Stellen in dem hier gegebenen Bericht erklärt worden, dass der DTK über das

Konstrukt des agitierten Verhaltens eine relativ hohe Affinität für komorbide Störungsbilder, insbesondere für ADHS, hat.

Der auf die drei Subskalen des DTK bezogene Vergleich zwischen den Daten aus der Anwendung des DTK in der griechischen Studie 2034GR einerseits und der Normstichprobe andererseits (Rossmann, 1993) führen zu dem nachstehenden Resultat. Die zu betrachtenden Verteilungen sind im Hinblick auf die Skala 1 des DTK und die Skala 2 sehr ähnlich. Für die Skalenaddition 1+2, aus deren Verteilung der Gesamtpunktwerte nach Empfehlung durch Pötz der Trennwert für die mit 75% korrekten Zuweisungen der Teilnehmer an die entsprechenden Risikogruppen zum höchsten mit dem DTK bislang erreichten Vorhersagewert von nachweislich mit klinischer Ausprägung depressiv belasteten Kindern führt, gilt auch, dass von einer beachtenswerten gewissen überregionalen Übereinstimmung ausgegangen werden darf. Unterschiede ergeben sich für die Skala 3 „Müdigkeit/ Autonome Reaktionen“, wo die griechischen Schulkinder deutlich mehr belastet als die deutsche Kinder zu sein scheinen.

Für die Einordnung der Bedeutung der DTK Ergebnisse sollte der nachstehenden Überlegung gefolgt werden. In dem auf die Untersuchung 2034GR beschränkten internen Vergleich der Befunde aus CDI und DTK ergibt sich (1.) ein signifikanter Unterschied (14,1% zu 25,9%).

Eine deutliche Übereinstimmung findet sich aber (2.) sowohl für den Vergleich zwischen den eigenen CDI Ergebnissen und denen aus internationalen Studien als auch für den Vergleich zwischen eigenen DTK Ergebnissen und der deutschsprachigen Normstichprobe des Verfahrens. Die Erklärung der intern auftretenden Dissonanz zwischen CDI und DTK Ergebnissen sollte demnach nicht in aktuellen Variationen der Messqualitäten der eingesetzten Verfahren gesucht werden. Wahrscheinlicher ist bei der gegebenen Datenlage, dass beide Verfahren unterschiedliche Objekte oder Objektkombinationen erfassen. Die Divergenz wird nämlich verständlich, wenn Zimmermann (in Rossmann, 2005) gefolgt wird. Der fand in einer Anwendung des DTK in klinischen Kleingruppen, dass zwei klinisch unterschiedliche Gruppen von Kindern höhere Gesamtpunktwerte im DTK erreichten. Zum einen waren dies Kinder, für die eine depressive Störung diagnostiziert worden war und des weiteren eine Gruppe von Kindern mit diagnostizierter ADHS. Demnach enthält der nach GPW aus dem DTK mitgeteilte Wert von 25,9% einen ersten Anteil von depressiv gestörten Kindern und einen zweiten Anteil von Kindern mit Verdacht auf ADHS.

Dieser Feststellung folgt der nachstehende 9. Befund.

9. Befund: Die über die Ergebnisse aus der Anwendung des DTK in der Untersuchung 2034GR bekannt werdende Einschätzung, 25,9% aller beteiligten Kinder seien wahrscheinlich psychisch gestört, ist nicht allein auf depressive Störungen zu beziehen. Der DTK ist sensitiv gegenüber dem Störungsbild der ADHS. Demnach ist eine mit hoher Wahrscheinlichkeit eher als andere zutreffende Interpretation des oben angegebenen Prozentsatzes die nachstehende Variante. Der Prozentsatz von 25,9% über die Anwendung des DTK als wahrscheinlich psychisch gestört, betrifft einerseits Kinder oder Jugendliche mit depressiven Störungen und andererseits auch solche, die von ADHS betroffen sind. Eine dritte komorbide Teilgruppe schließt wahrscheinlich die Kinder ein, die von beiden Störungen belastet werden.

Die bis hier für die Untersuchung 2034GR gebildeten Befunde übergehen die einige Hundert umfassenden Teilbefunde nicht. Jeder Teilbefund steht für die von ihm vertretene Aussage, die in dem ihm entsprechenden Zusammenhang die ihr zustehende Bedeutung hat. Die Befundbildung nähert sich dieser großen Gesamtmenge differenzierender Teilergebnisse aus der oberen hierarchischen Begriffsebene um Antworten für die dort abgeleitete Forschungsfrage nach Hinweisen auf möglicherweise in der untersuchten Stichprobe von griechischen Schülerinnen und

Schülern auftretenden depressiven Störungen mit klinischen Ausprägungen zu finden. Solche Hinweise wurden aufgefunden, quantifiziert und unter explizitem Hinweis auf die für deren Erklärungswert sprechenden Wahrscheinlichkeiten und die ihn negierenden Gegenwahrscheinlichkeiten in Befunden mitgeteilt. Diese Befunde sind sowohl wissenschaftlich als auch für die mit dem Auftreten von depressiven Störungen unter Schülerinnen und Schülern unmittelbar oder mittelbar befassten Praxisbereichen der griechischen Gesellschaft informativ. Das trifft auf beides, auf die generierten Befunde und die ihnen zu Grunde liegenden Teilergebnisse zu. Die Nutzbarkeit der mitzuteilenden Sachverhalte wird allerdings durch zwei mit ihnen verbundenen Besonderheiten erschwert.

Die eine ergibt sich aus dem mit vielen Aussagen verbundene Grad der Wahrscheinlichkeit, unter dem das zutrifft, was mitgeteilt wurde. Die den Aussagewert kennzeichnenden Wahrscheinlichkeiten und Gegenwahrscheinlichkeiten erschweren die Nutzung der weiterreichenden Aufschlüsse, weil sie sowohl bei der Aufnahme der Informationen als auch bei deren Nutzung spezifische Beachtung verlangen. Spezifisch ist jeweils die im gegebenen Zusammenhang vor Fehlern schützende Verwendung. So gilt beispielsweise, dass unter Anwendung von DTK oder CDI nach konservativen Einschätzungen die jeweils erfassten Kinder oder Jugendlichen nur zu etwa 75% entsprechend gebildeten Risikogruppen korrekt zugewiesen werden können. Somit wäre jeweils mit dem Auftreten von Fehlerquoten um 25% zu rechnen.

Die andere Besonderheit entsteht aus der Bindung empirisch aufgefunder Sachverhalte an die bei der Datenaufnahme und der Datenanalyse herrschenden Bedingungen. Neben anderen setzt die Art der Stichprobe solche Bedingungen. In der Studie 2034GR wurden griechische Schülerinnen und Schüler, nicht finnische oder spanische und auch keine in Arbeitsprozessen stehende griechische Jugendliche untersucht.

Für die Aufnahme und die Verwendung der aus dieser wie aus anderen Untersuchungen mitzuteilenden Arbeitsergebnisse ist über die eben genannten Schwierigkeiten hinaus mit weiteren Komplikationen zu rechnen, die sich individuell beispielsweise über die Wahrnehmung, die Urteilsbildung oder die Handlungsregulation ergeben. Die hier aus der Studie 2034GR mitzuteilenden Befunde richten im Interesse der zu Grunde liegenden Forschungsfrage die Aufmerksamkeit auf das mutmaßliche Bestehen depressiver Störungen und lenken möglicherweise von anderen Aspekten ab. So sind depressive Störungen psychologisch mit dem Auftreten von depressiven Episoden und dem einzelner depressiven Symptomen verbunden. Die Aussage, der Anteil der Teilgruppe depressiver Kinder und Jugendlichen an der Gesamtheit einer Stichprobe umfasse 25% zieht die Aufmerksamkeit auf sich und lässt doch gleichzeitig erfahren, dass ein Anteil von 75% nicht von der Störung betroffen sei. Die Frage, warum die einen von depressiven Störungen betroffen sind, steht der Frage gegenüber, warum die anderen es nicht seien. Die Wahrnehmung der komplementären Struktur der Befunde und Ergebnisse veranlasst unter anderem auch, nach dem Verbleib derjenigen Kinder oder Jugendlichen zu fragen, die, bezogen auf die hier zuvor in der Analysearbeit an den Daten aus DTK und CDI eingeführten Risikogruppen, fehlerhaft zugewiesen worden sind. Der Hinweis auf fehlerhafte Zuweisungen könnte einerseits als Bezugnahme auf die allgemein als zutreffend angesehene Aussage verstanden werden, Menschen seien permanent von der Möglichkeit bedroht, fehlerhaft zu urteilen und zu handeln, irren sei menschlich. Die fehlerhaften Zuweisungen an Risikogruppen nach GPW aus DTK oder CDI ist dennoch als ein spezielles psychologisches Problem zu sehen. Das ist zu veranschaulichen, indem drei Klassen von fehlerhaften Zuweisungen unterschieden werden. Alle Fehlzuweisungen basieren auf Selbstzuweisungen von kritischen Antwortmöglichkeiten aus dem jeweiligen Repertoire der Messverfahren. Dem für DTK und CDI zu denkenden Messprozess wird die Annahme unterlegt, wer von den Teilnehmern und Teilnehmerinnen von depressiven Störungen betroffen sei, werde das über seine Wahl unter der den Items beigegebenen Antwortmöglichkeiten innerhalb gewisser Toleranzgrenzen korrekt

mitteilen. Das führte zum Erreichen einer die Betroffenheit der Persönlichkeit zutreffend kennzeichnenden Zahl von Gesamtpunktwerten. Nun wird aber durch entsprechende Validitätsstudien bekannt, dass auch stärkere Abweichungen in beiden möglichen Richtungen auftreten. Die dabei zu beobachtenden Abweichungen werden hier nach drei Kategorien unterschieden. Die eine fehlerhafte Selbstzuweisung schaffen die befragten Jungen oder Mädchen, die in Abweichung von der diagnostisch ermittelten hohen individuellen Betroffenheit durch depressive Störungen, so antworten, dass sie nur eine geringe Zahl von GPW erreichen. Sie praktizieren Verneinungen. Eine zweite Kategorie fehlerhafter Selbstzuweisungen zeigt leichtere Abweichungen von der persönlichen Betroffenheit durch depressive Störungen. Die Befragten irrten bei der Antwort auf die eine oder andere Frage und sind nach der fehlerhaften Zuweisung unmittelbar in der numerischen Nähe des als Trennwert benutzten GPW zu finden. Die dritte fehlerhafte Selbstzuweisung betrifft Befragte, die sich zahlreiche depressive Symptome zuschreiben, ohne wirklich von ihnen betroffen zu sein. Sie stellen sich als belastet dar, eine klinische Diagnose ordnete sie als unbelastet ein.

Die drei Kategorien sind, um Übersichtlichkeit zu schaffen, auf die zuvor schon eingeführte höchste Risikogruppe für eine depressive Belastung bezogen. Die Qualität der Selbstzuweisungen ließe sich auch an individuellen Kriterien orientiert ordnen. Die klassische Testtheorie befasst sich auch mit den bei Testungen auftretenden fehlerhaften Zuweisungen und unterscheidet Fehler der einen und der anderen Art. Die hier dargestellten drei Kategorien ziehen jedoch, anders als die klassische Testtheorie, ausdrücklich die an Persönlichkeiten gebundene Existenz und Wirksamkeit unterschiedlicher und klinisch bedeutsamer Antworttendenzen in Betracht. Sie unterstellen die Existenz dreier unterschiedlicher persönlicher Tendenzen im Entscheid über die Selbstzuweisung von Testantworten, die Tendenz zum Negieren bestehender Symptome, die Tendenz zu einer ungefähr zutreffenden Selbstzuweisung und die Tendenz zur Überakzentuierung der eigentlich geringen eigenen Betroffenheit. Die Benennung dieser drei Kategorien beweist nicht die psychische Existenz der unterstellten psychischen Prozesse, wohl aber die Existenz von drei Personengruppen, die sich, bewiesen durch ihre Selbstzuschreibungen im DTK oder CDI, tendenziell so verhalten, wie dargestellt. Die Trennung der Negierenden praktizierenden Teilgruppe von der mittleren Teilgruppe kann ebenso wie die Trennung der überakzentuierenden Teilgruppe von der mittleren über die Bestimmung signifikanter Abweichungen von dem als Trennwert dienenden GPW an der Grenze zur höchsten Risikogruppe bestimmt werden.

Aus mehreren Gründen sind zumindest die Teilgruppe der Symptome negierenden und die der überakzentuierenden Persönlichkeiten als beratungs- oder behandlungsbedürftig einzuschätzen. Ihre Selbstzuschreibungen weisen nach, dass zumindest in mancher Situation der Interaktion von Person und Umwelt ihre Selbststeuerung nicht realitätskonform ist. Der Beratungs- oder Behandlungsbedarf ist durch die antizipierbare Gefahr einer Selbstschädigung in der aktuellen Lebenssituation oder der künftigen Entwicklung begründet. Da für die mittlere der drei dargestellten Kategorien das Vorliegen einer entsprechenden Selbstgefährdung nicht ausgeschlossen werden kann, wird die Teilgruppe der Kinder oder Jugendlichen, die durch diese Kategorie in ihren Selbstzuweisungen beschrieben wird, in die Folgerungen einbezogen, die mit dem nachstehenden Befund aus dem auftretenden Sachverhalt gezogen werden.

10. Befund: Durch DTK oder CDI wird eine Teilgruppe der jeweils untersuchten Kinder im Hinblick auf ihren mutmaßlichen diagnostischen Status fehlerhaft zugeordnet. Der Anteil dieser falsch zugeordneten Teilnehmer beträgt geschätzt etwa 25% der Stichprobengröße. Die Ursachen für die fehlerhafte Zuweisung können mit realitätsverzerrenden Komponenten der individuellen Selbststeuerung verbunden sein. Für diesen Fall besteht sehr wahrscheinlich psychologischer Beratungs- oder Behandlungsbedarf. Aus diesem Grund ist für die präventiv, kurativ oder pädagogisch arbeitende Praxis im Umfeld dieser

Kinder oder Jugendlichen auch Anlass gegeben, nach diesen im Allgemeinen unbekanntem Kindern oder Jugendlichen zu suchen. Eine erste damit zu verbindende wissenschaftliche Aufgabe wäre, Suchstrategien zu finden, die das Resultat solcher Suche optimieren helfen. Eine weitere Aufgabe sollte die empirisch verlässliche Beschreibung der psychischen Prozesse sein, die zur Beantwortung der Fragen im DTK und im CDI führen.

11. Befund: Aus der weiter oben unter Bezugnahme auf die Durchführung des DTK und des CDI bezogenen Schätzung des Anteiles mutmaßlich mit klinischer Ausprägung depressiv belasteter Kinder und Jugendlicher sowie unter Einbeziehung komorbider Störungen ist zusammen mit dem unter der 10. Aussage mitgeteilten Sachverhalt davon auszugehen, dass für etwa 1/3 der Kinder und Jugendlichen in der untersuchten Stichprobe – einschließlich ihres familiären und schulischen Umfeldes – eine psychologische Beratung oder Behandlung anzuraten wäre.

Von einer nachgehenden psychologischen Beratung wird angenommen, dass mit ihr sowohl im Interesse der Betroffenen als auch im wohlverstandenen Interesse des Sozialverbandes, dem die Kinder oder Jugendlichen angehören, fachkundig nachgeforscht wird, welche Unterstützung für die jeweils betrachtete Persönlichkeit notwendig wird und wie die zu geben sei.

Der Terminus Behandlung ist inhaltlich identisch mit dem Begriff der Psychotherapie. Wie groß der Behandlungsbedarf sein würde, wäre über die Beratung abzuklären.

Der mitgeteilte 11. Befund setzt, um Anlass zum Bedenken sozialpolitischer Maßnahmen geben zu können, das Bestehen einer gesellschaftlichen Struktur voraus, in der die Sorge um die Gesundheit der nachwachsenden Generation als unabweisbare politische Aufgabe anerkannt ist. Das ist nicht in allen Staaten gegeben.

Die bis an diese Stelle der Erörterung mitgeteilten Befunde ergaben sich unter Nutzung verfahrensspezifischer Empfehlungen zur Quantifizierung von Trennwerten auf der Erfahrungsgrundlage entsprechender empirischen Untersuchungen (Pötz, 2003; Kovacs, 2004). Die empfohlenen Trennwerte hatten sich dort intern als die Trennwerte mit der höchsten erreichbaren diskriminanten Leistung erwiesen. Insofern stehen die hier nach der Anwendung jener Trennwerte mitgeteilten Aussagen unter dem Vorbehalt, dass diese Werte zumindest relative intergrupale Stabilität hätten. Auf die bei der Analyse der Daten aus DTK und CDI unter Anwendung eines dort eingeführten verteilungsspezifischen Trennwert nach dem Kriterium der „Frequenzunterschreitung von $GPW=0$ “ erarbeiteten Ergebnisse (DTK-Skala1+2=6,7%, CDI=4,2%) wird an dieser Stelle nicht gesondert eingegangen, weil sie bereits in den Befund eingearbeitet wurden, der die Gesamtheit der für den Anteil der Teilgruppe mit den höchsten Risiko für die Betroffenheit durch depressiven Störungen erarbeiteten Ergebnisse aus DTK und CDI repräsentiert.

Die mitgeteilten Befunde 1 bis 11 nennen die Hinweise, die zur Beantwortung der dieser Studie zu Grunde liegenden Forschungsfrage in der untersuchten Stichprobe mit DTK oder CDI gefunden wurden. Diese Hinweise sprechen für das Auftreten von depressiven und komorbiden Störungen mit unterschiedlichen Schweregraden. Die Anteile der mutmaßlich betroffenen Kinder oder Jugendlichen variieren mit dem Ausprägungsgrad der selbstberichteten Betroffenheit von depressiven Symptomen zwischen Prozentsätzen um 5% bis etwa 25%. Dabei umfassen die Prozentsätze über 15% sehr wahrscheinlich einen deutlichen Anteil komorbider Störungen.

13.4 Hinweise zum Auftreten depressiver Symptome und depressiver Störungen in der untersuchten Stichprobe: Hinweise über Datenkombinationen aus zwei Messverfahren

Diese jeweils auf Ergebnissen aus einem von zwei Testverfahren beruhenden Aufschlüsse, werden nachfolgend um Befunde erweitert, die jeweils auf ein Ergebnis zurückgehen, das aus den Daten zweier Messinstrumente gebildet worden ist. Diese zwei Messinstrumente sind die Verfahren DTK und CDI. Dieser Art der Befundbildung unterliegt die Annahme, Messfehler des einen Verfahren würden durch das jeweils andere Verfahren gemildert oder ausgeglichen werden, was dann zum Erreichen einer höheren Messgenauigkeit führen sollte.

Die Möglichkeit der Bildung eines Befundes der Daten über die an der Untersuchung teilnehmenden Schulkinder aus Messungen in zwei messtheoretisch unterschiedlichen Perspektiven sammelt, ergibt sich aus der Tatsache, dass in der durchgeführten Studie zwei verschiedene depressionsspezifische Verfahren verwendet worden sind. Zwar wurden insgesamt vier Messverfahren eingesetzt, aber zwei von denen erwiesen sich aus messtheoretischen Gründen für eine gemeinsame Befundbildung als nicht geeignet.

Die Einführung zweier in der griechischen Sprache neuer depressionsspezifischen Messverfahren in ein und derselben Stichprobe, dies in Anlehnung an den in vielfältiger Weise nützlichen Multitrait-Multimethod-Ansatz, eröffnete einerseits die Möglichkeit zur wechselseitigen Kontrolle der Messqualitäten dieser Instrumente und gab dann andererseits auch Gelegenheit zum Vergleich der Ergebnisse aus beiden Verfahren. Zusätzlich war letzteres um eine transkulturelle Kontrolle zu erweitern, da auch Reicher und Rossmann (1991), die Daten aus der Anwendung von DTK und CDI in einer deutschsprachigen Stichprobe von Kindern und Jugendlichen analysiert haben. Darüber hinaus können nun auch die mit jedem einzelnen Verfahren aufgenommenen Selbsteinschätzungen der Schulkinder so miteinander kombiniert werden, dass sich ein weiteres und in mancher Hinsicht auch strengeres Kriterium zur Abschätzung der Anzahl mutmaßlich depressiv gestörter Kinder gewinnen lässt.

Eine Kombination von Daten aus zwei Verfahren wird messtheoretisch nützlich, wenn sich dadurch die Messqualität erhöhen lässt. Mit hoher Wahrscheinlichkeit sind die Messungen eines jeden der beiden Verfahren mit der an sich ungewollten Aufnahme von fehlerhaften Angaben verbunden. Eine aussagekräftig zu begründende Annahme ist, dass durch die Kombination der Daten aus beiden Verfahren die Schwächen der einen Messung mit Stärken der anderen reduziert oder gar aufgehoben werden könnten. Aber auch schon das gegenüber der Zahl in einem Verfahren deutliche Ansteigen der insgesamt benutzten Menge an Items aus beiden Verfahren führt nachweisbar zu Verbesserungen der Messqualität. Die daraus ableitbare Annahme besagt, dass diejenigen Kinder, die sowohl im DTK als auch im CDI höhere Gesamtpunktwerte erreichen, mit höherer Wahrscheinlichkeit von depressiven Störungen betroffen sein sollten als Kinder mit anderen denkbaren Kombinationen von Punktwerten. Die aber würden dann wiederum in gegeneinander abzuwägende unterschiedliche Positionen in einer Rangreihe der Wahrscheinlichkeiten für das Bestehen klinisch ausgeprägter depressiver Störungen aufzunehmen sein. Solche Überlegungen gelten hier den nach der Höhe der Gesamtpunktwerte gebildeten statistischen Klassen. Gegenüber der Addition von Punktwerten aus zwei unterschiedlichen Verfahren erbrächte eine an Werteklassen orientierte Kombination einige Vorteile, so etwa den Ausgleich von ansonsten sich möglicherweise kumulierenden Ungenauigkeiten der Selbstzuschreibung von Items durch die Schulkinder.

Die Annahme des Gewinns von Messqualitäten durch Kombination von Ergebnissen aus unterschiedlichen Verfahren bestätigen empirische Studien mit entsprechenden Ergebnissen. Die Kombination von Testverfahren zu einer so genannten „multi-instrument battery“ (Kazdin, Colbus und Rodgers, 1986) wird von vielen Autoren als die verlässlichste Vorgehensweise zur

Identifizierung depressiver Kinder durch Screeningsverfahren angesehen (Almqvist, 1999; Carey et al., 1987; Doerfler et al., 1988; Fundudis et al., 1991; Kazdin, Colbus und Rogers, 1986; Kresanov et al., 1998; Mattisson et al., 1990; Wendel et al., 1988). Der 12. Befund fasst die Ergebnisse aus der Kombination von Daten des DTK und des CDI zusammen.

12. Befund: Die Verwendung zweier spezifisch auf depressive Symptome gerichteter Messverfahren gestattet die Beantwortung der Frage, welche an der Untersuchung teilnehmenden Personen von beiden Instrumenten als mutmaßlich unbelastet oder belastet angeführt werden. Allerdings kann bei der Verwendung zweier unterschiedlicher Messverfahren auch jede teilnehmende Person in einem der Verfahren andere Gesamtpunktwerte als in dem zweiten erreichen. Bei daran zu orientierenden Zuweisungen zu Risikogruppen könnten solche Personen möglicherweise zugleich zwei inkompatiblen Gruppen zugerechnet werden. In der Untersuchung mit 2034 griechischen Schulkindern sind solche widersprüchlichen Zuweisungen an jeweils zwei inkompatible Risikogruppen aufgetreten. Sie ergaben sich über die beiden eingesetzten Verfahren, DTK und CDI, unter Anwendung der von Pötz für den DTK mit 17 GPW und von Kovacs für den CDI mit 19 GPW empfohlenen Trennwerte, in weniger als 1/5 aller möglichen Fälle. Die in Absetzung davon mit DTK und CDI übereinstimmend vorgenommenen Zuordnungen zur höchsten Risikogruppe führten zu den durch die nachstehende Tabelle angeführten prozentualen Anteilen der genannten Teilgruppen an der Teilnehmerzahl der gesamten Stichprobe.

Tabelle 13.1: Größe der nach GPW in DTK und CDI durch homogene und inhomogene Zuweisungen zu Risikogruppen gebildeten Teilgruppen

Risikogruppe	n	%
Niedrige oder mittlere Belastung nach DTK und CDI	1414	71,1
Hohe Belastung nach DTK oder nach CDI	340	17,1
Hohe Belastung nach DTK und nach CDI	236	11,9
	1990	

Zeichenerklärung: n = Anzahl, % = Prozentsatz zum Grundwert der untersuchten Stichprobe

Mit den dargestellten Werten erweisen sich etwa 71,1% der befragten Kinder, soweit ihren Selbsteinschätzungen zu vertrauen wäre, durch die Übereinstimmung der Ergebnisse aus den eingesetzten Verfahren DTK und CDI, als mutmaßlich nicht oder nur in einem mittleren Ausmaß und in diesem Sinne dann durch alltäglich anzutreffenden Episoden belastet. Dieser Feststellung kommt handlungsleitende Bedeutung zu. Die Schulkinder der untersuchten Stichprobe wären unter Vertrauen auf die mitgeteilten Ergebnisse in zwei Gruppen zu trennen. Eine erste Gruppe, in der sich 71,1% von ihnen befänden, ließe sich im Hinblick auf das Risiko der Entwicklung klinischer Ausprägungsgrade durch allgemein präventive Maßnahmen angemessen versorgen. Aber für eine zweite Gruppe, einer Risikogruppe mit 28,9% der Schulkinder, bestünde der Verdacht, sie seien durch das bei ihnen zu vermutende stärkere Auftreten depressiver oder komorbider Symptome auch stärker durch Folgerisiken gefährdet. Das sollte dazu veranlassen, spezielle Interventionen zu entwickeln und durchzuführen, die geeignet wären, die möglichst valide zu identifizierenden Belastungen dieser Kinder zu überwinden. Das entspräche der Intention, Schaden von diesen Schulkindern und der Gemeinschaft in der sie

aufwachsen abzuwenden. Die zweite der beiden angeführten Gruppen, die Risikogruppe, umschließt dann aber auch noch, den übereinstimmenden Ergebnisse von DTK und CDI zufolge, eine Teilgruppe ein, der 11,9% aller untersuchten Schulkinder zugehören. Für diese Kinder und Jugendlichen wird sich innerhalb der für die 2. Risikogruppe durchzuführenden Intervention hochwahrscheinlich die Notwendigkeit einer verstärkten Betreuung ergeben. In Erwartung dieser Anforderung sollte die für die Risikogruppe von 28,1% und die in ihr eingeschlossene Teilgruppe von 11,9% aller untersuchten Kinder durchzuführende Intervention, im Hinblick auf die sehr wahrscheinlich auftretenden starken Unterschieden im Ausprägungsgrad der individuellen Belastungen durch Störungen, bereits mehrstufig geplant sein, um den bei der Durchführung erkennbaren individuellen Erfordernissen nachkommen zu können.

Für 17,1 % aller untersuchten Kinder sind die von ihnen in den beiden Verfahren, DTK und CDI abgegebenen Selbsteinschätzungen insoweit unterschiedlich oder auch widersprüchlich, als aus den von ihnen jeweils erreichten Gesamtpunktwerten voneinander abweichende Zuweisungen an die eingeführten Risikogruppen hervorgingen. Diese Kinder werden von einem Verfahren als mutmaßlich belastet vom anderen als wahrscheinlich nur in einem geringeren Ausmaß betroffen eingestuft. Die durch die beiden eingesetzten Verfahren signifikant unterschiedlichen Zuweisungen der Schulkinder zu den eingeführten Risikogruppen verweisen auf bestehende Unterschiede zwischen DTK und CDI. Die an der Zuweisung zu den definierten Risikogruppen bestimmte Quote übereinstimmender Messdaten betrug für die durchgeführte Untersuchung sowohl für die Gruppe wahrscheinlich gering belasteter als auch für die Gruppe mutmaßlich stärker belasteter Schulkinder in beiden Fällen zumindest 80%, nämlich für Risikogruppe 1 = 83,3% und für Risikogruppe 2 = 80,8%. Diese Quoten sind sensibel gegenüber Änderungen an den Trennwerten. Die nachfolgende Tabelle zeigt zur Verdeutlichung der Verteilung aller nach DTK und CDI unterschiedlich vorzunehmenden Zuweisungen zu Risikogruppen eine Kreuztabelle zu den einerseits durchaus weit überlappenden im Übrigen dann aber hochsignifikant unterschiedlichen Messperspektiven der beiden Verfahren. Was damit an Unterschiedlichkeit den Verfahren zugeschrieben wird, könnte sich in Nachforschungen auch als intrapersonal determinierte Instabilität der Selbsteinschätzung einer Teilgruppe der Schulkinder in dieser Stichprobe erweisen.

Tabelle 13.2: Kreuztabelle der über DTK und CDI ermittelten Risikogruppen von Schulkindern mit und ohne mutmaßlich bestehender depressiver Störungen

<u>Zuweisungen an Risikogruppen durch Verfahren</u>			
CDI	DTK		Diff
	nB Gruppe 1	SB Gruppe 2	Summe
nB Gruppe 1	1414	284	1698
SB Gruppe 2	56	236	292
Summen	1470	520	1990

Zeichenerklärung: nB = mutmaßlich niedrig oder mittel belastete Gruppe; Sb = mutmaßlich stärker belastete Gruppe, Diff = Differenz DTK nB und DTKSB

Die Tabelle bietet Informationen über die duale Klassifizierung der untersuchten Schulkinder nach Risikogruppen und dies aus zwei Perspektiven, aus der des DTK und der des CDI. Bezogen auf die Zuweisungen durch den CDI werden aus dessen Gruppe niedrig oder mittel belasteter Kinder, der Risikogruppe 1, durch den DTK 16,3% nicht anerkannt und in Gruppe 2 verwiesen. Aus der CDI Gruppe 2 sollten nach den Messungen des DTK 19,7% in die Gruppe 1 kommen. Bezogen auf die Gesamtzahl der in beiden Messungen vertretenen Schulkinder mit $n=1990$ geben DTK und CDI für 17,1% von ihnen unterschiedliche Zuweisungen an. Wird das Bezugssystem so geändert, dass die Messungen des DTK als Grundwert erscheinen, finden sich die nachfolgend angeführten Relationen. Die nach DTK nur als alltäglich von depressiven Symptomen belastet einzustufen waren, erkannte der CDI mit einem Prozentsatz in Höhe von 96,2% ebenfalls als wenig belastet. Der Dissens zwischen beiden Verfahren ist im Hinblick auf die Risikogruppe 1 relativ gering. Im Gegensatz dazu wurden nur 45,4% aller nach DTK der Gruppe 2 zuzuweisenden Kinder auch vom CDI dieser Gruppe hoher Belastung zugeordnet. Die anderen Kinder, immerhin 54,6% der nach DTK in Gruppe 2 befindlichen, stuft der CDI als niedriger belastet ein. Das bedeutet, dass die CDI-Urteile für die nach DTK hochbelasteten Kinder nur in etwa 5 von 10 Fällen übereinstimmen. Die Übereinstimmung zwischen den beiden Verfahren ist für die als wahrscheinlich nur gering oder mittel durch depressive Symptome belasteten Kinder relativ hoch. In etwas mehr als 8 von 10 Fällen sind die Zuordnungen gleich.

Wird das Erkennen und Benennen von stärker belasteten Kindern als gemeinsame Aufgabe gesetzt, weichen die Befunde der zwei Verfahren in hohem Maße von einander ab. Wenn die CDI-Ergebnisse zutreffend genannt werden könnten, erreichte der DTK eine Erfolgsquote von etwa 81%. Sollten aber die DTK-Urteile realitätsnäher sein, erscheint die Leistung des CDI als niedrig, da mit ihm weniger als 5 von 10 Zuordnungen zutreffend wären.

Der dargestellte Unterschied zwischen CDI- und DTK-Messungen ist zunächst als Divergenz zwischen den von zwei unterschiedlichen Verfahren genannten Mengen an Selbstzuschreibungen von depressionsspezifischen Symptomen durch die beteiligten Schulkinder zu verstehen. Problemgehalt kommt den Messunterschieden zu, wenn die Daten helfen sollen Merkmale zu erkennen, über die eine valide Zuweisung der Kinder an zwei unterschiedliche Gruppen mit von einander zu trennenden Merkmalsträgern verhelfen soll. Die für die Problemlösung nutzbare Antwort auf die Frage, welches Verfahren die kritischen Merkmale besser erkenne, ist zum gegebenen Zeitpunkt und dann noch empirisch verlässlich nicht zu geben. Alle Entscheidungen darüber, welchen Angaben eher zu vertrauen sei, beruhen jeweils auf Annahmen, die zutreffen oder nicht zutreffen könnten. Eine solche Annahme wäre, dass den Messungen am ehesten zu vertrauen wäre, in denen die Verfahren relativ gut übereinstimmen. Wird dem gefolgt, ist die nachstehende Feststellung begründet:

Die Angaben der Verfahren zum Umfang einer Teilgruppe von wenig oder nur im mittleren Bereich durch depressive Symptome belasteter griechischer Schulkinder ist zu vertrauen. Sehr wahrscheinlich sind 71,1% aller beteiligten Kinder in keinem höheren Grade betroffen.

Danach gilt dann aber auch der Verdacht, dass die verbleibenden 28,9% in höherem und auch klinischem Ausmaß durch depressive Symptome belastet sein könnten. Wie schon an anderer Stelle ausgeführt, reichte dieses Ergebnis zu, spezifische Interventionen zur Reduktion des für diese Schulkinder der Stichprobe aktuell bestehenden Risikos der Belastung durch eine psychische Störung zielgerichtet ansetzen und aufbauen zu können.

Wenn die über DTK und CDI abgegebenen Angaben der nur wenig oder im mittleren Ausmaß belasteten Kinder (Teilgruppe 1) mit den Angaben der stärker betroffenen Kinder (Teilgruppe 2) verglichen werden sollen, ist zweckdienlich, so nahe am bestehenden Sachverhalt zu bleiben, wie nur möglich. Die dann zu treffende Feststellung ist ohne jede Zuhilfenahme von Interpretationen – immer wieder – die nämliche: Die Kinder der Teilgruppe 1 nennen mehr depressive Symptome

als die der Teilgruppe 2. Wenn DTK und CDI zu Übereinstimmungen hinsichtlich der Kinder aus Teilgruppe 1 aber nicht auf die Teilgruppe 2 kommen, sollte das im Zusammenhang mit speziellen Verursachungen stehend gesehen werden. Solche Verursachungen ließen sich zum einen in den unterschiedlichen Messinstrumenten vermuten. CDI und DTK folgen durch ihre inhaltlich unterschiedlichen Konstruktionen auch verschiedenartigen Messintentionen. Beide sind auf die Aufnahme von Informationen über das Auftreten von depressiven Symptomen ausgerichtet worden. Aber der DTK bezieht sich auf ein Störungsbild, in dem, nach DSM-III, auch ausagierendes Verhalten eingezogen wird. Damit öffnet sich der DTK messtheoretisch auch für die Aufnahme von Informationen über das Auftreten von ADHS. In der Skalenkombination DTK 1+2 finden sich für Kinder mit ausgeprägter ADHS, die im Zusammenhang damit auch depressive Symptomatik entwickelt haben, Messbedingungen, die sie beim Erwerb höherer Gesamtpunktwerte gegenüber Kindern mit Diagnosen über das Bestehen einer depressiven Störung nach DSM-IV begünstigen. Der Sachverhalt ist durch einfaches Auszählen der für depressive Störungen oder ADHS jeweils spezifischen Fragen in Skala 1+2 nachzuweisen. Solange andere Determinanten für die Unterschiede in den Messergebnissen nicht nachgewiesen sind, liegt nahe, diesen konstruktionsbedingten Unterschied zwischen DTK und CDI als Hauptquelle für nachzuweisende Messunterschiede anzusehen.

Von speziellem Interesse wären in diesem Zusammenhang aber auch Untersuchungen, die klären könnten, ob das in CDI und DTK unterschiedliche Antwortschema nachweisbaren Einfluss auf das Entstehen unterschiedlicher Messergebnisse hat.

Das jeder Erklärung der unterschiedlichen Messungen von CDI und DTK immer wieder entgegen stehende Grundproblem ist das Fehlen eindeutiger Nachweise für die Validität der Messergebnisse aus den Verfahren. Indem diese Feststellung ein Defizit benennt, erweckt sie auch den Anschein, als wäre die Überwindung dieses Defizits nur versäumt worden. Möglicherweise ist der Validitätsnachweis aber auch ein Problem, weil die dazu heran zu ziehenden psychiatrischen Diagnosen selbst nicht eindeutig zu validieren sind, denn anders als beispielsweise bei einer Knochenfraktur, sind aktuell keine validen Außenkriterien verfügbar.

Die Ergebnisse aus der Anwendung des DTK und des CDI in der untersuchten Stichprobe griechischer Schulkinder sollten zusammengefasst werden, weil damit, der einschlägigen Literatur folgende, ein gegenüber Einzelergebnissen höherer Grad an Gewissheit über das mutmaßliche Auftreten depressiver Störungen erreicht werden kann. Ein solches Ergebnis ist, trotz zahlreicher Anlässe zum Überdenken der diesem Versuch unterliegenden Bedingungen, zum Abschluss gebracht worden. Der Versuch, eine, gegenüber den von jeweils einzelnen Messinstrumenten hervorgebrachten Ergebnisse, durch Zusammenfassung der Daten aus unterschiedlichen Messinstrumenten eine höhere Eindeutigkeit der Befunde zu schaffen, führte zu dem angestrebten Ergebnis, aber auch zu einem zwangsläufig hinzunehmenden Verlust an differentiellen Aussagen. Nach CDI und DTK sind sehr wahrscheinlich, wie in Befund 11 dargestellt, 71,1% der untersuchten Schulkinder durchaus mit dem Auftreten einzelner depressiver Symptome vertraut, durch sie aber wenig oder nur in einem klinisch als eher unbedenklich erscheinenden Ausprägungsgrad belastet. Die nicht zu dieser Mehrheit der Stichprobe untersuchter Schulkinder gehörende 28,9%ige Minderheit ist sehr wahrscheinlich stärker von depressiven Symptomen betroffen und eine Teilgruppe innerhalb dieser Minderheit dürfte depressive Störungen oder eine ADHS mit klinisch relevanten Ausprägungsgraden ausgebildet haben. Der Anteil dieser Kinder an der Gesamtstichprobe wird durch den CDI auf 14,7% geschätzt. Welcher Prozentsatz für die Größe der Untergruppe von vermutlich durch eine ADHS mit komorbiden depressiven Symptomen zutrifft, wird aus den Messdaten nicht ersichtlich.

13. Befund: Durch die Kombination der Daten aus der Durchführung von DTK und CDI verstärkt sich der schon zuvor vermerkte Hinweis auf den Anteil mutmaßlich

nicht durch depressive oder komorbide Störungen belasteter Kinder oder Jugendlicher als 2/3 der Stichprobe. Die Kombination von Daten aus DTK und CDI quantifizieren diesen Anteil auf 71,1%. Über die Belastung der dann verbleibenden 28,9% ist auch durch die Kombination der Messdaten aus DTK und CDI kein weiterreichenden Schluss als zuvor schon bekannt, zu gewinnen gewesen.

Erkennbar ist erneut, dass weiterreichende Aufschlüsse an das Vorliegen entsprechender klinischer Daten gebunden sind, die erst über die Durchführung von Validitätsstudien mit Stichproben griechischer Kinder gewonnen werden können.

13.5 Auf personenspezifische Merkmale bezogene Hinweise zum Auftreten depressiver Symptome und depressiver Störungen in der untersuchten Stichprobe

Die Befunde 1 bis 13 werden in ihrem Erklärungswert durch Ergebnisse gestützt, die über die Anwendung von DTK oder CDI als für die untersuchte Stichprobe insgesamt zutreffend ermittelt worden sind. Die Mitglieder dieser Stichprobe unterscheiden sich aber voneinander durch eine Vielzahl von Merkmalen. Zu erwarten ist, dass die mit DTK und CDI ermittelten Daten mit dem einen oder anderen der persönlichen Merkmale variieren oder aber zumindest variieren könnten. Insoweit als persönliche Merkmale in dem als Grundlage der Analyse dienenden Datensatz Merkmale enthalten, so etwa Geschlecht, Alter, Schule, Schulort, Ortsgröße, Schulklasse und Schultyp. Nachfolgend werden Ergebnisse aus der Analyse der nach persönlichen Merkmalen differenzierten Verteilungen von Daten zu Befunden zusammengefasst, dargestellt und erörtert. Falls höhere GPW aus DTK oder CDI zusammen mit persönlichen Merkmalen überzufällig auftreten sollten, wären sie als potentielle Risikofaktoren einzuschätzen. Potentielle protektive Faktoren wären ermittelt, wenn Merkmale zusammen mit besonders niedrigen GPW auftreten würden.

Diese Analyseschritte wurden, ebenso wie alle andere, unter Beachtung der einschlägigen internationalen Forschungsergebnisse durchgeführt. Zu prüfen war dementsprechend auch, ob international bekannte Hinweise auf risikobeladene oder protektive Faktoren auch innerhalb der untersuchten griechischen Stichprobe nachweisbar wären.

13.5.1 Geschlechtsspezifische Variationen von Hinweisen auf depressive Störungen

Für eine Gesamtgruppe die besteht Möglichkeit, dass die sie kennzeichnenden Werte über signifikant unterschiedliche Daten aus ihren Teilgruppen entstehen. Dem zufolge besteht allgemein die Notwendigkeit, empirische Daten für eine Gesamtgruppe auf deren Herkunft aus Teilgruppen mit unterschiedlichen Merkmalsträgern zu prüfen.

In der internationalen Literatur finden sich sowohl Belege für geschlechtsspezifische Unterschiede im Auftreten von depressiven Störungen als auch Berichte über das Fehlen von Nachweisen dafür. Im Hinblick auf solche Diskrepanzen ist auf den Sachverhalt aufmerksam zu machen, dass jede neu betrachtete Gesamtheit, so dann auch die Gesamtheit der weiblichen Teilnehmer an einer Untersuchung, sehr wahrscheinlich wiederum aus Teilgruppen unterschiedlicher Merkmalsträgerinnen besteht, beispielsweise aus solchen unterschiedlichen Alters. Zunächst wird nur auf das Merkmal Geschlecht einzugehen sein. Um die Übersichtlichkeit zu erhalten, finden sich in den Befunden, wie auch schon zuvor, Hinweise auf Risikogruppen. Damit sind die zwei Teilgruppen gemeint, die über definierte Trennwerte (cut-offs) nach den

Gesamtpunktwerten aus der Durchführung der Verfahren gebildet worden waren. Dabei ist jedes Mädchen und jeder Junge der untersuchten Stichprobe entweder der Teilgruppe mit eher niedrigem GPW und niedrigerem Risiko für das Bestehen einer depressiven Störung oder der Teilgruppe mit höherem GPW und höherem Risiko für das Bestehen einer depressiven Störung zugewiesen worden. Höheres oder niedrigeres Risiko ist gleichbedeutend mit höherer oder niedrigerer Wahrscheinlichkeit für das in diagnostischen Nachprüfungen aufzufindende Bestehen einer depressiven Störung. Da jedes teilnehmende Kind und jeder teilnehmende Jugendliche sowohl einen GPW aus dem Verfahren DTK als auch einem aus dem CDI zugewiesen bekam, war die Zuweisung an Risikogruppen ebenfalls zweimal zu vollziehen, Risikogruppen nach DTK und Risikogruppen nach CDI. Welche Zuweisung in den Befunden oder den Erörterungen zu ihnen jeweils gemeint ist, findet sich an den entsprechenden Stellen angegeben.

Der nachstehende 14. Befund nennt die in dieser Studie 2034GR gefundenen Beziehungen zwischen dem Merkmal Geschlecht und den Hinweisen auf das Bestehen depressiver Störungen.

14. Befund: Der isolierte Effekt des Geschlechts der Kinder oder Jugendlichen auf deren Zuordnung zu den gebildeten zwei Risikogruppen erscheint in einer ersten Betrachtung zwischen den zwei Verfahren unterschiedlich zu sein. Aus der entsprechenden Analyse der DTK-Daten wurde gefunden, dass die Mädchen in der nach DTK gebildeten Risikogruppe höherer Belastung ($GPW \geq 17$) signifikant überrepräsentiert sind. Die Überrepräsentation der Mädchen in dieser Gruppe hoher Belastung beträgt gegenüber den über Randsummen bestimmten Erwartungswerten etwa +8%. Im Gegensatz dazu hat die geschlechtsspezifische Verteilung der Gesamtstichprobe auf zwei nach CDI-GPW gebildeten Risikogruppen keine signifikanten Variationen gezeigt.

Immerhin ist an dieser Stelle zu erwähnen, dass statistische Werte, die Hinweise auf die Stärke des Zusammenhanges zwischen Geschlecht und DTK-Risikogruppen geben, nur schwach ausgebildet sind. Das könnte darauf verweisen, dass der zuvor erwähnte geschlechtsspezifische Effekt auf die Größe der Stichprobe zurückzuführen ist. Warum dieser statistische Hinweis dann nicht auch bei dem entsprechenden CDI-Ergebnis auftritt, wird hier zwar gefragt, aber nicht zu beantworten versucht. Immerhin war zuvor mehrfach die Annahme zu äußern gewesen, dass der DTK nicht nur das Auftreten depressiver Symptome erfasst und sich messtheoretisch vom CDI unterscheidet.

13.5.2 Altersspezifische Variationen der Hinweise auf depressive Störungen

So wie zuvor nach Zusammenhängen zwischen der isolierten Variable Geschlecht und den aufgenommenen Hinweisen auf depressive Störungen gefragt worden war, so wird jetzt die Frage im Hinblick auf die Variable Alter wiederholt.

15. Befund: Im Hinblick auf den Risikofaktor Alter wurde über den Einsatz beider Verfahren (CDI und DTK) und auch in Übereinstimmung mit internationalen Ergebnissen gefunden, dass mit ansteigendem Alter die Wahrscheinlichkeit für das selbstberichtete Auftreten von Hinweisen auf mit mutmaßlich klinischen Ausprägungen bestehende depressive Störungen zunimmt.

Die Analyse an den DTK-Daten hat gezeigt, dass die 8-, 9- und 11jährigen Kinder in der DTK-Gruppe niedriger Belastung von depressiven Symptomen überrepräsentiert sind. Demgegenüber ergibt sich dann, dass die 13-, 14- und 15jährigen in der Gruppe vermuteter hoher depressiver

Belastung (depBel 2) deutliche Überbesetzung erreichen. Das vermehrte Auftreten in der Gruppe depBel 2 beträgt gegenüber den über Randsummen bestimmten Erwartungswerten ungefähr +25% bei den 13jährigen, etwa 37% bei den 14jährigen und etwa 67% bei den Jugendlichen, die älter als 15 Jahre sind.

Ein ähnliches Ergebnis zeigten auch die Analysen der CDI-Daten. Die dort gefundenen altersspezifischen und statistisch signifikanten Abweichungen sind, arithmetisch betrachtet, weniger als die im DTK. Das Verhältnis, bezogen auf Überrepräsentation von Altersgruppen, ist 3 im CDI zu 6 im DTK, aber jedenfalls sind die Richtungen der Abweichungen identisch. Die 8-jährige Kinder sind zu fast 45% in der Gruppe hoher und damit auch klinischer Belastung unterrepräsentiert. Hingegen übersteigen die Altersstufen der 14- und 15+-jährigen Kinder und Jugendlichen mit ihrer Präsenz in der CDI-Gruppe hoher Belastung die Erwartungswerte um 33% bzw. 64%.

13.5.3 Kombinierte alters- und geschlechtsspezifische Variationen der Hinweise auf depressive Störungen

Aus der internationalen Literatur zum Auftreten von Depressionen bei Kindern und Jugendlichen ist bekannt, dass der mit dem Alter zu beobachtende Anstieg des Auftretens depressiver Störungen geschlechtsspezifische Unterschiede aufweist. Die sind auch in unserer Studie 2034GR nachzuweisen.

16. Befund: Als kritisches Lebensjahr für die Mädchen der Stichprobe ergab sich aus den DTK-Antworten das mit dem Alter von 13 Jahren. Kritisch werden diese Altersstufen hier genannt, weil mit ihnen das Risiko für das Auftreten depressiver Störungen, beziehungsweise das Auftreten von Hinweisen auf sie, deutlich ansteigt. Nach den CDI-Daten setzt das kritische Alter mit 14 Jahren ein. Die Zahl der nach Selbstberichten von depressiven Symptomen hoch belasteten Jungen nimmt nach Daten aus beiden Verfahren erst ab dem Alter von 15 Jahren signifikant zu.

Damit stützten unsere Ergebnisse partiell die in der internationalen einschlägigen Literatur vertretene These von einer erhöhten Vulnerabilität für das Auftreten depressiver Störungen zwischen der 12. und 14. Altersstufe.

13.5.4 Schulische Merkmale und auf sie bezogene Variationen der Hinweise auf depressive Störungen

Die zuvor erwähnten alters- und geschlechtsspezifischen Variationen in der Produktion von depressionsspezifischen GPW aus der Durchführung von CDI und DTK wurden auch in entsprechenden, nach Schulstufen (Grundschule – Gymnasium) der Kinder dichotomisierten Aufteilungen der Stichprobe; über Chi-Quadrat Analysen wiedergefunden. Das sollte nicht überraschen, unterscheiden sich doch die Schulstufen im Hinblick auf das Alter der Kinder. Immerhin ergibt sich so ein Beispiel dafür, wie die Aussagenlogik von der Benennung des Merkmals geleitet wird, mit dem die Daten klassifiziert oder gruppiert werden.

17. Befund: Die nach Schulstufen (Grundschule – Gymnasium) unterschiedene geschlechtsspezifische Verteilung der untersuchten Kinder oder Jugendlichen auf die gebildeten Risikogruppen variierte erwartungsgemäß mit der Produktion von GPW. Die

entsprechenden Ergebnisse aus den zwei durchgeführten Verfahren stimmen überein. Für die jüngeren Mädchen und Jungen (8 bis 12 Jahre alt) der Stichprobe wurde gefunden, dass sie in den gebildeten Gruppen niedriger Belastung von selbstberichteten depressiven Symptomen überrepräsentiert und in den Gruppen hoher Belastung unterrepräsentiert sind. Die älteren Jungen (12 bis 15+) weisen mit ihren GPW im CDI und DTK keine signifikanten Abweichungen in Vergleich zu den für die Gesamtstichprobe gefundenen Zahlen auf. Die älteren Mädchen erreichen aber eine signifikante Überbesetzung im Verhältnis zu den über Randsummen bestimmten Erwartungswerten für die Risikogruppen mit potentiell hoher Belastung. Das ergibt sich aus beiden der zwei verwendeten Verfahren. Für diese Mädchen ist die Wahrscheinlichkeit, von einer behandlungsbedürftigen depressiven Störung belastet zu sein, mit 39% bzw. 46% (laut CDI- und DTK-Daten) höher als die Erwartungswerte vermuten ließen.

Zu berücksichtigen ist an dieser Stelle die Überlegung, dass in den zuvor erwähnten Ergebnissen zu einer die Mädchen aus Gymnasien treffenden höheren Wahrscheinlichkeit vom Auftreten einer vermutlich klinisch relevanten depressiven Störung belastet zu sein, neben Geschlecht und Alter sehr wahrscheinlich noch andere Variablen beigetragen haben, die dann auch in der Befundbildung zu berücksichtigen wären, hätte man sie identifiziert. Das sind alle jene, die der persönlichen Interaktion mit dem Schultyp die Ausprägung geben. Die zwei Schultypen, griechische Grundschulen und griechisches Gymnasium, unterscheiden sich inhaltlich und strukturell sehr stark voneinander. Entsprechend unterschiedlich sind auch die Anforderungen, die den Kindern gestellt werden. Wäre die depressogene Kraft der Gymnasien untersucht und nachgewiesen, dürfte man aus den zuvor dargestellten Ergebnissen die These ableiten, dass ältere griechische Mädchen bedeutsam größere Schwierigkeiten haben, sich auf das Leben in den von ihnen altersentsprechend besuchten staatlichen Bildungsinstitutionen einzustellen als Mädchen in Grundschulen. Die Frage, „warum speziell die Mädchen an Gymnasium und nicht die dort befindlichen Jungen“ solchen Schwierigkeiten zeigen, wird wegen der Schwierigkeiten, die sich bei dem Versuch einstellen, empirisch tragbare Aufklärung zu erreichen, zwar gestellt, aber nicht beantwortet. Die Vermutung, sowohl persönliche, psychophysiologische, entwicklungspsychologische und andere Aspekte seien zu berücksichtigen, stünde noch immer für die relative Hilflosigkeit der Forscherinnen und Forscher gegenüber den beobachtbaren Phänomenen.

Ein weiterer Befund ordnet das Wiederholen einer Schulklasse als potentiellen Risikofaktor ein.

18. Befund: Die Analysen aus den CDI- und DTK-Daten haben übereinstimmend gezeigt, dass die Verteilung der Gymnasiasten auf die zwei gebildeten Risikogruppen signifikant mit dem altersgerechten oder davon abweichenden Schulbesuch variiert. Bei überalterten Kindern in Gymnasien liegt die Wahrscheinlichkeit für eine Zuordnung in die Gruppen hoher Belastung nach GPW aus DTK oder CDI mit 55% beziehungsweise 33% deutlich über den statistischen Erwartungswerten.

Da die vorhandenen Ergebnisse keine kausale Zuschreibung gestatten, bestehen für den genannten Zusammenhang drei konkurrierende Thesen über die Entstehung des Zusammenhanges. Nach einer ersten These könnten depressiv belastete Kinder oder Jugendliche häufiger als andere in die Situation kommen, eine Klassenstufe wiederholen zu müssen. Nach einer zweiten wäre die Wiederholung einer Klassenstufe ein das Auftreten depressiver Störungen begünstigendes Ereignis. Schließlich könnten beide Thesen zutreffen.

Darüber hinaus wurde gefunden, dass die überalterten Kinder aus Gymnasien im DTK durchschnittlich über signifikant mehr Symptome von Dysphorie, Selbstwertproblemen und Agitiertheit berichten als ihre alterskomform beschulten Mitschüler und Mitschülerinnen.

Im Gegensatz zu dem Befund für die Gesamtgruppe dieser Altersspanne- ist die geschlechtsspezifische Verteilung der überalterten Kinder auf die gebildeten CDI- und DTK-Gruppen niedriger und hoher depressiver Belastung nicht signifikant unterschiedlich. Von dem höheren Risiko sind die in ihren Schulklassen überalterten Jungen und Mädchen gleichermaßen betroffen.

13.5.5 Wohnort und Variationen der Hinweise auf depressive Störungen

Der unterschiedliche Wohnort ist ebenfalls eines der Merkmale, über das Teilgruppen von untersuchten Kindern und Jugendlichen zu unterscheiden sind. Mit einer enger gefassten Bestimmung ist der Wohnorten genannte Platz eigentlich der Schulort. Für die große Mehrzahl der untersuchten Jungen und Mädchen wird gelten, dass beides namentlich übereinstimmt. Dann ist die Schule in dem Ort, in dem die Schüler und Schülerinnen wohnen. Mitunter könnten aber auch Jungen oder Mädchen im lokalen Randgebiet des Schulortes wohnen und dennoch die Ortsschule besuchen. Die Bezeichnung Wohnort wird beibehalten, weil die gegebene Unschärfe zum einen bei der gegebenen Datenlage nicht aufzulösen ist und sich zum anderen auslöst, wenn der Name des Wohnortes als für den Ort selbst und seine lokalen Randgebiete stehend verstanden wird.

Indem die aufgenommenen Daten auch unter Einbeziehung der Variablen des Wohnorts analysiert werden, folgt dies der These von einer potentiellen Abhängigkeit des Auftretens depressiver Störungen bei Kindern und Jugendlichen in Abhängigkeit vom Wohnsitz. In der einzuziehenden internationalen Literatur ist der Wohnsitz nur gelegentlich und vor allem im Zusammenhang mit extremen Ortslagen beachtet worden, so etwa im Zusammenhang mit extrem geringer Sonnenscheindauer in den Wintermonaten nördlich des Polarkreises.

Die Annahme, der Wohnort der Kinder habe eine Bedeutung für das selbstberichtete Auftreten depressiver Symptome oder könne sie besitzen, stünde theoretisch in deutlicher Übereinstimmung mit interaktionalen Ansätzen zur Erklärung der Entwicklung depressiver Störungen. Darüber hinaus ist in Griechenland wenig bekannt über die psychische Gesundheit von griechischen Kindern außerhalb der Universitätsstädte, was als hinderlich für die Planung und Realisierung einer angemessenen nationalen Gesundheitspolitik zu bewerten ist.

18. Befund: Die von den untersuchten Kindern und Jugendlichen erreichten Gesamtpunktwerte im CDI und im DTK variieren in signifikanter Weise mit den unterschiedlichen Wohnorten der Kinder.

Für die Gesamtstichprobe wurde gefunden, dass laut den DTK-Analysen in 5 von insgesamt 12 in die Studie aufgenommen besuchten Wohnorten, signifikante Abweichungen von Erwartungswerten auftraten. Aus entsprechenden Analysen in den über den DTK gesammelten Daten ergeben sich signifikante Abweichungen von den dort gerechneten Erwartungswerten für 2 Orte. Die Unterschiedlichkeit der Ergebnisse aus den beiden verglichenen Messinstrumenten durchzieht die gesamte Untersuchung und war schon mehrfach als Effekt der unterschiedlichen Spezifität dieser Verfahren für psychische Störungsbilder verstanden worden. Um so bedeutsamer ist die Nachfrage, ob die vom DTK nominierten Orte auch in der durch den CDI bekannt

werdenden Liste erscheinen. Im Hinblick auf diese zwei Orte – Stavroupoli und Heraklio – stimmen die Ergebnisse aus den zwei verwendeten Verfahren überein.

Die nachstehenden Tabellen stellen zusammengefasst die für die Gesamtstichprobe und für geschlechts- und altersspezifische Teilgruppen gefundene Ergebnisse dar. Die in der ersten Tabelle angegebenen Prozentsätze beziehen sich auf prozentuale Abweichungen von über Randsummen bestimmten Erwartungswerten.

Tabelle 13.3: Zusammenfassende Darstellung der mit den Wohnorten variierenden Zuweisungen der Gesamtstichprobe auf nach CDI und DTK Gesamtpunktwerten definierten Risikogruppen

CDI		DTK	
Stavroupoli	Plus 59,2%	Stavroupoli	Plus 52,5%
Heraklio	Plus 74,6%	Heraklio	Plus 56,5%
		N. Agioneri	Plus 33,9%
		Athen	Minus 16,3%
		Karditsa	Minus 37,7%

Tabelle 13.4: Zusammenfassende Darstellung der mit den Wohnorten variierenden Zuweisungen der Jungen und Mädchen aus den untersuchten Grundschulen auf nach CDI und DTK Gesamtpunktwerten definierten Risikogruppen

CDI		DTK	
n.s.		Jungen	Mädchen
		N. Agioneri	EW Plus 216%
		Heraklio	EW Plus 161,90%
		Karditsa	Minus 100% Minus 71%

Zeichenerklärung: n.s.= keine signifikante Abweichung von Erwartungswerten, EW: Erwartungswerte

Tabelle 13.5: Zusammenfassende Darstellung der mit den Wohnorten variierenden Zuweisungen der Jungen und Mädchen aus den untersuchten Gymnasien auf nach CDI und DTK Gesamtpunktwerten definierten Risikogruppen

	CDI		DTK		
	Jungen	Mädchen	Jungen	Mädchen	
Stavroupoli	EW	Plus 161,90%	Stavroupoli	EW	Plus 87%
Heraklio	EW	Plus 107,79%	Heraklio	EW	Plus 81%
N. Agioneri	EW	Minus 80,77%	Athen	Minus 42%	EW

Wie aus Tabelle 13.5 zu entnehmen ist, sind die Ergebnisse aus den zwei Verfahren für die Stichprobe der Grundschulkinder nicht übereinstimmig. Die Analysen der CDI-Daten zeigte keine signifikanten Variationen der nach Wohnort geschlechtsspezifisch unterschiedenen Verteilung der Grundschulkinder auf die definierten CDI-Risikogruppen. Im Gegensatz dazu wurden aus den DTK-Daten 3 Orte identifiziert, die von Erwartungswerten signifikant abweichen. In zwei von denen (N. Agioneri und Heraklio) sind die Mädchen in der DTK Gruppe hoher Belastung

signifikant überrepräsentiert. Interessanterweise wurde dabei auch ein Ort gefunden (Karditsa), wo sowohl die Jungen als auch die Mädchen die Erwartungswerte signifikant unterschreiten. Die Frage, warum für die Teilstichprobe der Grundschul Kinder die Ergebnisse zwischen den zwei Verfahren nicht übereinstimmen, sollte wiederum als ein Effekt der unterschiedlich konstruierten Messverfahren verstanden werden.

Für die Teilstichprobe der Gymnasiasten sind die Ergebnisse zwischen den zwei Verfahren relativ homogen. Bei beiden Verfahren wurden signifikante Variationen in 3 Orten gefunden, wobei 2 von beiden Verfahren genannt werden (Stavroupoli und Heraklio). Darüber hinaus sind es jeweils die Mädchen in diesen Orten, die mit ihren GPW im CDI und DTK die jeweiligen Erwartungswerte signifikant überschreiten.

Die mit dieser Darstellung hervorgehobenen Ergebnisse aus der mit dem Wohnort der Kinder signifikant variierenden Verteilung der Gesamtpunktwerte aus DTK und CDI geben Anlass zu vielen Nachfragen, vor allem aber zu der nach der Bedeutung dieser Unterschiede. Versuche, nach Antworten zu suchen, werden an dieser Stelle unterlassen, weil nur Spekulationen zu äußern wären. An anderer Stelle sollte die Diskussion darüber erneut aufgenommen werden. Für die Studie 2034GR die auftretenden Unterschiede allein schon wegen der mit ihnen gegebenen lokalen und regionalen Informationen von Bedeutung.

Der nachstehende Befund hebt das hervor.

19. Befund: Für die systematisch über Griechenland verteilten Wohnorte der in dieser Studie untersuchten 2034 griechischen Schulkinder verweisen deren Daten aus DTK und CDI mit signifikanten Unterschieden in der Varianz der Hinweise zum Auftreten depressiver Störungen auf die Möglichkeit der Existenz lokaler Unterschiede in den Bedingungen für die Entwicklung oder den Erhalt depressiver Störungen. Zur Ursache für die auftretenden lokalen Unterschiede sind mehrere und auch sehr unterschiedliche Hypothesen vorzubringen. Eine von diesen ist die hier nachstehend geäußerte: „Das von Kindern und Jugendlichen erlebte soziokulturelle Milieu ihres Wohn- und Schulortes variiert mit den Orten in seiner über die Interaktion von Person und Umwelt pathologisierenden oder protektiven Wirkung auf die individuelle emotionale Befindlichkeit.“

Für den Fall, dass die beobachteten Variationen nicht zufällig aufgetreten sind – was nach den vorliegenden Ergebnissen zu vermuten ist - scheint der „Wohn- oder Schulort“ einen beachtenswerten Einfluss und vor allem auf Mädchen und besonders auf die älteren Mädchen zu haben. In diesem Zusammenhang erwähnen Madianos und Mit. (1999) Ergebnisse aus Studien, die dafür sprechen, dass „...socioeconomic factors such as social disadvantage or deprivation, lack of infrastructure and support, emigration, social mobility, and population density might have an effect on the geographical variation of rates of treated psychiatric illness in different parts of the world“ (Madianos, Zacharakis, Tsitsa & Stefanis, 1999, S. 477).

Bedauert werden sollte das bislang zu bemerkende Fehlen von praxisnahen Aufschlüssen und Einsichten in das entsprechende Bedingungsgefüge. Das vor allem, weil der Beachtung der mitgeteilten Informationen über lokale Unterschiede im Auftreten von Hinweisen auf depressive Störungen unter Schülerinnen und Schülern für die Planung und Realisierung präventiver Interventionen eine handlungsleitende Funktion zukäme.

In der auf den Wohn- oder Schulort bezogenen Befundbildung ist diese Ortskategorie mit ihren Variationen das die zu Grunde liegende Datenmatrix umorganisierende Merkmal. Die Wohn-

oder Schulorte selbst sind dann weiterhin aber auch Träger von Merkmalen, über die sie voneinander unterschieden werden können. Wenn – wie im 20. Befund unterstellt - das „von Kindern und Jugendlichen erlebte soziokulturelle Milieu“, also die so kategorisierte Eigenheit eines Ortes und damit etwas für ihn Charakteristisches, spezifische Bedingungen für ein vermehrtes oder vermindertes Auftreten depressiver Belastungen setzt, dann variieren Orte folgerichtig in eben dieser Qualität. Die ist in der Studie 2034GR nicht gemessen worden. Da diese Qualität mit den Orten variiert, wurde sie bei einer nach Wohnorten organisierten Datenanalyse wirksam, ohne selbst identifiziert worden zu sein. Die Möglichkeit wäre gegeben, dass dies „von Kindern und Jugendlichen erlebte soziokulturelle Milieu“ eines Ortes mittelbar mit Merkmalen variierte, die Orte unterscheiden, so etwa mit der Zahl der Einwohner. Die Lebensumstände in kleineren griechischen Orten sind andere als die in größeren Städten. Variierten die Hinweise auf depressive Störungen unter Kindern und Jugendlichen mit der Einwohnerzahl der Orte, wäre zumindest eine Information über die mit den Orten variierenden Bedingungen für das Auftreten depressiver Störungen unter Schulkindern bekannt. In den durchgeführten Analysen war dem nachgegangen worden.

20. Befund: Die Zahl der Einwohner eines Ortes hatte als Variable in unserer Untersuchung keinen nachweisbaren Einfluss auf die Variation der Gesamtpunktwerte aus CDI und DTK und somit auch keinen Zusammenhang mit dem Auftreten der Hinweise auf depressive Störungen unter Schülerinnen und Schülern.

Der 21. Befund widerspricht damit der möglicherweise plausibel erscheinenden These von einer mit der Einwohnerzahl von Städten variierenden protektiven oder pathogenisierenden Einflussnahme der Lebensbedingungen an den unterschiedlichen Wohnplätzen. Der 21. Befund gilt auch für die geschlechtsspezifische Verteilung der Kinder aus Grundschulen auf die definierten CDI- und DTK-Risikogruppen. In der Verteilung der Gymnasiasten traten jedoch signifikante Unterschiede auf. Allerdings stimmten die Ergebnisse der Analysen von Daten aus DTK und CDI einmal mehr nicht überein. Nach den DTK-Analysen ergab sich, dass ältere Jungen aus Groß- und Mittelstädten in der Hochrisikogruppe unterrepräsentiert und die Mädchen aus Großstädten überrepräsentiert sind. Diese Feststellung gibt Gelegenheit zu Spekulationen über die hinter den Fakten zu vermutenden Ursachen. In der Verteilung der Mädchen und Jungen auf die CDI-Risikogruppen nach GPW aus dem CDI zeigten sich keine signifikanten Abweichungen von Erwartungswerten.

13.6. Beantwortung der Frage nach dem Wissen von Lehrkräfte über das Auftreten selbstberichteter depressiver Belastungen unter ihren Schulkindern

Die in der Studie 2034GR eingesetzten Verfahren sind zwei unterschiedlichen Kategorien von Messinstrumenten zuzuordnen. Die einen – DTK, CDI, SDQ – nehmen von Kindern oder Jugendlichen selbst berichtete Angaben zum Eigenerleben depressiver Symptome auf. Die anderen – dazu gehört in dieser Studie die Skala ESDM - sammeln Fremdbeobachtungen zum Auftreten depressiver Symptome bei Kindern und Jugendlichen.

Da der SDQ wegen unzureichender Messqualitäten des Verfahrens nicht zur Befundbildung herangezogen wurde, sind die Ergebnisse aus der Durchführung der Skala ESDM eine letzte Gruppe von Informationen über das mutmaßliche Auftreten von depressiven Symptomen, Episoden oder Störungen unter den griechischen Schulkindern der untersuchten Stichprobe.

Die Skala ESDM verhalf zu Daten, über deren Analyse auf die für Prävention und Intervention gegenüber depressiven Störungen bei Schulkindern durchaus bedeutungsvolle Frage zu antworten

war, ob die griechischen Lehrkräfte in den Schulen spezifisches und insoweit dann nutzbares Wissen über depressive Symptome bei den Kindern ihrer Schulklasse verfügbar hätten. In einen solchen Fall wäre dann begründet darüber nachzudenken, in welcher Weise Lehrkräfte zu Helfern in der Versorgung der durch depressive Störungen bedrohten Schulkinder werden könnten. Als Verfahren zur Untersuchung des verfügbaren Wissens der Lehrkräfte über die depressiv belasteten Kinder ihrer Schulklassen wurde eine griechische Version der von Domènech (1985) zu diesem Zweck entwickelte Skala, der Skala ESDM, eingesetzt. Die Skala stellt den Lehrkräften eine Aufgabe, die sie als eine aktuelle Stichprobe aus der Menge aller denkbaren Anwendungen des definierten spezifischen Wissens zieht und die sie nachfolgend der jeweils intendierten Analyse zuführt. Die verwendete Aufgabe bietet den Lehrkräften die Gelegenheit, das ihnen verfügbare spezielle Wissen dadurch zu beweisen, dass sie zu 16 dem CDI nahen depressionsspezifischen Items angeben, wie die von ihnen unterrichteten Schülerinnen und Schüler einer Schulklasse jeweils auf eben diese Items antworten würden. Der jeweils erreichte Grad der Übereinstimmung stünde als Beleg für das zum Messzeitpunkt von den Lehrkräften anwendbare Wissen zur Verfügung. Die von der Skala vorgegebenen Antwortmöglichkeiten sind dreistufig ausgelegt, so dass $16 \cdot 3 = 48$ Wahlmöglichkeiten gegeben sind. Das Erreichen von Übereinstimmung durch Antworten, die nicht auf spezifisches Wissen über die jeweils gemeinten Kinder oder Jugendlichen beruhen, ist extrem unwahrscheinlich.

Um die Frage nach dem Grad des depressionsspezifischen Wissens der griechischen Lehrkräften über das Ausmaß der Belastung ihrer Schulkinder durch depressive Symptome annähernd korrekt einschätzen zu können, war zunächst ein Leistungskriterium einzusetzen. Als Kriterium für den Wissensstand wurde die Übereinstimmung der Antworten der Lehrkräfte mit den von den Kindern und Jugendlichen abgegebenen Selbsteinschätzungen benutzt. Des Weiteren umfasste die Auswertung der Daten explorative Analysen zum Antwortverhalten der beteiligten Lehrkräfte auf die Fragen der Skala ESDM.

Untersuchungen zum Auftreten von Störungsbildern mit Daten aus Selbst- und Fremdbeobachtungen sind in der Regel mit der Frage konfrontiert, welche Bedeutung den unterschiedlichen Daten und der etwa erreichten Übereinstimmung der Daten aus beiden Perspektiven aktuell zuerkannt werden soll. Liegt eine externe Diagnose zum Störungsbild vor, kann die als Bezugspunkt gelten. Fremd- und Selbsteinschätzung könnten jeweils mit der geltenden Diagnose und zudem miteinander übereinstimmen. Sind, wie in der Studie 2034GR keine klinisch erstellten Diagnose für jedes der Schulkinder verfügbar, stehen die Daten aus Selbst- und Fremdbeobachtung zunächst nur für sich selbst und als so wie dargestellt gegeben. Gefragt werden könnte dann auch, ob die Lehrkräfte empathisch urteilen. Dann sollten sie mit den Selbsteinschätzungen der von ihnen unterrichteten Schulkinder übereinstimmen. Eine denkbare Variante des Einsatzes der Skala wäre auch, die Kinder und Jugendlichen über ESDM danach zu fragen, welche Antwort die Lehrkraft auf die ihr im ESDM gestellten Fragen als zutreffend für die jeweiligen Schülerinnen und Schüler geben werden. Der dann erreichte Grad der Übereinstimmung wäre zum einen ein Maß für Empathie seitens der Schulkinder und Jugendlichen, andererseits aber auch, wenn die Selbsteinschätzung der Kinder und Jugendlichen damit verglichen wird, ein denkbare Maß für das Erleben, zutreffend erkannt zu werden. Das Gegenteil. Das Erleben verkannt zu werden, wäre nicht nur pädagogisch eine wichtige und nutzbare Information.

In der Untersuchung 2034GR wurde nach der Übereinstimmung des Urteils der Lehrkraft mit der Selbsteinschätzung der Schulkinder gefragt. Eine solche Übereinstimmung wäre kein Nachweis für das Bestehen einer diagnostizierbaren depressiven Störung, wohl aber der Beweis einer

Übereinstimmung zwischen den Urteilenden. Diese Übereinstimmung zwischen Selbst- und Fremdbeobachtung ist nicht zwingend zu erwarten.

Die Selbsteinschätzung der Schüler und Schülerinnen ist als Produkt des eigenen verbalen Handelns mit sich und seinen Wiederholungen in der Regel in größerer Übereinstimmung als jede personale Fremdbeurteilung. Im Bereich der Depressionsforschung weisen die empirischen Daten diese allgemein akzeptierte Einschätzung als zutreffend nach (s.u.a. Edelbrock, Costello et al., 1986; Polaino & Domènech, 1993).

Die im Folgenden zusammenfassend darzustellende Befunde zur Beantwortung der Frage nach dem Wissen von griechischen Lehrkräften über die selbstberichtete emotionale Befindlichkeit ihrer Schulkinder sind als Belege für den Grad der speziellen Treffsicherheit der durch Lehrkräfte gegebenen Einschätzung der selbstberichteten emotionalen Befindlichkeit ihrer Schülerinnen und Schüler valide. Die mit dem ESDM gestellte Aufgabe simuliert den in einem entsprechenden Handlungszusammenhang auftretenden speziellen Beurteilungsprozess.

Eine erste Annäherung an die Beantwortung der Frage nach der speziellen empathischen Treffsicherheit der von den Lehrkräften (N= 114) aus 30 griechischen Grundschulen und Gymnasien abgegebenen Urteile die selbstberichtete emotionale Befindlichkeit ihrer Schulkinder bieten die Ergebnisse aus der Prüfung der Übereinstimmung zwischen den Urteilen der Lehrkräfte und Selbstangaben der Kinder und Jugendlichen in den depressionsspezifischen Verfahren CDI, DTK und SDQ. Die gefundene Korrelationswerte (Tabelle 12.9) sind – bezogen auf Gesamtscores und für die Gesamtstichprobe der Lehrkräfte – schwach und überschreiten nicht den Wert von $r = 0,27$.

21. Befund: Die innerhalb dieser Studie gefundenen sehr schwachen Korrelationen zwischen Selbstberichten der Kinder im CDI, DTK und SDQ und Urteilen der Lehrkräfte im ESDM geben einen ersten Hinweis darauf, dass die Informiertheit der beteiligten Lehrkräfte als Gesamtstichprobe über die selbstberichtete depressive Belastung der Kinder ihrer Schulklassen sehr begrenzt ist und nur weniger als 10% der Gesamtvarianz der Selbsteinschätzungen zu erklären vermag.

Unter Kenntnisnahme des 22. Befundes ist anzunehmen, dass die Lehrkräfte der untersuchten Stichprobe sehr wenig darüber wissen, wie ihre Schülerinnen und Schüler sich fühlen.

In Zusammenhang mit diesem Befund wurde an entsprechende Stelle des Kapitels 12 die Hypothese geäußert, der Skala ESDM fehle eine hinreichende konvergente Validität gegenüber den die Selbsteinschätzungen der Kinder und Jugendlichen erfassenden Verfahren CDI, DTK und SDQ. Der mitgeteilte Befund über niedrige Korrelationen zwischen Fremdurteilen von Lehrkräften und Selbstangaben der Schülerinnen und Schüler steht in Übereinstimmung mit Ergebnissen aus internationalen Studien der Depressionsforschung bei Kindern und Jugendlichen in unterschiedlichen Populationen und nach Verwendung verschiedenartiger Messinstrumente (s.u.a. Almquist et al., 1999; Groen et al., 1997; Kashani et al., 1983; Moor et al., 2000; Reicher, 1998). Die hier gefundenen Korrelationswerte bestätigen das Ergebnis aus einer Meta-Analyse entsprechender Studien durch Achenbach, McConaughy und Howell (1987), die über eine durchschnittliche Korrelation zwischen Selbstangaben und Fremdurteilen von Lehrkräften von etwa 0,27 berichten.

Der hier formulierte Verdacht einer geringen Informiertheit der an unserer Studie beteiligten Lehrkräfte über die emotionale Befindlichkeit ihrer Schüler und Schülerinnen scheint auch in Übereinstimmung mit dem Ergebnis aus einer Studie von Kleftharas und Didaskalou (2006) mit Lehrkräften von 333 Kindern aus Volos, Griechenland, zu stehen. Die Autoren haben - anhand von Kreuztabellen zwischen Selbstangaben im CDI und Lehrurteilen in einer von den Autoren

entwickelten Skala – gefunden, dass die an jener Studie beteiligten Lehrkräfte nur 14% der Kinder, die nach Selbstangaben im CDI von einer „mild“ oder „severe“ genannten Form depressiver Belastungen ($\text{CDI-GPW} \geq 10$) betroffen waren, auch als solche erkannt und in ihren Berichten als solche benannt haben.

Eine vertiefte Exploration des Grades der Informiertheit und der Erkennensleistung unserer Lehrkräften hinsichtlich der nach Selbstzuschreibungen emotional belasteten Kinder oder Jugendlichen ihrer Schulklassen ergab sich aus dem Vergleich der Lehrurteile in der Skala ESDM mit den Selbstangaben der Kinder und Jugendlichen in den zwischen beiden Verfahren inhaltsgleichen Items. Um die Diskussion der Ergebnisse zu erleichtern, wird an dieser Stelle die im Kapitel 12 erstellte Tabelle zu den Analyseergebnissen wiedergegeben.

Tabelle 13.6: Prozentuale Trefferliste für alle im ESDM verwendeten 9 CDI-Items

Item	PrZ „Selten“	PrZ „manchmal“	PrZ „sehr oft“
„Sadness“	59,2%	45,3%	12,9%
„crying spells“	81,7%	22,9%	0 %
„fatigue“	62,8%	38,7%	4,6 %
„school dislike“	79,1%	22,1%	9,6 %
„self depreciation“	68,0%	35,3%	9,9 %
„fighting“	70,7%	28,8%	17,6 %
„self-blame“	72,8%	72,6%	1,5 %
„indiciviness“	51,7%	25,7%	8,8 %
„school performance decrement“	67,1%	46,9%	41,3 %

Zeichenerklärung: PrZ = Prozentsatz richtiger Zuordnungen durch die Lehrkräfte zum Grundwert der durch die Kinder erfolgten Selbstzuschreibung

Aus der Tabelle 13.6 ist zu entnehmen, dass der Grad der Übereinstimmung der einerseits über ESDM-Urteile von den Lehrkräften gegebenen Einschätzungen des andererseits durch Selbstzuweisungen von CDI-Items seitens der Kinder und Jugendlichen bewirkte Auftreten von Hinweisen auf möglicherweise bestehende depressive Störungen von Item zu Item und weiterhin auch mit den Ausprägungsgraden „selten“, „manchmal“ und „sehr oft“ variiert. Beim „seltenen“ Auftreten eines Symptoms wechselt die Übereinstimmung im Bereich zwischen 52% und 82%. Bei der Einschätzung des Auftretens als „gelegentlich“ variiert die Übereinstimmung zwischen 22% und 72%. Zum Ausprägungsgrad „sehr oft“ liegen die Übereinstimmungsquoten zwischen den 9 gemeinsamen Items im CDI und ESDM zwischen 0% und 41,3%, wobei, in 8 der 9 gemeinsamen Items die Übereinstimmung zum Ausprägungsgrad „sehr oft“ unter 20% bleibt. Durchschnittlich ergibt sich eine Übereinstimmung von 68% beim seltenen Auftreten eines Symptoms, von 38% für den Fall der Auskunft, „manchmal“ eines der depressionsrelevanten Symptome zu erleben und nur von 12% bei Symptomen, die Kinder oder Jugendliche der Stichprobe als „sehr oft“ erlebt einstufen.

22. Befund: Die Informiertheit der an unserer Untersuchung beteiligten Lehrkräfte über die selbst berichtete emotionale Befindlichkeit ihrer Schüler und Schülerinnen ist im hohen Maße sowohl von der jeweils angesprochenen Qualität der Befindlichkeit abhängig, repräsentiert durch die verschiedenen Items, als auch vom ausgewählten Ausprägungsgrad. Die Übereinstimmung der von den Lehrkräften gegebenen

Urteile mit den Selbstberichten ist für sich weniger depressiv gestimmt darstellende, eher fröhliche Schülerinnen und Schüler deutlich höher als für die Gruppen von Kindern, die laut ihren Angaben „manchmal“ oder „sehr oft“ von bestimmten depressiven Symptomen betroffen sind.

Von Interesse – insbesondere auch bezüglich der Diskussion um eine Mitarbeit der Lehrkräfte am Erkennen depressiv gefährdeter oder schon betroffenen Kinder in ihrer schulischen Umgebung – ist die Tatsache, dass der einzige Probleminhalt, für den eine relativ hohe Übereinstimmung zwischen den Urteilen der Lehrkräfte und den Selbstangaben der Kinder oder Jugendlichen nachgewiesen wird, beim Ausprägungsgrad „sehr oft“ (41,3%) für die zufrieden stellende Schulleistung („school performance decrement“) der Kinder auftritt.

Der zuvor formulierte Satz über eine Variation der Urteile mit den Items und ihrer von den Lehrkräften angegebenen Ausprägung – die über die in Tabelle 13.6 dargestellten Angaben zunächst nur vermutet werden darf – konnte durch Analysen der Verteilung der Urteile der Lehrkräfte für die einzelne ESDM-Items als empirisch gesichertes Arbeitsergebnis nachgewiesen werden. In diesem Zusammenhang haben entsprechende Analysen auch gezeigt, dass bei der Vergabe des Urteils „sehr oft“ hinsichtlich des Ausprägungsgrads eines ESDM-Symptoms in der Gesamtmenge der durch Lehrkräfte gegebene ESDM-Urteile solche Aussagen vorherrschen, die sich auf schulübliche Kategorien von Verhaltenseinstufungen – wie z.B. Aufmerksamkeit und Leistung – beziehen. Die Dominanz der explizit unterrichtsbezogenen ESDM-Items bleibt erhalten, auch wenn als Vergleichswert die Relation zwischen „nicht betroffenen“ (ESDM-Antwort „selten“) und „betroffenen“ (ESDM-Antworten „manchmal“ und „sehr oft“) herangezogen wird.

Darüber hinaus verweisen die Ergebnisse aus der durchgeführten Faktorenanalyse des ESDM darauf, dass die an unserer Untersuchung beteiligten Lehrkräfte in ihrer Wahrnehmung, belegt durch deren Wahrnehmungsurteile, bei der Beantwortung der ESDM-Items einer durch ihre schulische Praxis berufsspezifischen und unterrichtsrelevanten Akzentuierung folgen.

Alle zuvor erwähnten Befunde sprechen zudem dafür, dass die Lehrkräfte bei der Beantwortung der ESDM-Fragen sich weniger durch die in der Schule wahrnehmbaren depressionsspezifischen Hinweise leiten lassen, sondern deutlich eher den unter Zuhilfenahme von schulpädagogischen Kategorien strukturierten schulischen Erfahrungen mit den Kindern ihrer Klasse folgen. Ihre Urteilsweise ist berufsabhängig entwickelt sind folgt weitaus weniger klinisch relevanten Beobachtungen. Die über eine Faktorenanalyse gefundenen Urteilungskategorien entsprechen oberbegrifflich der „Bewertung von Schulleistungen“, der „Bewertung des sozialen Verhaltens in der Schule“, einer „Einstufung des Störverhaltens im Unterricht“, sowie der „Einstufung oppositionellen und streitbaren Verhaltens“. Da die Faktorenanalyse nicht prüft, wohl aber die Bildung von Hypothesen nahelegt, wird die genannte Feststellung als Hypothese in die Liste der Befunde aufgenommen.

23. Befund: Die Ergebnisse aus Analysen der ESDM-Daten sprechen dafür, dass die Lehrkräfte bei der Beantwortung der ESDM-Fragen sich weniger durch die in der Schule wahrnehmbaren depressionsspezifischen Hinweise leiten lassen, sondern deutlich eher den unter Zuhilfenahme von schulpädagogischen Kategorien strukturierten schulischen Erfahrungen über die Kinder ihrer Klasse folgen. Die Ergebnisse aus Faktorenanalysen stützen die nachstehende Hypothese: „Die im ESDM abgegebenen Urteile der Lehrkräfte folgen nur partiell klinischen Urteilskategorien und dann weiterhin vor allem 4 bekannten schulpädagogischen

Bewertungsdimensionen so 1. der „Bewertung von Schulleistungen“, 2. der „Bewertung des sozialen Verhaltens in der Schule“, 3. der „Einstufung des Störverhaltens im Unterricht“, sowie 4. der „Einstufung oppositionellen und streitbaren Verhaltens“.

Diese Hypothese könnte in Zusammenhang mit Wissen über weitere Einflussfaktoren die niedrige Übereinstimmung zwischen Selbstangaben und Lehrurteilen erklären helfen.

Die dargestellte Befundlage differenziert und erweitert sich über die Analyse der individuellen Urteile der Lehrkräfte im ESDM. Der insgesamt geringe Wissensstand der Lehrkräfte als Stichprobe verdeckt das ihn konstituierende Zusammentreffen von drei in ihren Trefferquoten unterschiedlichen Teilgruppen von Lehrkräften. Eine erste etwa 20% der Lehrerinnen und Lehrer umfassende Teilgruppe erreicht ganz im Gegensatz zu ihren Kolleginnen und Kollegen signifikant vom Zufallstreffern abweichende Übereinstimmungen mit den Kindern oder Jugendlichen ihrer Schulklasse. Die von einzelnen Lehrkräften erreichte Übereinstimmung erklärt, je nach individueller Leistung, zwischen 20% und 50% der gemeinsamen Varianz mit den Selbstberichten der Schülerinnen und Schüler. Demgegenüber sind die Urteile einer weitaus größeren und etwa 80% umfassenden Teilgruppe von Lehrkräften statistisch nicht von dem mit $r=0$ definierten Nichtwissen zu unterscheiden. Die von ihnen erreichten Trefferquoten könnten auch durch zufällig mit den Selbstberichten der Kinder oder Jugendlichen gleichsinnig abgegebene Urteile zustande gekommen sein. In dieser Teilgruppe findet sich sogar eine kleine Untergruppe von Lehrkräften, die negative Korrelationen erreichen und somit eine deutliche Verkennung der emotionalen Befindlichkeit der Schülerinnen und Schüler demonstrieren.

Die damit nachgewiesenen drei Qualitätsstufen des Urteils der Lehrkräfte lassen zwei Fragen aufwerfen. Eine erste wäre, ob der relativ geringe Anteil von 20% jener Lehrkräfte, die spezifisches Wissen über die emotionale Situation ihrer Schülerinnen und Schüler zu nutzen wissen, durch Training zu steigern wäre. Die zweite Frage drängt auf empirisch gesicherte Aufschlüsse über die Bedingungen, unter denen solche unterschiedlichen Beurteilungen zustande kommen. Dabei steht nicht die Annahme im Hintergrund, eine der zwei die Urteile abgebenden Gruppen, Lehrkräfte oder Schülerinnen und Schüler, verfügte über diagnostisch valides Wissen. Der Mangel an Übereinstimmung in 80% der Urteile verweist vor allem anderen auf ein gewichtiges pädagogisches Problem. Von den für einen Großteil eines Schuljahres mit Schülerinnen und Schülern arbeitenden pädagogischen Fachkräften verfügen in der untersuchte 4n Stichprobe nur 20% über handlungsrelevantes Wissen zur emotionalen Befindlichkeit der Kinder oder Jugendlichen, die sie unterrichten. Diese Feststellung hat, wenn sie replizierbar sein sollte – was durchaus wahrscheinlich ist – erhebliche pädagogische und psychologische Bedeutung. Aus diesem Grund wird sie durch den nachstehenden Befund hervorgehoben.

24. Befund: Mit einer das Ergebnis zum Wissen der beteiligten Lehrkräfte über die emotionale Befindlichkeit ihrer Schülerinnen und Schüler von der Aussage über die Gesamtgruppe der Lehrkräfte zur individuellen Urteilsleistung hin aufschließenden Analyse wird deren interindividuelle Inhomogenität erkennbar. Eine Teilgruppe von etwa 20% aller beteiligten Lehrkräfte erreicht in Absetzung von dem geringen Wissensstand der anderen beachtenswerte Übereinstimmung mit den Selbsteinschätzungen der Kinder oder Jugendlichen in ihren Schulklassen. Die bis zu einem Spitzenwert von $r = 0,71$ reichenden Korrelationen stehen für individuell erreichte Aufklärungsraten zwischen etwa 20% bis 50% der gemeinsamen Varianz der Variablen. Demgegenüber gilt für etwa 80% der Lehrkräfte ein statistisch dem Nichtwissen gleichzusetzender Status. Eine Untergruppe in dieser überaus deutlichen Mehrheit tendiert entsprechend den für sie errechneten negativen

Korrelationen zur Verkennung und Verzerrung der emotionalen Situation von Schülerinnen und Schülern.

Dem 25. Befund kommt, wenn er replizierbar sein sollte – was durchaus wahrscheinlich ist – wegen seiner Aussagekraft über wichtige regulative Bedingungen schulischer Interaktion erhebliche pädagogische und psychologische Bedeutung zu.

Angesichts der nachgewiesenen und danach überaus unterschiedlichen Informiertheit der Lehrkräfte dieser Untersuchung über die emotionale Befindlichkeit ihrer Schülerinnen und Schüler sind empirisch begründete Hinweise auf die Bedingungen unter denen die Urteile im ESDM abgegeben werden, erwünscht und systematisch aufzusuchen.

Eine Bedingung wird durch den schulischen Leistungsstand der Kinder oder Jugendlichen gesetzt.

25. Befund: Zwischen den Urteilen der Lehrkräfte zum beobachteten Auftreten von depressiven Symptomen in ihren Schulklassen, erfasst über die 16 ESDM-Fragen, und der durch Lehrkräfte vollzogenen Zuweisung der Kinder oder Jugendlichen zu vier unterschiedlichen und auf Schule bezogenen Leistungsgruppen (Leistungsstand „sehr niedrige“, „niedrige“, „durchschnittliche“ und „überdurchschnittlich“ im Vergleich zu Mitschülern) ergab sich eine statistisch hochsignifikante linear verlaufende negative Beziehung. Dieses Ergebnis war stabil sowohl für die Beurteilung durch Lehrerinnen oder Lehrer als auch für Grundschulkindern und Gymnasiasten.

Auf einige weitere Bedingungen hinweisende Ergebnisse waren unter anderem die nachfolgend genannten.

Die über die Skala ESDM von den beteiligten Lehrkräften gegebene Urteile zum Auftretens depressiver Symptome bei ihren Schulkindern weisen sowohl auf Gesamtscoreebene als auch auf Itemebene und für beide der besuchten Schulformen – Grundschulen und Gymnasien - geschlechtsspezifische Unterschiede auf. Für die Stichprobe der Grundschulkindern wurde gefunden, dass männliche Lehrkräfte von einer signifikant höheren Belastung ihrer Schüler und Schülerinnen berichten als ihre weiblichen Kolleginnen. Für die Gymnasien fand sich eine in entgegengesetzter Richtung verlaufende geschlechtsspezifische Urteilstendenz. Vermehrt heben dort weibliche Lehrkräfte die Belastung der Kinder und Jugendlichen hervor.

Weibliche Lehrkräfte an Gymnasien berichten von einer erhöhten depressiven Belastung mit zunehmendem Alter ihrer Schüler und Schülerinnen.

Einige Hinweise lassen auch vermuten, dass die Urteile in nicht weiter quantifizierbarem Maße auch vom Geschlecht der zu beurteilenden Kinder oder Jugendlichen beeinflusst werden.

Die geschlechtsspezifische Einflüsse auf die Urteile der Lehrkräfte sind im Kapitel 12 unter 12.6.1. ausführlich erörtert.

Die über das Verfahren ESDM gewonnenen Lehrurteile variieren weiterhin sowohl bei weiblichen als auch bei männlichen Lehrkräften signifikant mit der Klassenstufe, über die sie urteilen. Die ESDM-Urteile variieren auch signifikant mit den untersuchten einzelnen Schulklassen.

Eine in ihrem Einfluss nicht näher einschätzbare Quelle der Varianz ist den Analyseergebnissen zufolge auch das von der Lehrkraft vertretene Unterrichtsfach.

26. Befund: Die von den Lehrkräften mit ESDM-Urteilen gegebenen Einschätzungen der emotionalen Befindlichkeit ihrer Schülerinnen und Schüler variieren in unterschiedlichem Ausmaß auch

- mit dem Geschlecht der Lehrkräfte,
- mit dem Geschlecht der Kinder und Jugendlichen,
- mit der Klassenstufe,
- mit dem Unterrichtsfach der Lehrkräfte,

Von speziellem Interesse könnten Items sein, für die keine geschlechtsspezifischen Unterschiede, weder für die Stichprobe der Lehrkräfte an Grundschulen noch für die an Gymnasien, auftreten. Das sind 6 Items. Davon betreffen 2 die soziale Akzeptanz von Schülerinnen oder Schülern untereinander. Die anderen 4 sind Aussagen über die schulische Bewährung der Kinder oder Jugendlichen.

Die Urteile der Lehrkräfte stehen im Zusammenhang mit besonderen von den Lehrkräften angeführten Lebensumständen der Schülerinnen und Schüler.

27. Befund: Die ESDM-Beurteilungen der Lehrkräfte variieren signifikant mit den nach ihren Angaben zu speziellen Lebensumständen der Kinder hergestellten Teilgruppen. Die drei nach Urteilen der Lehrkräfte am stärksten von depressiven Symptomen belasteten Teilgruppen von Grundschulkindern sind

- die Kinder mit Lernproblemen, die in gestörten familiären Situation lebenden Kinder und
- aufmerksamkeitsgestörte Kinder.

Unter den Gymnasiasten wurden von den Lehrkräften über ESDM die nachstehend aufgeführten Gruppen als emotional stärkere belastet eingestuft,

- die, nach Angaben der Lehrkräfte, sehr introvertierten und überempfindlichen,
- die mit Lernproblemen, sowie
- Kinder und Jugendliche aus Familien getrennter lebender

Die oben genannten Bedingungen für die signifikante Variation von ESDM Urteilen der Lehrkräfte beziehen sich auf Merkmale der Lehrkräfte, Merkmale der Kinder sowie auf weitere schulisch wirksame Bedingungen, die mit nicht bekannt gewordener Stärke und in nicht erfassten psychischen Zusammenhängen auf die Interaktionen zwischen Lehrkräften und deren Schülern und Schülerinnen einwirken. Die dargestellte Befunde beziehen sich zumeist auf den isolierten Einfluss der untersuchten Variablen auf die über den ESDM gegebene Lehrurteile. Anzunehmen ist, dass auch diese Wirkfaktoren miteinander und mit ungenannten anderen interagieren.

Die aus dieser Untersuchung mitgeteilten Ergebnisse zu der über ESDM festgestellten Erkennungsleistung von Lehrkräften im Hinblick auf die vermutlich von depressiven Symptomen belasteten Schulkindern stimmen insgesamt sehr gut mit Ergebnissen aus anderen griechischen und internationalen Studien überein. Stasinou (1995) hat beispielsweise nach einer Untersuchung mit 205 griechischen Lehrkräften über die Verhaltensauffälligkeiten ihrer Klassenkinder gefunden, dass die von den Lehrern gegebene Urteile über verhaltensproblematische Kinder in ihren Schulklassen, von verschiedenen Faktoren abhängig sind. Genannt werden das Geschlecht der Kinder, deren Schulleistungen, das Geschlecht der Lehrkraft, deren Alter und ihre pädagogische Spezialisierung, sowie die von ihnen angebotene Unterrichtsfächer.

Hinsichtlich des Einflusses des Geschlechts der Kinder auf die Urteile der Lehrkräfte stehen in die Ergebnisse der einschlägigen Literatur in Übereinstimmung mit den von uns gefunden. Insgesamt werden von Lehrkräften den Jungen mehr Probleme zugeschrieben als den Mädchen (Bibou-Nakou et al., 2001; Ekblad, 1990; Harris, Tyre & Wilkinson, 1993; Motti-Stefanidi et al., 1996; Rutter & Garmezy, 1993; Rutter, Tizard & Whitmore, 1970; Sourander et al., 1995). Dabei überwiegen Verhaltensauffälligkeiten. Für die Mädchen berichten Lehrkräfte allgemein mehr über das Auftreten emotionaler Probleme (Rutter & Garmezy, 1993). Die Ergebnisse in Bezug auf den Einfluss des Alters der Schüler auf die Lehrurteile sind die Berichte uneinheitlich. Im griechischen Raum konnten diesbezüglich weder Bibou-Nakou und Mit. (2001) noch Tsiantis und Mit. (1994) einen deutlichen Zusammenhang finden. Die Schulleistung der Kinder wird ebenfalls oft als ein Einflussfaktor auf die Urteile von Lehrkräften über die psychische Gesundheit ihrer Schulkinder identifiziert (McGee, Williams & Silva, 1995). Zuletzt zeigen Studien aus dem griechischen Raum, dass Lehrkräfte dazu tendieren, die von ihnen beobachtete emotionale Probleme und Verhaltensauffälligkeiten ihrer Schulkinder auf Faktoren außerhalb des schulischen Kontextes zurückzuführen, vor allem auf eigene Defizite des Kindes und auf ungünstige familiäre Umstände (Bibou-Nakou, 2003; Mavropoulou und Padeliadu, 2000). Andererseits scheinen Faktoren wie Disziplin der Kinder gegenüber Klassenregeln sowie deren Motivation zur guten schulischen Leistung die Urteile der Lehrkräfte zu beeinflussen (Merret & Wheldall, 1993; Trovato et al., 1992).

Stasinou (1993, s.o.) betont in der Zusammenfassung seiner Ergebnisse insbesondere für Griechenland die Notwendigkeit der Entwicklung von Programmen, die als Ziel die Informiertheit von Lehrkräften über die psychische Gesundheit ihrer Schüler und Schülerinnen haben sollten, sowie die Entwicklung einer Netzes zur Unterstützung von Schulen und Lehrkräften durch psychologische Einrichtungen, die Beratung und geeigneten Interventionen anbieten.

Die vorangegangene Übersicht zeigt im Hinblick auf die Bedingungen von Urteilen der Lehrkräfte die tendenziell hohe Übereinstimmung der Befunde aus der Studie 20034GR mit dem nationalen und internationalen Stand der einschlägigen Literatur.

13.7 Faktizität. Nutzbarkeit und potentielle Nützlichkeit der Untersuchungsergebnisse aus der Studie 2034GR

Der Begriff Faktizität fasst das in einem gemeinten Zusammenhang zu beobachtende Aufkommen von Fakten oberbegrifflich zusammen. Ein Faktum ist die Durchführung der Studie 2034GR. Ein weiteres Faktum ist der aus der Untersuchung hervorgegangene Datensatz. Schließlich sind auch die nun mit diesen Ausführungen abzuschließenden Darlegungen und Erörterungen als Bericht über die durchgeführte Untersuchung ein Faktum.

Fakten sind, solange sie bestehen, nicht mehr als das Faktische an ihnen. Der Hinweis auf eine prinzipiell nicht gesicherte Dauerhaftigkeit ist berechtigt und auch nützlich. Schon das erstgenannte Faktum, die Durchführung der Studie, ist mit ihrer jeweils aktuellen Realisierung bereits in den Prozess geraten, der die Notwendigkeit schafft, die Faktizität der vorübergehenden und vergänglichen aktuellen Ereignisse durch Dokumente oder Zeugnisse und deren Speicherung glaubhaft bleiben zu lassen. Das werden sie durch das Gespeicherte nur mehr oder minder zulänglich, weil keine Speicherung das Original erhält. All das gilt auch für das zweite hier zuvor angeführte Faktum, den durch die Untersuchung erbrachten und über sie geschaffenen Datensatz als eine spezielle Speicherung von Untersuchungsereignissen, von Antworten, die in dieser Untersuchung 2034 griechische Kinder oder Jugendliche und deren Lehrkräfte gegeben hatten. Von ihnen allen sind nur die durch die untersuchende Person beachteten Antworten als kodierte

Partikel in einem Hunderttausende davon enthaltenden Datensatz erhalten. Auch die mit diesem Bericht gegebene Darstellung der Untersuchung dient dem Versuch, eine über Sprache und die Schrift symbolisierte Übersicht zum Untersuchungsgeschehen zu geben und durch mediale Speicherung zu bewahren.

Die mit dieser Untersuchung direkt und mittelbar geschaffenen Fakten sind danach nicht mehr und auch nicht weniger als andere Fakten, nämlich sie selbst und insoweit inhaltsvalide Zeugnisse für sich selbst.

Das Faktische umschließt ungezählte Ereignisse, von denen, falls sie gespeichert wurden, Symbolisierungen erhalten bleiben. Symbolisierungen sind zu Zeichen verschlüsselte Informationen, die aus den Zeichen durch Entschlüsselung aufgeschlüsselt und verstanden werden könnten, wenn jemand, der zu entschlüsseln fähig wäre, das auch wollte. Solche Informationen sind auf Sachverhalte bezogen, die, weil sie sich neben anderem auch so denken lassen, dann als Entitäten, Subjekte, Objekte, Personen oder Gegenstände bezeichnet werden. Von manchen Gegenständen des Denkens verbleibt allerdings auch nur der Hinweis auf ihre einmal vorhandene Gegenwart. Das gilt im Rückblick auf die hier erörterte Untersuchung so auch schon für die in ihrem Lebensalltag zurückgelassenen Teilnehmer, die Schülerinnen, Schüler, Lehrer und Lehrerinnen, von deren, durch die Untersuchungsereignisse ausgelösten, persönlichen Empfindungen, Gedanken, Hoffnungen und Fragen die darzulegenden Untersuchungsbefunde nur wenig oder sogar nichts mitteilen.

Fakten als reale oder symbolisierte Gegenstände bleiben, wenn sie bleiben, was sie sind und sind so denen überlassen, die sich ihnen zuwenden und gegebenenfalls einen Aspekt ihrer Nutzbarkeit entdecken.

So könnte beispielsweise eine im Sand aufgefundene steinzeitliche Speerspitze aus Feuerstein liegen gelassen, oder in einem heimatkundlichen Museum ausgestellt oder von Kindern zum Einritzen von Geheimzeichen in Mauerwerk benutzt werden, bevor Eltern sie als gefährliches Spielzeug beschlagnahmten und dem großen Bruder übergeben würden, der dann auch diesen Stein mit gekonntem Wurf, ihn sechs, sieben mal auf dem glatten Wasserspiegel aufschlagen und im See versinken ließe. Genutzt wird, was von irgendwen als hilfreich zum Erreichen von Zielen eingeschätzt und aufgenommen wird.

Die Nutzbarkeit von Fakten entsteht erst, wenn sie von jemandem entdeckt wird, der sie, zu dem von ihm gemeinten Zweck als nutzbar ansieht und den Gegenstand damit zumindest probeweise in Gedanken nutzt.

Auf diesem Gedankenweg kann man auch zu Betrachtung wissenschaftlicher Untersuchungen gelangen, die erfolgreich realisiert wurden und Fakten geschaffen haben, die niemand nutzt. Das ereignet sich auch in der Psychologie und somit ebenfalls in der klinischen Psychologie. Die Existenz von Gegenständen als Fakten in dieser oder jener Form leitet nicht zwangsläufig zu deren Nutzung über, jede Nutzung beginnt damit, dass jemand das faktisch Verfügbare in die Ausführung seiner Handlung, was auch immer die zum Ziel haben mag, übernimmt.

Der vorstehende Exkurs ist niedergeschrieben worden, um, zum einen, belegen zu können, dass auch empirisch erarbeitetes faktisches Wissen zwar grundsätzlich nutzbar ist, aber nachweisbar nicht nur in Griechenland, aber auch dort, ungenutzt bleibt.

Nun ist aber, zum anderen, in Beantwortung der dieser Untersuchung zu Grunde liegenden Forschungsfrage, faktische Wissen zum Auftreten von depressiven Störungen bei griechischen Schulkindern erarbeitet worden. Die Grundlage dieses Wissens bilden die von Schülerinnen, Schülern und Lehrkräften aufgenommenen Informationen, die in der weiteren Untersuchungsarbeit symbolisiert und analysierten wurden, bevor sie auf einer ersten Stufe der

notwendigen Dekodierung als Ergebnisse der Datenanalysen darzustellen waren. Diese, einige Hundert umfassenden Teilergebnisse und Ergebnisse sind die Menge datennaher Fakten. Auf einer zweiten Stufe der Entschlüsselung sind die datennahen Fakten zu Befunden zusammengefasst worden, die als oberbegriffliche Bündelungen von Fakten verstanden werden. Auf einer nun hier erreichten dritten Stufe der oberbegrifflichen Repräsentation aller zuvor mitgeteilten Befunde ist deren Zusammenfassung nahe an der Umgangssprache und entsprechend verständlich. Damit wird eine wesentliche, allerdings nur eine, Voraussetzung für die nicht nur wissenschaftsinterne Kenntnisnahme des Hauptergebnisses dieser Studie mit griechischen Schulkindern geschaffen. Eine wesentliche Voraussetzung ist dies, wenn alle von der Antwort auf die Untersuchungsfrage potentiell Betroffenen auch Gelegenheit finden sollten, Kenntnis zu nehmen.

Die Betroffenen sind alle Menschen in Griechenland, deren Lebensführung mittelbar oder unmittelbar mit der körperlichen und psychischen Gesundheit griechischer Schülerinnen und Schüler verbunden ist. Das sind zum einen die Kinder und Jugendlichen selbst, deren Eltern und Verwandte, die Lehrkräfte an Schulen, Fachschulen und Universitäten, die in der Berufsausbildung tätigen Personen, die Arbeitgeber in den kleineren, mittleren und den Großbetrieben, die künftigen Arbeitskollegen und –kolleginnen, die Verwaltungen, die Armee, die Krankenhäuser, der Gesundheitsdienst, Polizei, Feuerwehr, Forschung, Luftfahrt, Schifffahrt und Landwirtschaft.

Ganz Griechenland ist in seiner Gegenwart und Zukunft auf die körperliche und psychische Gesundheit der jetzt in Griechenland aufwachsenden Kinder und Jugendlichen angewiesen und auch auf die Gesundheit der weiterhin nachwachsenden Generationen.

Alle Bewohner Griechenlands gehören zu den unmittelbar oder mittelbar Betroffenen. Dann aber sind vor allem die anzusprechen, die kraft ihrer gesellschaftlichen Position oder ihres Amtes verantwortlich für die politische Gestaltung des gesellschaftlichen Lebens in Griechenland sind.

Als Antwort auf die zu Beginn dieser Studie gestellte Forschungsfrage sind die nachstehenden Feststellungen zur Kenntnis zu nehmen:

1. Die als 2034GR bezeichnete Untersuchung weist nach, dass mit an Sicherheit grenzender Wahrscheinlichkeit depressive Symptome, depressive Episoden und depressive Störungen unter griechischen Schülerinnen und Schülern auftreten.
2. Der Anteil der von depressiven Störungen betroffenen Kinder oder Jugendlichen an der untersuchten Stichprobe von insgesamt 2034 Kindern oder Jugendlichen wird, je nach Schweregrad der individuellen Ausprägung der Störung, als mit großer Wahrscheinlichkeit zwischen 5% und 14% liegend geschätzt.
3. Mit hoher Wahrscheinlichkeit ist anzunehmen, dass für etwa $\frac{1}{4}$ der untersuchten Kinder und Jugendlichen, die zuvor genannten 5% bis 14% darin eingeschlossen, akuter Bedarf an medizinischer, psychologischer und pädagogischer Betreuung besteht.
4. Die für die untersuchte Stichprobe geltenden Feststellungen dürfen bei dem gegebenem Wissensstand als eine beste aktuell erreichbare Schätzung des Auftretens depressiver Störungen unter griechischen Schulkindern angesehen und genutzt werden.

Die vierte Feststellung wird unter Berufung auf Bortz und Döring (2002) als methodisch gerechtfertigt aufgeführt.

Dargelegt worden war zuvor, dass die Existenz faktischen Wissens nicht zwangsläufig in dessen Nutzung übergeht. Dann ist unter ausdrücklicher Bezugnahme auf die durchgeführte Untersuchung 2034GR, aber auch unter Hinweis auf andere griechische Untersuchungen zur psychischen Gesundheit von Kindern und Jugendlichen in Griechenland zu erklären, dass die Nichtbeachtung der mit dieser und mit anderen wissenschaftlichen Untersuchungen erarbeiteten

Fakten durch die oben aufgeführten Betroffenen ein Fehler wäre, der mit an Sicherheit grenzender Wahrscheinlichkeit absehbaren und schwer hinzunehmenden Schaden nach sich ziehen wird.

Zur Verdeutlichung der bestehenden Zusammenhänge kann auf vergleichbare Ereignisse verwiesen werden.

Die unbestechliche Faktizität der Ereignisse belehrt die Bewohner Griechenlands in jeder Generation aufs Neue darüber, dass die präventive Vorsorge gegenüber bedrohlichen Katastrophen, so etwa die von auch künftig zu erwartenden Erdbeben und in Hitzeperioden ausbrechenden Waldbränden notwendig, realisierbar und nützlich ist. Vertraut sind die Bewohner Griechenlands auch mit Warnungen vor aktuell sich ausbreitenden Epidemien und den präventiven und kurativen Folgen daraus.

Das alles liegt im Eigeninteresse aller Bewohner Griechenlands, die ihre oder die Gesundheit und Unversehrtheit ihrer Lieben bedroht sehen.

Weit weniger verbreitet ist in Griechenland auf Grund der gesellschaftlichen Entwicklung im 20. Jahrhundert die Einsicht in die Bedeutung der psychischen Gesundheit für Lebensführung und körperliche Leistungsfähigkeit. Unmissverständlich sind beispielsweise die Warnungen der WHO vor dem Anwachsen des Auftretens depressiver Störungen und deren die Gesellschaft bedrohenden Schädigungspotential.

So wie die Warnungen vor drohenden Erdbeben, Waldbränden, Vogelgrippe, Tuberkulose oder Aids beachtet und mit der Frage nach notwendig werdenden und vorsorglichen Maßnahmen und Hilfsmöglichkeiten aufwerfen, ist die griechische Gesellschaft gut beraten, auch Warnungen wegen des akuten Auftretens psychischer Störungen als Anlass zum Überdenken und zur Realisierung von Gegenmaßnahmen zu verstehen. Das ist in Griechenland noch keine gesellschaftlich akzeptierte und verbreitete Selbstverständlichkeit. Demnach besteht, was alle gesellschaftliche Verantwortung tragende Persönlichkeiten beunruhigen sollte, die Gefahr, dass Warnungen vor aktuell als bestehend erkannten Gefährdung der psychischen Gesundheit von Einwohnern Griechenlands nicht gehört, nicht gelesen und nicht beachtet werden. Die Gefahr, einem solchen Risiko zu erliegen ist aus zwei Gründen groß. Zum einen schleichen sich die Symptome psychischer Störungen leicht unbemerkt in die Wahrnehmung ein, erscheinen als alltäglich und werden erst bemerkt, wenn die Episoden kumulieren und Störungen des Zusammenlebens unerträglich scheinen und selbst dann noch verkannt werden können. Dazu trägt zum anderen die defizitäre Bewusstheit von den Eigenheiten, Merkmalen und Folgen psychischer Störungen bei.

Die Mehrheit der griechischen Bevölkerung könnte relativ ungeübt im Umgang mit psychischen Störungen sein, die nur deswegen als nicht existent gelten würden, weil sie alltäglich verkannt werden.

Die Untersuchung 2034GR hat Fakten erarbeitet, die vor dem aktuellen Auftreten depressiver Störungen unter Schülerinnen und Schülern warnen lassen, weil sie die Gesundheit von Kindern und Jugendlichen belasten und damit wiederum deren persönliche Entwicklung beeinträchtigen. Warum depressive Störungen den Entwicklungsgang und potentiell auch den gesamten Lebenslauf von Kindern, Jugendlichen und Erwachsenen bis hin zum Scheitern im alltäglichen Leben zu schädigen vermögen, ist in der einschlägigen Literatur umfassend beschrieben und kann dort nachgelesen werden. Die Ergebnisse aus der hier zuvor umfassend dargestellten Untersuchung weisen nach, dass unter griechischen Schulkindern wegen des Auftretens depressiver Störungen ein akuter Notstand besteht, der sofortige gesellschaftliche Anstrengungen

um dessen Eindämmung und Überwindung verlangt. Die eingehende und Erfolg versprechende Entwicklung einer landesweit einsetzbaren Strategie und die Planung der daraufhin zu realisierenden Maßnahmen ist unter Rückgriff auf den internationalen Forschungsstand und unter Nutzung nationaler Forschung unmittelbar möglich. Der Notstand, auf den verwiesen wird, kann durch sofort realisierbare Maßnahmen eingedämmt werden.

Der weiter oben vorangestellte Exkurs über Aspekte vorhandenen faktischen Wissens, stellte unter anderem dar, dass die Nutzbarkeit solchen Wissens nicht zwangsläufig in Nutzung übergeht und so das Erreichen von Nutzen und Nützlichkeit sichert. Daraus ist angesichts des nachgewiesenen Notstandes unter Schulkindern die Folgerung zu ziehen, dass für die Aufnahme wirkungsvoller gesellschaftlicher Anstrengungen zur Eindämmung des Auftretens depressiver und anderer psychischer Störungen gearbeitet werden muss. Falls Psychiater und Psychologen die einzigen Berufsgruppen sein sollten, die diese Dringlichkeit erkennen, ist das Bemühen um die Aktivierung weiterer gesellschaftlicher Gruppen zunächst auch deren Aufgabe. Insoweit ist diese Folgerung auch ein Ergebnis der durch die Untersuchung 2034GR vollzogenen Befassung mit dem Auftreten depressiver Störungen unter griechischen Schulkindern.

Die Untersuchungsarbeit mit der Studie 2034GR setzte an einer Frage, der nach dem mutmaßlichen Auftreten depressiver Episoden und Störungen unter griechischen Schulkindern. Mit der Beantwortung der Ausgangsfrage werden neue Fragen aufgeworfen. So endet die an einer Frage aufgenommene Untersuchungsarbeit nach deren Beantwortung mit daraufhin als dringlich zu erkennenden neuen Fragen, die Ansätze für weiterführende Arbeiten markieren. Die Zahl solcher Fragen vergrößert sich im Nachdenken über sie und nötigt zu dem Versuch, deren Aufzählung in eine strukturierte Ordnung zu überführen. Aus der werden drei sequentiell-hierarchische Handlungs- oder Entscheidungsbäume ausgewählt, die, im Interesse des notwendigen Vorgehens gegen depressive Störungen unter Schulkindern, für Handlungsbereiche stehen, in denen eine Aufschiebung der Realisierung notwendiger Maßnahmen nicht toleriert werden sollte. Da Handlungen von den für sie antizipierten Zielen her reguliert werden, steht am Ursprung jedes dieser Handlungsbäume die Frage nach dem übergeordneten Leitziel für den angesprochenen Handlungsbereich. Über die Antwort darauf ergeben sich die jeweils möglichen Ausdifferenzierungen der weiteren Handlungsstruktur.

Zu den anzusprechenden drei Handlungsbereichen gehört erstens der für die gesundheitliche und psychosoziale Versorgung der in Griechenland lebenden Kinder und Jugendlichen verantwortliche Teil der Legislative und der Exekutive des griechischen Staates, vertreten durch die jeweiligen Entscheidungsträger oder Entscheidungsträgerinnen. Der zweite Bereich ist das griechische Schulwesen, ebenfalls vertreten durch entscheidungsbefugte Persönlichkeiten. Der dritte Bereich ist die internationale Psychologie.

An die für den staatlichen Bereich der gesundheitlichen und psychosozialen Versorgung der in Griechenland lebenden Kinder und Jugendlichen entscheidungsberechtigte Persönlichkeit wird die folgende Frage gerichtet: „Mit welcher Zielsetzung soll der staatliche Bereich, den Sie vertreten, auf den zur Kenntnis gegebenen Untersuchungsbericht über das Auftreten depressiver Störungen unter Schülerinnen und Schülern in griechischen Schulen reagieren?“

Der für den Bereich der in Griechenland arbeitenden Schulen zu Entscheidungen berechtigten Persönlichkeit wird die nachstehende Frage gestellt: „Mit welcher aktuellen Zielsetzung wird das staatliche Schulwesen versuchen, dem Auftreten depressiver Störungen unter Schülerinnen und Schülern in griechischen Schulen entgegen zu wirken?“

Da sowohl für den ersten als auch für den zweiten Bereich die Verteilung der Zuständigkeiten auf mehrere Persönlichkeiten anzunehmen ist, sind die angeführten Fragen entsprechend zu wiederholen.

Für den Bereich der Psychologie wird die Frage aufgeworfen, welche Steigerungen der bislang erreichten und oft nicht befriedigenden Wirksamkeit präventiver und kurativer Programme zur Versorgung regionaler oder lokaler Populationen von durch depressive Störungen bedrohter oder schon betroffener Schülerinnen und Schüler unterschiedlicher Altersstufen auf welche Weise erreicht werden könnten. Dem Anraten zur vermehrten Zuwendung zu dieser Problematik und der Intensivierung und Ausdehnung des entsprechenden Forschungs- und Erprobungsaufwandes sollte gefolgt werden, weil die in zahlreichen Ländern, so auch in Griechenland, bestehenden Systeme der medizinischen und psychosozialen Versorgung depressiv gestörter Kinder und Jugendlicher weder für angemessene Präventionen noch für die kurative Versorgung hinreichend leistungsfähig sind. Dieser Zustand wird sich mittelfristig, aller Voraussicht nach, wegen der defizitären staatlichen Haushalte, kaum ändern. Danach ist der für den Bereich der Psychologie unter Kenntnisnahme der epidemiologischen Zahlen zum Auftreten von depressiven Störungen unter Schulkindern zu beschreibende dringliche Handlungsbedarf nicht auf die Suche nach gegenüber dem gegenwärtigen Stand wirksameren Verfahren und Strategien zu beschränken, viel mehr liegt das Auffinden angemessener ökonomischer Interventionen auch im existentiellen Interesse aller und besonders der in Vergleichsweise wirtschaftlich schwachen Staaten und Regionen lebenden Kinder und Jugendlichen.

Literaturverzeichnis

- Achenbach, T.M. (1991a). *Manual for the Teacher's Report Form and 1991 Profile*. Burlington: University of Vermont, Department of Psychiatry.
- Achenbach, T.M. (1991b). *Manual for the Youth Self-Report and Profile*. Burlington. V.T.: University of Vermont, Department of Psychiatry.
- Achenbach, T.M. & Edelbrock, C. (1983). *Manual for the Child Behavior Checklist and Revised Child Behavior Profile*. Burlington: University of Vermont, Department of Psychiatry.
- Achenbach, T.M. & Edelbrock, C. (1987). *Manual for the Youth Self Report and Profile*. Burlington: University of Vermont, Department of Psychiatry.
- Achenbach, T.M., McConaughy, S.H. & Howell, C.T. (1987). Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*, 213-232.
- Achermann, D. (1995). *Belastungsbewältigung und Depression bei Jugendlichen unter den erweiterten Bedeutungsaspekten von Gesundheit und Angst*. Unveröff. Diplomarbeit, Karl-Franzens-Universität Graz.
- Akiskal, H.S. (1990). Toward a clinical understanding of the relationship of anxiety and depressive disorders. In J.D. Maser & C.R. Cloninger (Eds.), *Comorbidity of mood and anxiety disorders* (597-609). Washington: American Psychiatric Press.
- Almqvist, F., Kumpulainen, K., Ikäheimo, K., Linna, S.-L., Henttonen, I., Huikko, E., Tuompo-Johansson, E., Aronen, E., Puura, K., Piha, J., Tamminen, T., Räsänen, E. & Moilanen, I. (1999). Behavioural and emotional symptoms in 8-9-year-old children. *European Child and Adolescent Psychiatry*, *8*, (suppl. 4), IV/7-IV/16.
- American Psychiatric Association (1990). *Diagnostic and statistical manual of mental disorders*. 3th edition. Washington, DC: American Psychiatric Association.
- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of mental disorders*. 4th ed. (DSM IV). Washington, DC: American Psychiatric Association.
- Amtliches Statistisches Handbuch für Griechenland (Εθνική Στατιστική Υπηρεσία). *Προσωρινά Αποτελέσματα Απογραφής Πληθυσμού Έτους 2001. Πραγματικός Πληθυσμός ανά Δήμο ή Κοινότητα*. Εθνική Στατιστική Υπηρεσία Ελλάδος. Αθήνα, 2001.
- Anderson, J.C., Williams, S., McGee, R. & Silva, P.A. (1987). DSM-III disorders in preadolescent children: prevalence in a large sample from the general population. *Archives of General Psychiatry*, *44*, 69-76.
- Angold, A. (1988). Childhood and adolescent depression. I. Epidemiological and aetiological aspects. *British Journal of Psychiatry*, *152*, 601-617.
- Angold, A., Costello, E.J. & Pickles, A. (1987). *The development of a questionnaire for use in epidemiological studies of depression in children and adolescents*. Unpublished manuscript, London University.
- Angold, A., Erkanli, A. & Rutter, M. (1996). Precision, reliability and accuracy in the dating of symptom onsets in child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, *37*, 657-664.
- Angold, A., Prendergast, M., Cox, A., Harrington, R., Sumonoff, A. & Rutter, M. (1995). The Child and Adolescent Psychiatric Assessment (CAPA). *Psychological Medicine*, *25*, 739-753.
- Argyriadou, S., Melissopoulou, H., Krania, E., Karagiannidou, A., Vlachonicolis, I. & Lionis, C. (2001). Dementia and depression: Two frequent disorders of the aged in primary health care in Greece. *Family Practice*, *18*, 87-91.
- Asarnow, J.R. & Carlson, G.A. (1985). Depression Self-Rating Scale: utility with child psychiatric inpatients. *Journal of Consulting and Clinical Psychology*, *53*, 491-499.
- Bahls, S. (2002). Epidemiology of depressive symptoms in adolescents of a public school in Curitiba, Brazil. *Revista Brasileira de Psiquiatria*, *24*, 63-67.

- Barrett, M.L., Bhate, S.R., Berney, T.P., Famuyima, O.O., Fundudis, T., Kolvin, I. & Tyrer, S. (1991). Diagnosing childhood depression. Who should be interviewed- parent or child? The Newcastle Child Depression Project. *British Journal of Psychiatry*, 159 (suppl. 11), 22-27.
- Beck, A.T. (1976). *Cognitive therapy and emotional disorders*. New York: International Universities Press.
- Beck, A. T. (1978). *Depression inventory*. Philadelphia, PA: Center for Cognitive Therapy.
- Becker, A., Hagenberg, N., Roessner, V., Woerner, W. & Rothenberger, A. (2004). Evaluation of the self-reported SDQ in a clinical setting: Do self-reports tell us more than ratings by adult informants? *European Child and Adolescent Psychiatry*, 13, II/17-II/24.
- Bedford, A., Foulds, G.A. & Sheffield, B.F. (1976). A new personal disturbance scale (DSSI /sAD). *British Journal of Social and Clinical Psychology*, 15, 387-394.
- Beratis, S., Gourzis, P. & Gabriel, I. (1996). Psychological factors in the development of mood disorders with a seasonal pattern. *Psychopathology*, 29, 331-339.
- Besevegis, E. (1996). *Strategies in coping with problems scale*. Psychometric Laboratory, University of Athens.
- Bibou-Nakou, I. (2003). *Ο ρόλος της οικογένειας στο λόγο των εκπαιδευτικών για τη "δύσκολη συμπεριφορά" των παιδιών στο σχολείο*. Επιστημονική Επιτηρίδα - Σχολή Ψυχολογίας V, 439-467.
- Bibou-Nakou, I., Kioseoglou, G. & Stogiannidou, A. (2001). Δυνατότητες και δυσκολίες παιδιών σχολικής ηλικίας στο οικογενειακό και σχολικό πλαίσιο. *Ψυχολογία*, 8, 4, 506-525.
- Bibou-Nakou, I. & Kioseoglou, G. (2001a). Προσδιορισμός καταθλιπτικών συμπτωμάτων σε παιδιά και εφήβους σχολικής ηλικίας. I. Δεδομένα από τη χρήση του Ερωτηματολογίου Διάθεσης και Συναισθημάτων. *Παιδί και Έφηβος. Ψυχική Υγεία και Ψυχοπαθολογία*, 3, 1, 71-95.
- Bibou-Nakou, I. & Kioseoglou, G. (2001b). Προσδιορισμός καταθλιπτικών συμπτωμάτων σε παιδιά και εφήβους σχολικής ηλικίας. II. Παράγοντες φύλου και ηλικίας στην αξιολόγηση παιδιών και εφήβων με σημαντική παρουσία καταθλιπτικών συμπτωμάτων. *Παιδί και Έφηβος. Ψυχική Υγεία και Ψυχοπαθολογία*, 3, 1, 96-109.
- Bird, H.R., Canino, G., Rubio-Stipec, M., Gould, M.S., Ribera, J., Sesman, M., Woodbury, M., Huertas-Goldman, S., Pagan, A., Sanchez-Lacay, A. & Moscoso, M. (1988). Estimates of the prevalence of childhood maladjustment in a community survey in Puerto Rico: the use of combined measures. *Archives of General Psychiatry*, 45, 1120-1126.
- Block, J. (1978). *The Q-short Method in Personality Assessment and Psychiatric Research*. Consulting Psychologists, Palo Alto, CA.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5. Auflage). Berlin: Springer.
- Bortz, J. & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (3. Auflage). Berlin: Springer.
- Bortz, J. & Lienert, G. A. (1998). *Kurzgefasste Statistik für die Klinische Forschung. Ein praktischer Leitfaden für die Analyse kleiner Stichproben*. Heidelberg: Springer.
- Bowi, U. & Krampe, R. (1987). *Eine deutsche Fassung des Children's Depressions Inventory (Kovacs, 1983)*. Abteilung für Kinder- und Jugendpsychiatrie, Psychiatrische Klinik der Heinrich-Heine-Universität Düsseldorf, 1987.
- Brosius, G. & Brosius, F. (1998). *SPSS. Base system und Professional Statistics* (2. Unver. Nachdruck). Bonn, Albany (u.a.): International Thomson Publishing.
- Buchmann, C. (2000). *Emotionale Befindlichkeit und Verhaltensauffälligkeiten im frühen Schulalter aus der Perspektive von Kindern und ihren Lehrern: Erprobung eines Bildverfahrens*. Unveröff. Diplomarbeit, Karl-Franzens-Universität Gratz.
- Bühler, K. (1934). *Sprachtheorie: die Darstellungsfunktion der Sprache*. Jena: Fischer.
- Campbell, D.T. & Fiske, D.W. (1959). Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin*, 103, 276-279.
- Caine, T.M., Foulds, G.A. & Hope, K. (1967). *Manual of the Hostility and Directions of Hostility Questionnaire*. London: University of London Press.
- Canals, J., Domènech, E., Carbajo, G. & Bladé, J. (1977). Prevalence of DSM-III-R and ICD-10 psychiatric disorders in a Spanish population of 18-years-old. *Acta Psychiatrica Scandinavica*, 96, 287-294.

- Canals, J., Domènech-Llaberia, E., Fernández-Ballart, J. & Martí-Henneberg, C. (2002). Predictors of depression at eighteen. A 7-year follow-up study in a Spanish nonclinical population. *European Child and Adolescent Psychiatry*, 11, 226-233.
- Canals, J., Martí-Henneberg, C., Fernández-Ballart, J. & Domènech, E. (1995). A longitudinal study of depression in an urban Spanish pubertal population. *European Child and Adolescent Psychiatry*, 4, 102-111.
- Carey, M.P., Faulstich, M.E., Gresham, F.M., Ruggiero, L. & Enyart, P. (1987). Children's Depression Inventory: construct and discriminant validity across clinical and nonreferred (control) populations. *Journal of Consulting and Clinical Psychology*, 55, 755-761.
- Carlson, G.A. & Cantwell, D.P. (1980a). A survey of depressive symptoms, syndrome, and disorder in a child psychiatric population. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 21, 19-25.
- Carlson, G.A. & Cantwell, D.P. (1980b). Unmasking masked depression in children and adolescents. *American Journal of Psychiatry*, 137, 445-449.
- Casanovas, M. (2003). *Aspekte erlebter Ethnizität und das Auftreten depressiver Symptome bei spanischen Kindern und Jugendlichen in der Bundesrepublik Deutschland*. Unveröff. Dissertation, Freie Universität Berlin.
- Chan, D.W. (1997). Depressive symptoms and perceived competence among Chinese secondary school students in Hong Kong. *Journal of Youth & Adolescence*, 26, 303-319.
- Chen, X., Rubin, K.H. & Li, B. (1995). Depressed mood in Chinese Children: Relations With School Performance and Family Environment. *Journal of Consulting and Clinical Psychology*, 63, 938-947.
- Christopoulou, A. (1996). Θέματα ψυχικής υγείας και προσαρμογής των εφήβων στην Ελλάδα. *Ψυχολογία*, 3, 2, 101-109.
- Compass, B.E. & Hammen, C.L. (1996). Child and adolescent depression: Covariation and comorbidity in development. In R.G. Haggerty, L.R. Sherrod, N. Garmezy & M. Rutter (Eds.), *Stress, risk and resilience in children and adolescents* (pp. 225-267). Cambridge University Press: Cambridge.
- Costello, E. & Angold, A. (1988). Scales to assess child and adolescent depression. Checklists, screens and nets. *Journal of the American Academy of Child and Adolescent Psychiatry*, 27, 726-737.
- Costello, E., Angold, A., Burns, B., Erkanli, A., Stangl, D. & Tweed, D. (1996). The Great Smoky Mountains study of youth: functional impairment and severe emotional disturbance. *Archives of General Psychiatry*, 53, 1137-1143.
- Costello, E., Edelbrock, C., Costello, E.J. (1985). Validity of the NIMH diagnostic interview schedule for children: A comparison between psychiatric and pediatric referrals. *Journal of Abnormal Child Psychology*, 13, 579-595.
- Craighead, W.E., Curry, J.F. & Ilardi, S.S. (1995). Relationship of Children's Depression Inventory factors to major depression among adolescents. *Psychological Assessment*, 7, 171-176.
- Craighead, W. E., Smucker, M. R., Craighead, L. W. & Ilardi, S. S. (1998). Factor analysis of the Children's Depression Inventory in a community sample. *Psychological Assessment*, 10, 156-162.
- Cooper, P.J. & Goodyer, I. (1993). *A community study of depression in adolescent girls. I: Estimates of symptom and syndrom prevalence*. *British Journal of Psychiatry*, 163, 369-374.
- Cytryn, L. & McNew, D.H. (1972). Proposed classification of childhood depression. *American Journal of Psychiatry*, 129, 149-155.
- Cytryn, L., McKnew, D.H. & Bunney, W.E. (1980). Diagnosis of depression in children: A reassessment. *American Journal of Psychiatry*, 137, 22-25.
- del Barrio, V. (1989). *Epidemiología de la depresión infantil en Valencia*. Conselleria de Sanitat i Consum.
- Dikeos, D.G., Papadimitriou, G.N., Avramopoulos, D., Karadima, G., Daskalopoulou, E.G., Souery, D., Mendlewicz, J., Vassilopoulos, D. & Stefanis, C.N. (1999). Association between the dopamine D3 receptor gene locus (DRD3) and unipolar affective disorder. *Psychiatric Genetics*, 9, 189-195.
- Dilling, H., Mombour, W. & Schmidt, M.H. (Hrsg.) (2000). *Internationale Klassifikation psychischer Störungen ICD-10, Kapitel V (F). Klinisch-diagnostische Leitlinien* (4. korr. und erg. Auflage). Bern: Huber.
- Dimitriou, E.C. (1996). Depressive symptomatology in schizophrenia. *Psychiatriki*, 7, 261-270.

- Dimitriou, E.C. & Dimitriou, C.E. (1998). Buspirone augmentation of antidepressant therapy. *Journal of Clinical Psychopharmacology*, 18, 465-469.
- Doerfler, L.A., Felner, R.D., Rowlison, R.T., Raley, P.A. & Evans, E. (1988). Depression in children and adolescent: a comparative analysis of the utility and construct validity of two assessment measures. *Journal of Consulting and Clinical Psychology*, 56, 769-772.
- Domènech, E., Canals, J. & Fernandez-Ballert, J. (1992). Suicidal ideation among Spanish schoolchildren: a three-year follow up study of pubertal population. *Personality and Individual Differences*, 13, 1055-1057.
- Domènech, E., Monreal, P. & Ezpeleta, L. (1985). "Escala de Depresión para Maestros: ESDM." Universitat Autònoma de Barcelona. Unitat de Psicopatologia Infantil. 08193 Bellaterra, Barcelona. Spain (Unpublished manuskript).
- Domènech, E., & Polaino-Lorainte, A. (1990). La escala ESDM como instrumento adicional en el diagnóstico de la depresión infantil. *Rev. Psiquiatria Fac. Med. Barcelona*, 17, 3, 105-113, 235.
- Domènech, E., & Polaino, A. (1990). *Epidemiología de la depresión infantil*. Barcelona: Espaxs S.A.
- Domènech, E., Canals, J., Fernández-Billart, J. & Martí-Henneberg, C. (1991). Epidemiología de la depresión Mayor a la Pubertad: Resultats de tres anys de seguiment. *Rev. R. Acad. Med. Catalunya*, 7, 35-40.
- DuBois, D. L. Felner, R. D., Bartels, C. L. & Silverman, M. M. (1995). Stability of self-reported depressive symptoms in a community sample of children and adolescents. *Journal of Clinical Child Psychology*, 24, 386-396.
- Doumanis, M. (1983). Mothering in Greece: From collectivism to individualism. In R. Shaffer (Ed.), *Behavioral development: A series of monographs*. London: Academic Press.
- Economou, M. & Angelopoulos, N. (1989). Dysthymic symptoms, hostility and scholastic achievement in a group of high-school children. *Educational Psychology*, 9, 4, 331-337.
- Edelbrock, C., Costello, A.J., Dulkan, M.K., Conover, N.C. & Kalas, R. (1986). Parent-child agreement on child psychiatric symptoms assessed via structured interviews. *Journal of Child Psychology and Psychiatry*, 27, 181-190.
- Ekblad, S. (1990). The Children's Behaviour Questionnaire for completion by parents and teachers in a Chinese sample. *Journal of Child Psychology and Psychiatry*, 31, 775-791.
- Elander, J. & Rutter, M. (1996). Use and development of the Rutter parents' and teachers' scales. *International Journal of Methods in Psychiatric Research*, 6, 63-78.
- Engel, E. (1994). *Depressionsspezifische Attributionsstile bei Kindern unter besonderer Berücksichtigung der persönlichen Bedeutsamkeit negativer Ereignisse*. Unveröff. Diplomarbeit, Karl-Franzens-Universität Graz.
- Epstein, S. (1981). Entwurf einer integrativen Persönlichkeitstheorie. In Philipp. S. H. (Ed.), *Selbstkonzept-Forschung. Probleme, Befunde, Perspektiven*. Klett-Cotta.
- Essau, C.A. (2002). *Depression bei Kindern und Jugendlichen*. München: Reinhardt.
- Essau, C.A., Karpinski, N.A., Peterman, F., Conradt, J. (1998). Häufigkeit, Komorbidität und psychosoziale Beeinträchtigung von Depressiven Störungen bei Jugendlichen: Ergebnisse der Bremer Jugendstudie. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie*, 46, 316-329.
- Eysenck, H.J. (1953). *The structure of human personality*. London: Methuen.
- Ezpeleta, L. (1987). *Instrumentos de evaluación en epidemiología de la depresión infantil*. Dissertation. Universitat Autònoma de Barcelona. Departamento de Psicología de la Salud.
- Fan, X. (1998). Item Response Theory and Classical Test Theory: An empirical comparison of their Item/Person Statistics. *Educational and Psychological Measurement*, 58, 357-381.
- Fichter, M.M., Xepapadakis, F., Quadflieg, N., Georgopoulou, E. & Fthenakis, W. (2004). A comparative study of psychopathology in Greek adolescents in Germany and in Greece in 1980 and 1998 – 18 years apart. *European Archive of Psychiatry and Clinical Neurosciences*, 254, 27-35.
- Fleming, J.E., Offord, D.R. & Boyle, M.H. (1989). Prevalence of childhood and adolescent depression in the community: Ontario Child Health Study. *British Journal of Psychiatry*, 155, 647-654.
- Fotiou, F., Fountoulakis, K.N., Iacovides, A. & Kaprinis, G. (2003). Pattern-reversed visual evoked potentials in subtypes of major depression. *Psychiatry Research*, 118, 259-271.
- Foulds, G.A. (1965). *Personality and Personal Illness*. London: Tavistock.

- Freehan, M., McGee, R., Raja, S.N. & Williams, S.M. (1994). DSM-III-R disorders in New Zealand 18-years-old. *Australian and New Zealand Journal of Psychiatry*, 28, 87-99.
- Friedman, R. J. & Butler, L. F. (1979). *Development and evaluation of a test battery to assess childhood depression*. Final report to Health and Welfare, Canada, for Project 606-1533-44.
- Frigerio, A., Pesenti, S., Molteni, M., Snider, J. & Battaglia, M. (2001). Depressive symptoms as measured by the CDI in a population of northern Italian children. *European Psychiatry*, 16, 33-37.
- Fristad, M.A., Emery, B.L. & Beck, S.J. (1997). Use and abuse of the Children's Depression Inventory. *Journal of Consulting and Clinical Psychology*, 65, 699-702.
- Fry, P.S. (1983). Process measures of problem and non-problem children's classroom behaviour variables. *British Journal of Educational Psychology*, 53, 79-88.
- Fuchs, P. (1995). *Depression im Kindesalter: Eine clusteranalytische Untersuchung von Kindern mit und ohne Separationserlebnis*. Unveröff. Diplomarbeit, Karl-Franzens-Universität Graz.
- Fountoulakis, K.N., Karamouzis, M., Iacovides, A., Nimatoudis, J., Diakogiannis, J., Kaprinis, G., Demitriadou, A. & Bech, P. (2001). Morning and evening plasma melatonin and dexamethasone suppression test in patients with nonseasonal major depressive disorder from Northern Greece (latitude 40-41.5 deg.). *Neuropsychobiology*, 44, 113-117.
- Fundudis, T., Berney, T.P., Kolvin, I., Famuyiwa, O.O., Barrett, L., Bhate, S. & Tyrer, S.P. (1991). Reliability and validity of two self-rating scales in the assessment of childhood depression. *British Journal of Psychiatry* 159 (suppl. 11), 36-40.
- Garrison, C.Z., Jackson, K.L., Marsteller, F., McKeown, R. & Addy, C. (1990). A longitudinal study of depressive symptomatology in young adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 29, 581-585.
- Garyfallos, G., Adamopoulou, A., Karastergiou, A., Voikli, M., Sotiropoulou, A., Donias, S., Giouzepas, J. & Paraschos, A. (1999). Personality disorders in dysthymia and major depression. *Acta Psychiatrica Scandinavica*, 99, 332-340.
- Garyfallos, G., Adamopoulou, A., Voikli, M. & Saitis, M. (1994). DSM-III-R personality disorders among patients with depressive and/or anxiety disorders. *Journal of Personality Disorders*, 8, 320-332.
- Gittelman-Klein, R. (1977). Definitional and methodological issues concerning depressive illness in children. In J.G. Schulterbrandt and A. Raskin (Eds.), *Depression in Childhood: Diagnosis, Treatment and Conceptual Models*. New York: Raven Press.
- Glaser, K. (1968). Masked depression in children and adolescents. *Annual Progress in Child Psychiatry and Child Development*, 1, 345-355.
- Goldberg, D.P. & Hillier, V.F. (1979). A scaled version of the General Health Questionnaire. *Psychological Medicine*, 9, 129-145.
- Goodman, R. (1994). A modified version of the Rutter Parent Questionnaire including extra items on children's strengths: A research note. *Journal of Child Psychology and Psychiatry*, 35, 8, 1483-1494.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 28, 581-586.
- Goodman, R. (1999). The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry*, 40, 791-801.
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire (SDQ). *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 1337-1345.
- Goodman, R., Ford, T., Richards, H., Gadward, R. & Meltzer, H. (2000). The development and well-being assessment: Description and initial validation of integrated assessment of child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, 41, 5, 645-655.
- Goodman, R., Ford, T., Simmons, H., Gatward, R. & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry*, 177, 534-539.
- Goodman, R., Meltzer, H. & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *European Child and Adolescent Psychiatry*, 7, 125-130.

- Goodman, R., Renfrew, D. & Mullick, M. (2000). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child and Adolescent Psychiatry*, 9, 129-134.
- Goodman, R. & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? *Journal of Abnormal Child Psychology*, 27, 17-24.
- Goodyer, I.M. & Cooper, P.J. (1993). A community study of depression in adolescents girls. II. The clinical features of identified disorder. *British Journal of Psychiatry*, 163, 374-380.
- Gotlib, I. H., Lewinsohn, P. M. & Seeley, J. R. (1995). Symptoms versus a diagnosis of depression: differences in psychosocial functioning. *Journal of Consulting and Clinical Psychology*, 63, 90-100.
- Gournellis, R., Lykouras, L., Fortos, A., Oulis, P., Roubos, V. & Christodoulou, G.N. (2001). Psychotic (delusional) major depression in late life: A clinical study. *International Journal of Geriatric Psychiatry*, 16, 1085-1091.
- Graham, P. J. (1980). Depressive disorders in children – A re-consideration. *Acta Paedopsychiatrica*, 46, 285-296.
- Gratz, A. (1994). *Attributionsstil und Depressivität im Kindesalter: Eine empirische Untersuchung unter besonderer Berücksichtigung der Kontrollattribution*. Unveröff. Diplomarbeit, Karl-Franzens-Universität Graz.
- Green, B. J. (1980). Depression in early adolescence: An exploratory investigation of its frequency, intensity, and correlates (Doctoral dissertation. Pennsylvania State University). *Dissertation Abstracts International*, 41, 3890-B.
- Groen, G. & Petermann, F. (2002). *Depressive Kinder und Jugendlichen*. Göttingen: Hogrefe.
- Groen, G., Scheithauer, H., Essau, C.A. & Petermann, F. (1997). Epidemiologie depressiver Störungen im Kindes- und Jugendalter: Eine kritische Übersicht. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie*, 45, 115-144.
- Hacker, W. (1986). *Arbeitspsychologie. Psychische Regulation von Arbeitstätigkeiten*. Bern: Huber.
- Hafner, S. (2001). *Stressbewältigung und Depressivität im Kindesalter (SSK und DTK)*. Unveröff. Diplomarbeit, Karl-Franzens-Universität Graz.
- Haley, G., Fine, S., Marriage, K., Moretti, M. & Freeman, R. (1985). Cognitive bias and depression in psychiatrically disturbed children and adolescents. *Journal of Consulting and Clinical Psychology*, 53, 535-537.
- Handford, H.A., Mattison, R.E., Humphrey, F.J. & McLaughlin, R.E. (1986). Depressive syndrome in children entering a residential school subsequent to parent death, divorce, or separation. *Journal of the American Academy of Child Psychiatry*, 25, 409-414.
- Harris, J., Tyre, C. & Wilkinson, C. (1993). Using the Child Behavior Checklist in ordinary primary schools. *British Journal of Educational Psychology*, 63, 245-260.
- Harkness, S. & Super, C.M. (1990). Culture and Psychopathology. In M. Lewis & S.M. Miller (Eds). *Handbook of Developmental Psychopathology*, (pp. 41-52). New York: Plenum Press.
- Hautzinger, M., Stark, W. & Treiber, R. (1989). *Kognitive Verhaltenstherapie bei Depressionen*. München: Weinheim.
- Hecht, D.B., Inderbitzen, H.M. & Bukowski, A.L. (1998). The relationship between peer status and depressive symptoms in children and adolescents. *Journal of Abnormal Child Psychology*, 26, 153-160.
- Helsel, W.J. & Matson, J.L. (1984). The assessment of depression in children: the internal structure of the Child Depression Inventory (CDI). *Behavioural Research and Therapy*, 22, 289-298.
- Herjanick, B. & Reich, W. (1982). Development of a structured psychiatric interview for children: Agreement between child and parent on individual symptoms. *Journal of Abnormal Child Psychology*, 10, 307-324.
- Hershberg, S.G., Carlson, G.A., Cantwell, D.P. & Strober, M. (1982). Anxiety and depressive disorders in psychiatrically disturbed children. *Journal of Clinical Psychiatry*, 43, 358-361.
- Hodges, K. (1990). Depression and anxiety in children: a comparison of self-report questionnaires to clinical interview. *Psychological assessment*, 2, 376-381.

- Hodges, K., Kline, J., Stern, L., Cytryn, L. & McKnew, D. (1982). The development of a child interview for research and clinical use. *Journal of Abnormal Child Psychology*, 10, 173-189.
- Hodges, K., McKnew, D., Cytryn, L., Stern, L., & Kline, J. (1982). The Child Assessment Schedule (CAS) Diagnostic Interview: A report on reliability and validity. *Journal of the American Academy of Child Psychiatry*, 21, 468-473.
- Hoffmann, V. (1991). *Die Entwicklung depressiver Reaktionen in Kindheit und Jugend: Eine entwicklungspsychopathologische Längsschnittuntersuchung*. Stuttgart: Klett-Cotta.
- Hoffmann, K.B., Cole, D.A., Martin, J.M., Tram, J. & Seroczynski, A.D. (2000). Are the discrepancies between self- and others' appraisals of competence predictive or reflective of depressive symptoms in children and adolescents: a longitudinal study, part II. *Journal of Abnormal Psychology*, 109, 651-662.
- Hörmann, H. (1967). *Psychologie der Sprache*. Berlin, Heidelberg: Springer.
- Iacovides, A., Fountoulakis, K.N. & Ierodiakonou, C. (1999). Is it possible to predict burnout in nursing staff and its possible progression to depression? *Primary Care Psychiatry*, 5, 77-78.
- Iacovides, A., Fountoulakis, K.N., Kaprinis, St. & Kaprinis, G. (2003). The relationship between job stress, burnout and clinical depression. *Journal-of-Affective-Disorders*, 75, 209-221.
- Iacovides, A., Fountoulakis, K.N., Moysidou, C. & Ierodiakonou, C. (1999). Burnout in nursing staff: Is there a relationship between depression and burnout? *International Journal of Psychiatry in Medicine*, 29, 421-433.
- Ierodiakonou, B.I. (1998). Psychoanalytical psychotherapy of anxiety depression in everyday psychiatric practice. *Socijalna Psihijatrija*, 26, 175-177.
- Ierodiakonou, C.S. & Iacovides, A. (1987). Somatic manifestations of depressive patients in different psychiatric settings. *Psychopathology*, 20, 136-143.
- Ierodiakonou, C.S., Iacovides, A. & Ierodiakonou-Benou, I. (1998). Changing patterns of attempted suicide in Greece: clinico-epidemiological and psychodynamic data. *Psychopathology*, 31, 281-292.
- Ines ,T.M. & Sacco, W.P. (1992). Factors related to correspondence between teacher ratings of elementary student depression and student self-ratings. *Journal of Consulting and Clinical Psychology*, 60, 140-142.
- Janisch, P. (1994). *Empirischer Beitrag zur Validierung des Depressionstests für Kinder DTK: Eine Multitrait-Multiinformant-Studie*. Unveröff. Diplomarbeit, Karl-Franzens-Universität Gratz.
- Kanfer, F.H. (1975). Self-management methods. In F. H. Kanfer & A. P. Goldstein (Eds.), *Helping people change: A textbook of method* (pp.309-356). New York: Pergamon.
- Kanfer, F.H., Reinecker, H. & Scmelzer, D. (Hrsg.) (2006). *Selbstrmanagement-Therapie. Ein Lehrbuch für die Klinische Praxis* (4., durchgesehene Auflage). Heidelberg: Springer.
- Kashani, J.H., Anderson, J.C. & Robins, A.J. (1983). Depression in a sample of of 9-year-old children. *Archives of General Psychiatry*, 40, 1217-1223.
- Kashani, J.H., Beck, N.C., Hooper, E.W., Fallaki, C., Corcoran, C.M., McAllister, J.A., Rosenberg, T.K. & Reid, J.C. (1987). Psychiatric disorders in a community sample of adolescents. *American Journal of Psychiatry*, 144, 584-589.
- Kashani, J.H., McGee, R.O., Clarkson, S.E., Anderson, J.C., Walton, L.A., Williams, S. et al. (1983). Depression in a sample of 9-year-old children: Prevalence and associated characteristics. *Archives of General Psychiatry*, 40, 1217-1223.
- Kaslow, N.J., Rhem, L.P. & Siegal, A.W. (1984). Social-cognitive and cognitive correlates of depression in children. *Journal of Abnormal Child Psychology*, 12, 605-620.
- Kazdin, A.E. (1981). Assessment techniques for childhood depression: a critical appraisal. *Journal of the American Academy of Child Psychiatry*, 20, 358-375.
- Kazdin, A.E. (1987). Children's depression scale: Validation with child and parent ratings. *Journal of the American Academy of Child Psychiatry*, 28, 29-41.
- Kazdin, A.E. (1989). Identifying depression in children: a comparison of alternative selection criteria. *Journal of Abnormal Child Psychology*, 17, 437-454
- Kazdin, A.E. & Petti, T.A. (1982). Self-report and interview measure of childhood and adolescent depression. *Journal of Child Psychology and Psychiatry*, 22, 157-164.

- Kazdin, A.E., Colbus, D. & Rodgers, A. (1986). Assessment of depression and diagnosis of depressive disorder among psychiatrically disturbed children. *Journal of Abnormal Child Psychology*, 14, 499-515.
- Kazdin, A.E., French, N.H., Unis, A.S. & Esveldt-Dawson, K. (1983). Assessment of childhood depression: correspondence of child and parent ratings. *Journal of the American Academy of Child Psychiatry*, 28, 29-41.
- Kelz, I. (2000). *Dependenz und Depressivität von Kindern in der Selbst- und Fremdbeurteilung*. Unveröff. Diplomarbeit, Karl-Franzens-Universität Graz.
- Kendall, P. C., Cantwell, D. P. & Kazdin, A. E. (1989). Depression in children and adolescents: assessment issues and recommendations. *Cognitive Therapy and Research*, 13, 109-146.
- Kendell, R. E. (1976). The classification of depression: a review of contemporary confusion. *British Journal of Psychiatry*, 129, 15-28.
- Klasen, H., Woerner, W., Wolke, D., Meyer, R., Overmeyer, S., Kaschnitz, W., Rothenberger, A. & Goodman, R. (2000). Comparing the German versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the Child Behavior Checklist. *European Child and Adolescent Psychiatry*, 9, 271-276.
- Klauer, K.J. (2001). Wie misst man Schulleistungen? In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (2., unveränd. Aufl., S. 103-116). Weinheim: Beltz.
- Kleftaras, G. & Didaskalou, E. (2007). Incidence and teachers' perceived causation of depression in primary school children in Greece. *School Psychology International*, 27, 296-314.
- Klein, M. (1960). Zur Psychogenese der manisch-depressiven Zustände. *Psyche*, 14, 256-283.
- Knight, D., Hensley, V.R. & Waters, B. (1988). Validation of the Children's Depression Scale and the Children's Depression Inventory in a prepubertal sample. *Journal of Child Psychology and Psychiatry*, 29, 853-863.
- Kokkevi, A., Loukadakis, M., Plagianakou, S., Politikou, K. & Stefanis, C. (2000 a). Sharp increase in illicit drug use in Greece: trends from a general population survey on licit and illicit drug use. *European Addiction Research*, 6, 42-49.
- Kokkevi, A., Terzidou, M., Politikou, K. & Stefanis, C. (2000 b). Substance use among high school students in Greece: outburst of illicit drug use in a society under change. *Drug and Alcohol Dependence*, 58, 181-188.
- Kokolakis, M. (2003). Μια αυτοκρατορία σε κρίση, στο *Ιστορία του Νέου Ελληνισμού 1770-2000* (1ος τόμος Η Οθωμανική Αυτοκρατορία 1770-1821, σελ. 41-58). Αθήνα: Ελληνικά Γράμματα.
- Kolko, D. J. & Kazdin, A. E. (1993). Emotional/behavioural problems in clinic and nonclinic children: Correspondence among child, parent and teacher reports. *Journal of Child Psychology and Psychiatry*, 34, 991-1006.
- Kontaxakis, V.P. & Christodoulou, G.N. (2000). Late life depression: Epidemiological data and clinical manifestations. *Psychiatriki*, 11, 166-169.
- Kontaxakis, V.P., Havaki-Kontaxaki, B.J., Margariti, M.M., Stamouli, S.S., Kollias, C.T., Angelopoulos, E.K. & Christodoulou, G.N. (2000). The Greek version of the Calgary Depression Scale for Schizophrenia. *Psychiatry Research*, 94, 163-171.
- Kontaxakis, V.P., Havaki-Kontaxakis, B.J., Stamouli, S.S., Margariti, M.M., Kollias, C.T. & Christodoulou, G.N. (2002). Depression measures and motor side-effects in patients with acute schizophrenia. *Schizophrenia Research*, 56, 197-198.
- Koskelainen, M., Sourander, A., & Kaljonen, A. (2001) The Strengths and Difficulties Questionnaire among Finnish school-aged children and adolescents. *European Child and Adolescent Psychiatry*, 9, 277-284.
- Kovacs, M. (1980/1981). Rating scales to assess depression in school aged children. *Acta Paedopsychiatrica*, 46, 305-315.
- Kovacs, M. (1982). *The Interview Schedule for Children (ISC)*. Unpubl. Manuscript. Pittsburgh: University of Pittsburgh, Department of Psychiatry.
- Kovacs, M. (1983). *The Children's Depression Inventory: A self-rated depression scale for school-aged youngsters*. Unpublished manuscript. University of Pittsburgh School of Medicine.
- Kovacs, M. (1985). The Children's Depression Inventory (CDI). *Psychopharmacology Bulletin*, 21, 995-998.
- Kovacs, M. (1985). The Interview Schedule for Children. *Psychopharmacology Bulletin*, 21, 991-994.

- Kovacs, M. (1992). *The Children's Depression Inventory (CDI) Manual*. Toronto, ON: Multi-Health-Systems.
- Kovacs, M. (2004). *Children's Depression Inventory (CDI) Technical Manual*. Toronto, ON: Multi-Health-Systems.
- Kovacs, M., Akiskal, H.S., Gatsonis, C. & Parrone, P.L. (1994). Childhood-onset dysthymic disorder: clinical features and prospective naturalistic outcome. *Archives of General Psychiatry*, *51*, 365-374.
- Kovacs, M. & Beck, A.T. (1977). An empirical clinical approach toward a definition of childhood depression. In J. G. Schulterbrandt und A. Raskin (Eds.), *Depression in childhood* (pp. 1-25). New York: Raven.
- Kovacs, M. & Devlin, B. (1998). Internalizing disorders in childhood. *Journal of Child Psychology and Psychiatry*, *39*, 47-63.
- Kresanov, K., Tuominen, J., Piha, J. & Almqvist, F. (1998). Validity of child psychiatric screening methods. *European Child and Adolescent Psychiatry*, *7*, 85-95.
- Kypraiou, M. (2002). *Einschätzung der Schwierigkeiten und Möglichkeiten von Kindern beim Übergang zur Pubertät aus der Perspektive von Jugendlichen, Eltern und Lehrer*. Unveröff. Diplomarbeit, Aristoteles Universität zu Thessaloniki.
- Laux, G. (Hrsg.) (2002). *Depression 2000*. Berlin: Springer.
- Lefkowitz, M.M. & Tesiny, E.P. (1984). Rejection and depression: prospective and contemporary analysis. *Developmental Psychology*, *20*, 776-785.
- Leon, G.R., Kendall, P.C. & Garber, J. (1980). Depression in children: parent, teacher, and child perspectives. *Journal of Abnormal Child Psychology*, *8*, 221-235.
- Lepine, J.P., Gastpar, M., Mendlewicz, I. & Tylee, A. (1997). Depression in the community: the first pan-European study DEPRES (Depression Research in European society). *International Journal of Clinical Psychopharmacology*, *12*, 19-29.
- Lewin, K. (1953). *A dynamic theory of personality*. New York: McGraw-Hill.
- Lewinsohn, P.M., Clarke, G.N., Seeley, J.R. & Rohde, P. (1994). Major depression in community adolescents: Age of onset, episode duration, and time to recurrence. *Journal of the American Academy of Child and Adolescent Psychiatry*, *33*, 809-818.
- Lewinsohn, P.M., Rohde, P., Seeley, J.R. & Fischer, S.A. (1993). Age-cohort changes in the lifetime occurrence of depression and other mental disorders. *Journal of Abnormal Psychology*, *102*, 110-120.
- Lewinsohn, P.M., Rohde, P., Seeley, J.R. & Hops, H. (1991). Comorbidity of unipolar depression: I. Major Depression and Dysthymia. *Journal of Abnormal Psychology*, *100*, 2, 205-213.
- Lewinsohn, P.M. & Seeley, J.R. (1998). Major depressive disorder in older adolescents: Prevalence, risk factors, and clinical implications. *Clinical Psychology Review* *18*, 765 – 794.
- Lewinsohn, P.M., Hops, H., Roberts, R.E., Seeley, J.R. & Andrews, J.A. (1993). Adolescent psychopathology: I. Prevalence and incidence of depression and other DSM-III-R disorders in high school students. *Journal of Abnormal Psychology*, *102*, 133-144.
- Lewinsohn, P.M. & Seeley, J.R. (1998). Major depressive disorder in older adolescents: Prevalence, risk factors, and clinical implications. *Clinical Psychology Review* *18*, 765 – 794.
- Liakos, A., & Giannitsi, S. (1984). E axiopistia kai egkurotita tis tropopoemenis Ellinikis klimakas aghous tou Spielberger. (The reliability and validity of the Greek version of Spielberger's anxiety scale). *Egkefalos*, *4*, 71-76.
- Lienert, G. A. & Raatz, U. (1994). *Testaufbau und Testanalyse* (5. Auflage). Weinheim: Beltz.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Auflage). Weinheim: Beltz.
- Lilienfeld, A.M. & Lilienfeld, D.E. (1980). *Foundations of Epidemiology*. New York: Oxford University Press.
- Linden, M., Maier, W., Achberger, M. et al. (1996). Psychische Erkrankungen und ihre Behandlung in Allgemeinarztpraxen in Deutschland. Ergebnisse aus einer Studie der Weltgesundheitsorganisation (WHO). *Nervenarzt*, *67*, 205-215.
- Linsbogen, U. (1996). *Integration in der Schulklasse und Depressivität: Eine empirische Untersuchung an 10-jährigen Mädchen und Knaben*. Unveröff. Diplomarbeit, Karl-Franzens-Universität Graz.

- Lobert, W. (1989). Untersuchung von Merkmalen depressiver Verstimmung in der Pubertät mit dem Kinder-Depressions-Inventar nach Kovacs. *Zeitschrift für Kinder- und Jugendpsychiatrie*, 17, 194-201.
- Lobert, W. (1990). Untersuchung zur Struktur der depressiver Verstimmung in der Pubertät mit dem GCDI (German Children's Depression Inventory). *Zeitschrift für Kinder- und Jugendpsychiatrie*, 18, 18-22.
- Lobovits, D.A. & Handal, P.J. (1985). Childhood depression: prevalence using DSM-III criteria and validity of parent and child scales. *Journal of Pediatric Psychology*, 10, 45-54.
- Lösel, F., Bliesener, T., Klünder, A. & Köferl, P. (1988). *Verhaltens- und Erlebens-Probleme bei deutschen Jugendlichen: Anwendungen der Youth Self Report Form der Child Behavior Checklist und ein Vergleich mit Daten aus den Vereinigten Staaten*. Unveröff. Manuscript. Bielefeld: Universität, Sonderforschungsbereich 227.
- Lyketsos, G., Richardson, C., Mouzakis, D., & Paterakis, P. (1988). Depressive syndromes and symptoms in elderly schizophrenics. *New Trends in Experimental and Clinical Psychiatry*, 4, 17-23.
- Lykouras, L., Gournellis, R., Fortos, A., Oulis, P. & Christodoulou, G.N. (2002). Psychotic (delusional) major depression in the elderly and suicidal behaviour. *Journal of Affective Disorders*, 69, 225-229.
- Lykouras, L., Gournellis, R., Fortos, A., Oulis, P. & Christodoulou, G.N. (2000). Psychotic major depression in the elderly and suicide behavior. *Psychiatriki*, 11, 177-183.
- Madianos M.G., Economou M, Hatjiandreu A., Papageorgiou A. & Rogakou E. (1999). Changes in public attitudes towards mental illness in the Athens area (1979/1980-1994). *Acta Psychiatrica Scandinavica* 99, 73-78.
- Madianos, M.G., Gefou-Madianou, D. & Stefanis, C.N. (1993). Depressive symptoms and suicidal behavior among general population adolescents and young adults across Greece. *European Psychiatry*, 8, 139-146.
- Madianos, M.G., Gournas, G. & Stefanis, C.N. (1992). Depressive symptoms and depression among elderly people in Athens. *Acta Psychiatrica Scandinavica*, 86, 320-326.
- Madianos, M. & Stefanis, C. (1992). Changes in the prevalence of symptoms of depression and depression across Greece. *Social Psychiatry and Psychiatric Epidemiology*, 27, 211-219.
- Magnusson, D. & Endler, N. S. (Eds.) (1977). *Personality at the crossroads: Current issues in interactional psychology*. Hillsdale: Lawrence Erlbaum.
- Mahler, M.S. (1961). On sadness and grief in infancy and childhood. *Psychoanalytic Study of the Child*, 16, 332-354.
- Maikranz, J.M., Steele, R.G. & Forehand, R. (2003). Longitudinal correlates of depressive symptoms among urban african american children: II. Extension of findings across 3 years. *Journal of Clinical Child and Adolescent Psychology*, 32, 606-612.
- Malmquist, C. (1971). Depressions in childhood and adolescence. *New England Journal of Medicine*, 284, 887-892, 955-961.
- Malovic, D. (2000). *Mütterlicher Erziehungsstil und Depressivität bei Kindern*. Unveröff. Diplomarbeit, Karl-Franzens-Universität Graz.
- Markianos, M., Hatzimanolis, J. & Lykouras, L. (2002). Relationship between prolactin responses to ECT and dopaminergic and serotonergic responsivity in depressed patients. *European Archives of Psychiatry and Clinical Neuroscience*, 252, 166-171.
- Marmanidis, H., Holme, G. & Hafner, R.J. (1994). Depression and somatic symptoms: A cross-cultural study. *Australian and New Zealand Journal of Psychiatry*, 28, 274-278.
- Marzocchi, G.M., Capron, C., di Pietro, M., Tauleria, E.D., Duyme, M., Friggerio, A., Gaspar, M.F., Hamilton, H., Pithon, G., Simões, A. & Théron, C. (2004). The use of the Strengths and Difficulties Questionnaire (SDQ) in Southern European countries. *European Child and Adolescent Psychiatry*, 13, II/40-II/46.
- Mastaler, M. (2001). *Somatoforme Störungen bei Jugendlichen und jungen Erwachsenen: Epidemiologische Validierungsstudie eines umstrittenen Störungsbildes*. Roderer Verlag: Regensburg.
- Mattison, R.E., Handford, H.A., Kales, H.C., Goodman, A.L. & McLaughlin, R.E. (1990). Four-year predictive value of the Children's Depression Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 2, 169-174.
- Mavropoulou, S. & Padelidu, S. (2000). Teacher's causal attributions for behaviour problems in relation to perceptions of control. *Educational Psychology*, 22, 191-202.

- McDonald, V. M., Tsiantis, J., Achenbach, T. A., Motti-Stefanidi, F. & Richardson, S. C. (1995). Competencies and problems reported by parents of greek and amerian children, ages 6-11. *European Child and Adolescent Psychiatry*, 4, 1-13.
- McGee, R., Williams, S. & Silva, P. A. (1995). Factor structure and correlates of ratings of inattention, hyperactivity, and antisocial behavior in a large sample of 9-years-old children from the general popualtion. *Journal of Consulting and Clinical Psychology*, 53, 480-490.
- Merret, F. & Wheldall, K. (1993). How do teachers learn to manage classroom behaviour? A study of teachers' opinions about their initial training with special reference to classroom behavior management. *Educational Studies*, 19, 91-106.
- Mestre, V., Garcia, R., Frias, D. & Moreno, C. (1993). Validación y factorización de la Escala de sintomatología depresiva para maestros (ESDM) en población adolescente. *Revista de psiquiatría infanto-juvenil, Número 1*, 29-33.
- Monreal, P. (1988). *Estudio de la sintomatología depresiva infantil en la provincia de Gerona: Análisis de algunos factores individuales y psicosociales*. Dissertation. Universidad Autónoma de Barcelona. Departamento de Psicología de la Salud.
- Moor, S., Sharrock, G., Scott, J., McQueen, H., Wrate, R., Cowan, J. & Blair, C. (2000). Evaluation of a teaching package designed to improve teachers' recognition of depressed pupils – a pilot study. *Journal of Adolescence*, 23, 331-342.
- Moretti, M.M., Fine, S., Haley, G. & Marriage, M.B. (1985). Childhood and adolescent depression: Child-report versus parent-report information. *Journal of the American Academy of Child Psychiatry*, 24, 298-302.
- Motti-Stefanidi, F., Besevegis, E. & Giannitsas, N. (1996). Teachers' perceptions of school-age children's competence and mental health. *European Journal of Personality*, 10, 263-282.
- Motti-Stefanidi, F., Tsiantis, J. & Richardson, S.C. (1993). Epidemiology of behavioural and emotional problems of primary schoolchildren in Greece. *European Child and Adolescent Psychiatry*, 2, 2, 111-118.
- Mullick, MSI. & Goodman, R. (2001). Questionnaire screening for mental health problems in Bangladeshi children: a preliminary study. *Social Psychiatry and Psychiatric Epidemiology*, 36, 94-99.
- Muris P., & Maas, A. (2004). Strengths and difficulties as correlates of attachment style in institutionalized and non-institutionalized children with below-average intellectual abilities. *Child Psychiatry and Human Development*, 34, 317-328.
- Muris, P., Meesters, C. & van den Berg, F. (2003). The Strengths and Difficulties Questionnaire (SDQ): Further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *European Child and Adolescent Psychiatry*, 12, 1-8.
- Nelson, W.M. III, Politano, P.M., Finch, A.J., Wendel, N. & Mayhall, C. (1987). Children's Depression Inventory: Normative data and utility with emotionally disturbed children. *Journal of the American Academy of Child and Adolescent Psychiatry*, 26, 43-48.
- Nevermann, C. (1989). *Depression in children: Self-reports and reports from peers and teachers in a normal elementary school population*. Paper presented at the 19th Annual Congress of the European Association of Behaviour Therapy, Sept. 20-24, Vienna, Austria.
- Nevermann, C. (1992). *Depressive Störungen im Kindes- und Jugendalter: Intra- und interpersonale Wahrnehmungen depressiver Symptome in Schulklassen*. Unveröff. Dissertation, Freie Universität Berlin.
- Nissen, G. (2002). *Seelische Störungen bei Kindern und Jugendlichen: Alters- und entwicklungsabhängige Symptomatik und ihre Behandlung*. Stuttgart: Klett-Cotta.
- Nolen-Hoeksema, S., Seligman, M.E.P. & Girgus, J.S. (1986). Learned helplessness in children: a longitudinal study of depression, achievement, and explanatory style. *Journal of Personality and Social Psychology*, 51, 435-442.
- Ntonias, S. & Demertzis, I. (1983). Stathmisis tis katathliptikis sympomatologias me to erotimatologio katathilipsis tou Beck. (Standardization of depressive symptomatology using the Beck Depression Inventory). *Practica tou 10ou senedriou neurologias kai psychiatrikis*, 1, 486-492.
- Ofner, P. (1995). *Depressive Verstimmung und Kompetenzeinschätzungen bei 10- bis 13-jährigen*. Unveröff. Diplomarbeit, Karl-Franzens-Universität Gratz.
- Ollson, G.I. & von Knorring, A.L. (1999). Adolescent depression: prevalence in Swedish high-school students. *Acta Psychiatrica Scandinavica*, 99, 324-331.

- Papageorgiou, C., Grapsa, E., Christodoulou, N.G., Zerefos, N., Stamatelopoulos, S. & Christodoulou, G.N. (2001). Association of serum nitric oxide levels with depressive symptoms: A study with end-stage renal failure patients. *Psychotherapy and Psychosomatics*, 70, 216-220.
- Papakostas, Y.G., Markianos, M., Papadimitriou, G.N. & Zervas, I.M. (1995). Effects of methysergide and ritanserin on the prolactin and thyrotropin responses to TRH in depressed patients. *European Neuropsychopharmacology*, 5, 15-18.
- Paparrigopoulos, T. (2002). Melatonin response to atenolol administration in depression: Indication of beta-adrenoreceptor dysfunction in a subtype of depression. *Acta Psychiatrica Scandinavica*, 106, 440-445.
- Paparrigopoulos, T., Psarros, C., Bergiannaki, J.D., Varsou, E., Dafni, U. & Stefanis, C. (2001). Melatonin response to clonidine administration in depression: Indication of presynaptic alpha-sub-2-adrenoceptor dysfunction. *Journal of Affective Disorders*, 65, 307-313.
- Parashos, I.A., Stamouli, S., Rogakou, E., Theodotou, R., Nikas, I. & Mougias, A. (2002). Recognition of depressive symptoms in the elderly: What can help the patient and the doctor. *Depression and Anxiety*, 15, 111-116.
- Papatheofilou, R., Sokou-Bada, K., Michelogiannis, J. & Pantelakis, S. (1988). Σχολική επίδοση: κοινωνικές, ψυχολογικές και φυσικές παράμετροι; ψυχιατρική διαταραχή και σχολική επίδοση. *Ψυχολογικά Θέματα*, 1, 211-229.
- Parker, G. & Lipscombe, P. (1981). Influences of maternal overprotection. *British Journal of Psychiatry*, 2, 111-118.
- Pearce, J. (1977). Depressive disorder in childhood. *Journal of Child Psychology and Psychiatry*, 18, 79-82.
- Pisaniyas, P. (2000). Επανάστασι και Έθνος, στο *Ιστορία του Νέου Ελληνισμού 1770-2000* (3ος τόμος Μια ιστορική-κοινωνιολογική προσέγγιση του 1821, σελ 33-52). Αθήνα: Ελληνικά Γράμματα.
- Plutarch. *Moralia. Von der Seelenruhe*. Verdeutscht und hrsg. von Ax, Wilhelm. Leipzig: Dieterich, 1942.
- Polaino-Lorente, A. & Domènech, E. (1993). Prevalence of childhood depression. Results of the first study in Spain. *Journal of Child Psychology and Psychiatry*, 34, 1007-1017.
- Popper, K. R. (2005). *Alles Leben ist Problemlösen* (9. Aufl.). München: Piper.
- Pötz, M. (2003). *Depressionstest für Kinder – Untersuchungen im klinischen Bereich*. Unveröff. Diplomarbeit, Karl-Franzens-Universität Graz.
- Poulou, M. & Norwich, B. (2000). Teachers' perceptions of students with emotional and behavioural difficulties: severity and prevalence. *European Journal of Special Needs Education*, 15, 2, 171-187.
- Puig-Antich, J. & Chambers, W. (1978). *The Schedule for Affective Disorders and Schizophrenia for school-aged children*. New York: State Psychiatric Institute.
- Puig-Antich, J., Orvachle, H., Tabrizi, M. & Chambers, W.J. (1983). *Schedule for Affective Disorders and Schizophrenia for School-Age Children Epidemiological Version*. Unpublished manuscript, New York State Psychiatric Institute, Department of Child and Adolescent Psychiatry, New York.
- Puura, K., Tamminen, T., Almqvist, F., Kresanov, K., Kumpulainen, K., Koivisto, A.M. & Moilanen, I. (1997). Should depression in young school-children be diagnosed with different criteria? *European Child and Adolescent Psychiatry*, 6, 12-19.
- Radloff, L.S. (1991). The use of the Center for Epidemiological Studies Depression Scale in adolescents and young adults. *Journal of Youth and Adolescence*, 20, 149-166.
- Reicher, H. (1998). *Depressionen bei Kindern und Jugendlichen*. Münster: Waxmann.
- Reicher, H. & Rossmann, P. (1991). Zu den psychometrischen Eigenschaften einer deutschen Version des Children's Depression Inventory. *Diagnostica*, 37, 236-251.
- Reinhard, H.-G., Bowi, U. & Rulcovius, G. (1990). Stabilität, Reliabilität und Faktorenstruktur einer deutschen Fassung des Children's Depression Inventory (Kovacs 1983). *Zeitschrift für Kinder- und Jugendpsychiatrie*, 18, 185-191.
- Reinherz, H.Z., Giaconia, R.M., Pakiz, B., Silverman, A.B., Frost, A.K. & Lefkowitz, E.S. (1993). Psychosocial risks for major depression in late adolescence: A longitudinal community study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32, 1155-1163.

- Reitbauer, E. (1999). *Erklärungsstil und Depressivität im Kindesalter*. Unveröff. Diplomarbeit, Karl-Franzens-Universität Graz.
- Renner, E. (1993). *Validierung einer deutschen Form des Self-Description Questionnaire I (SDQ-I) von H. W. Marsch: Das Selbstkonzept von 10- bis 14-jährigen Schülern/innen und Beziehungen zu Depressivität, Prüfungsängstlichkeit und Schulklima*. Unveröff. Diplomarbeit, Karl-Franzens-Universität Graz.
- Reynolds, C.R. (1987). *Assessment of depression in adolescents: Manual for the Reynolds Adolescent Depression Scale (RADS)*. Odessa, FL: Psychological Assessment Resources.
- Rie, H.E. (1966). Depression in childhood. A survey of some pertinent contributions. *Journal of the American Academy of Child Psychiatry*, 5, 653-685.
- Rochlin, G. (1959). The loss complex: A contribution to the aetiology of depression. *Journal of the American Psychoanalytic Association*, 7, 299-316.
- Romano, E., Tremblay, R.E., Vitaro, F., Zoccolillo, M. & Pagani, L. (2001). Prevalence of psychiatric diagnoses and the role of perceived impairment: Findings of an adolescent community sample. *Journal of Child and Adolescent Psychiatry*, 42, 451-461.
- Rønning, J.A., Handegaard, B.H. & Mørch, W.-T., (2004). The Strengths and Difficulties Questionnaire as a screening instrument in Norwegian community sample. *European Child and Adolescent Psychiatry*, 13, 73-82.
- Rossmann, P. (1991). *Depressionsdiagnostik im Kindesalter: Grundlagen, Klassifikation, Erfassungsmethoden*. Bern: Huber.
- Rossmann, P. (1993). *Depressionstest für Kinder - DTK*. Bern: Huber.
- Rossmann, P. (2005). *DTK – Depressionstest für Kinder* (2., überarb. und erweiter. Auflage). Bern: Huber.
- Rost, J. (1999). Was ist aus dem Rasch-Modell geworden? *Psychologische Rundschau*, 50, 140-156.
- Rothenberger, A. & Woerner, W. (2004). Editorial. Strengths and Difficulties Questionnaire (SDQ) – Evaluations and applications. *European Child and Adolescent Psychiatry*, 13, II/1-II2 (suppl. 2).
- Rotundo, N. & Hensley, V.R. (1985). The Children's Depression Scale. A study of its validity. *Journal of Child Psychology and Psychiatry*, 26, 917-927.
- Roussos, A., Karantanos, G., Richardson, C., Hartman, C., Karajiannis, D., Kyprianos, S., Lazaratou, H., Mahera, O., Tassi, M. & Zoubou, V. (1999). Achenbach's Child Behaviour Checklist and Teachers' Report Form in a normative sample of Greek children 6-12 years old. *European Child and Adolescent Psychiatry*, 8, 165-172.
- Rutter, M. (1967). A children's behaviour questionnaire for completion by teachers: Preliminary findings. *Journal of Child Psychology and Psychiatry*, 8, 1-11.
- Rutter, M. & Garmezy, N. (1983). Developmental psychopathology. In F.M. Hetherington (Ed.), *Handbook of clinical psychology: Socialization, personality and social development* (pp. 775-911). New York: Wiley.
- Rutter, M., Tizard, J., Yule, W., Graham, P., & Whitmore, K. (1976). Research report: Isle of White studies 1964-1974. *Psychological Medicine*, 6, 313-332.
- Rutter, M., Tizard, J. Whitmore, K. (Eds). (1970). *Education, Health and Behaviour*. London: Longmans.
- Sacco, W.P. & Graves, D.J. (1985). Correspondence between teacher ratings of childhood depression and child self-rating. *Journal of Clinical Child Psychology*, 14, 353-355.
- Sandler, J. & Joffe, W.G. (1965). Notes on childhood depression. *International Journal of Psychoanalysis*, 46, 88-96
- Sarantidis, D., Thomas, A., Iphantis, K., Katsaros, N., Tripodianakis, J. & Katsabouris, G. (1997). Levels of anxiety, depression and denial in patients with myocardial infarction. *European Psychiatry*, 12, 149-151.
- Saß, H., Wittchen, H.-U. & Zaudig, M. (1996). *Diagnostisches und Statistisches Manual Psychischer Störungen DSM-IV*. Göttingen: Hogrefe.
- Saylor, C.F., Finch, A.J., Spirito, A. & Bennett, B. (1984). The Children's Depression Inventory: a systematic evaluation of psychometric properties. *Journal of Consulting and Clinical Psychology*, 52, 955-967.
- Schein, G. (1996). *Depressive Verstimmung, soziale Kompetenz und soziometrischer Status, eine empirische Untersuchung an 11- bis 13-jährigen Kindern*. Unveröff. Diplomarbeit, Karl-Franzens-Universität Graz.

- Schönplflug, W. (2004). *Geschichte und Systematik der Psychologie* (2. Auflage). Weinheim: Beltz PVU.
- Schulz von Thun, F. (2005). *Miteinander reden 1: Störungen und Klärungen. Allgemeine Psychologie der Kommunikation* (Sonderausgabe). Reinbek: Rowohlt.
- Seligman, M. (1979). *Erlernte Hilflosigkeit*. München, Wien, Baltimore: Urban und Schwarzenberg.
- Seligman, M. (1999). *Kinder brauchen Optimismus*. Reinbek bei Hamburg: Rowohlt.
- Seligman, M., Kaslow, N., Alloy, L., Peterson, C., Tananbaum, R. & Abramson, L. (1984). Attributional style and depressive symptoms among children. *Journal of Abnormal Psychology*, 93, 235-238.
- Shaffer, D., Gould, M.S., Fisher, P., Trautman, P., Moreau, D., Kleinman, M. & Flory, M. (1996). Psychiatric diagnosis in child and adolescent suicide. *Archives of General Psychiatry*, 53, 339-348.
- Simos, G. & Dimitriou, E. (1995). Is there a seasonal pattern in non-seasonal affective disorders? *European Journal of Psychiatry*, 9, 167-171.
- Sitzenfrey, E. (1994). *Depressivität und Selbstkonzept im Kindes- und Jugendalter: Eine empirische Untersuchung an Hauptschülern und Gymnasiasten aus der 5. Und 6. Schulstufe*. Unveröff. Diplomarbeit, Karl-Franzens-Universität Graz.
- Slezak, A. (1993). *Stabilität und innere Struktur des Depressionstests für Kinder bei Volksschülern*. Unveröff. Diplomarbeit, Karl-Franzens-Universität Graz.
- Smedje, H., Broman, J-E., Hetta, J. & von Knorring, A.L. (1999). Psychometric properties of a Swedish version of the "Strengths and Difficulties Questionnaire". *European Child and Adolescent Psychiatry*, 8, 63-70.
- Smucker, M.R., Craighead, W.E., Craighead, L.W. & Green, B.J. (1986). Normative and reliability data for the Children's Depression Inventory. *Journal of Abnormal Child Psychology*, 14, 25-39.
- Sourander, A., Helenius, H. & Piha J. (1995). Parent and teacher reports of problem behaviors in child psychiatric inpatients: Cross-informant correlations on admission and at 5-month follow up. *Child Psychiatry and Human Development*, 26, 85-94.
- Spielberger, C. D., Gorsuch, R. C., Lushene, R. E. (1970). *Manual for the State Trait Anxiety Inventory*. Palo alto, CA: Consulting psychologists Press.
- Spitz, R.A. & Wolf, K.M. (1946). Anaclitic depression: An inquiry into the genesis of psychiatric conditions on early conditions, II. *Psychoanalytic Study of the Child*, 2, 313-342.
- Stasinou, D. P. (1995). Teacher's perceptions of adolescent behavior problems in school: Findings from Greece. *International Journal of Adolescent Medicine and Health*, 8, 271-284.
- Steiger, A. (2002). Neuroendokrinologie der Depression. In G Laux (Hrsg.), *Depression 2000* (S. 29-36). Berlin: Springer.
- Steinhausen, H.C. (1993). *Psychische Störungen bei Kindern und Jugendlichen. Lehrbuch der Kinder und Jugendpsychiatrie* (2., erw. u. überarb. Auflg.). München: Urban & Schwarzenberg.
- Steinhausen, H.C. (1996). *Psychische Störungen bei Kindern und Jugendlichen. Lehrbuch der Kinder und Jugendpsychiatrie* (3. aktualisierte Auflg.). München: Urban & Schwarzenberg.
- Steinhausen, H.C., Winkler Metzke, C., Meier, M. & Kannenberg, R. (1998). Prevalence of child and adolescent psychiatric disorders: The Zürich Epidemiological Study. *Acta Psychiatrica Scandinavica*, 98, 262-271.
- Stiensmeier, J. (1988). Die Erfassung von Depressionen bei Kindern: Eine deutsche Version des Children's Depression Inventory (CDI-d). *Diagnostica*, 34, 320-331.
- Stiensmeier-Pelster, J., Schürmann, M. & Duda, K. (1989). Das Depressions-Inventar für Kinder und Jugendliche (DIJK). Göttingen: Hogrefe.
- Stiensmeier-Pelster, J. Schürmann, M. & Duda, K. (2000). *Depressions-Inventar für Kinder und Jugendliche (DIJK). ndanweisung* (2., überarb. und neunormierte Auflage). Göttingen: Hogrefe.
- Sullivan, J. L. & Felam, S. (1979). *Multiple Indicators. An Introduction*. Beverly Hills: Sage.
- Taylor, S. E. (1999). *Health Psychology*. Boston: McGraw-Hill.
- Trilivas, S., Vassilaki, E. & Chimienti, G. (1998). Coping strategies among Greek high school students and their relationship to anxiety and depression. In Ralf Schwarzer (Ed.), *Advances in health psychology research. Volume 1*. Berlin: Freie Universität.

- Trott, E.G. (2002). Depression im Kindes- und Jugendalter. In G. Laux (Hrsg.), *Depression 2000* (pp. 129-137). Berlin, Heidelberg: Springer.
- Trovato, J., Harris, J., Pryor, C.W. & Wilkinson, C. (1992). Teachers in regular classrooms: An applied setting for successful behavior programming. *Psychology in the Schools*, 29, 52-61.
- Tselebis, A., Moulou, A. & Ilias, I. (2001). Burnout versus depression and sense of coherence: Study of Greek nursing staff. *Nursing and Health Sciences*, 3, 69-71.
- Tsiantis, J., Markidian-Gazerian, B., Sipitanou, A. & Tata-Stamoupolpoulou, L. (1982). *Child Mental Health and Psychosocial Development: National Case Study*. Unpublished report to WHO.
- Tsiantis, J., Motti-Stefanidi, F., Richardson, C., Schmeck, K. & Poustka, F. (1994). Psychological problems of school-age German and Greek children: a cross cultural study. *European Child and Adolescent Psychiatry*, 3, 4, 209-219.
- Twenge, J.M. & Nolen-Hoeksema, S. (2002). Age, gender, socioeconomic status, and birth cohort differences on the Children's Depression Inventory: a meta-analysis. *Journal of Abnormal Psychology*, 11, 578-588.
- Tzanakaki, M., Guazzelli, M., Nimatoudis, I. & Zissis, N.P. (2000). Increased remission rates with venlafaxine compared with fluoxetine in hospitalized patients with major depression and melancholia. *International Clinical Psychopharmacology*, 15, 29-34.
- Ustün, T.B. & Sartorius, N. (1995). *Mental illness in primary care: an international study*. WHO, Wiley, Chichester.
- van Widenfelt, B.M., Goedhart, A.W., Treffers, P.D.A. & Goodman, R. (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child and Adolescent Psychiatry*, 12, 281-289.
- Vaslamatzis, G. (1995). On the process of the psychoanalytic psychotherapy with the borderline depressive patient. *International Forum of Psychoanalysis*, 4, 111-118.
- Vassiliou, G., & Vassiliou, V. (1982). Promoting psychological functioning and preventing malfunctioning. *Paediatrician*, 11, 90-98.
- Vassiliou, G., & Vassiliou, V. (1984). On group theory developments in context: A hellenic view. *International Journal of Group Therapy*, 34, 377-385.
- Verhulst, F.C. & Koot, H.M. (1992). Child psychiatric epidemiology: concepts, methods, and findings. In A.E. Kazdin (Ed.), *Developmental Clinical Psychology and Psychiatry* (vol. 23). Newbury Park, London, New Delhi: Sage Publications.
- Verhulst, F.C., van der Ende, J., Ferdinand, R.F. & Kasius, M.C. (1997). The prevalence of DSM-III-R diagnoses in a national sample of dutch adolescents. *Archives of General Psychiatry*, 54, 329-336.
- von Cranach, M. (1992). The multi-level organisation of knowledge and action. An integration of complexity. In C. M. von, W. Doise & G. Mugny (Eds.), *Social representations and the social bases of knowledge* (pp. 10-22). Lewiston: Hogrefe.
- Watzlawick, P., Beavin, J.H. & Jackson, D.D. (1969). *Menschliche Kommunikation: Formen, Störungen, Paradoxien*. Bern: Huber.
- Weiss, B. & Weisz, J.R. (1988). Factor structure of self-reported depression: clinical-referred children versus adolescents. *Journal of Abnormal Psychology*, 97, 492-495.
- Weiss, B., Weisz, J. R., Politano, M., Carey, M., Nelson, W. M. & Finch, A. J. (1991). Developmental differences in the factor structure of the Children's Depression Inventory. *Psychological Assessment*, 3, 38-45.
- Weisz, J.R. & Weiss, B. (1991). Studying the „referability“ in child clinical problems. *Journal of Consulting and Clinical Psychology*, 59, 266-273.
- Wendel, N.H., Nelson, W.M., Politano, P.M., Mayhall, C.A. & Finch, A.J. (1988). Differentiating inpatient clinically-diagnosed and normal children using the Children's Depression Inventory. *Child Psychiatry and Human Development*, 19, 98-108.
- Wertheimer, M. (1987). *A brief history of psychology (3rd ed.)*. New York, NY: Holt, Rinehart & Winston Inc.
- Wittchen, H.-U., Höfler, M. & Meister, W. (2001). Prevalence and recognition of depressive syndromes in German primary care settings: poorly recognized and treated? *International Clinical Psychopharmacology*, 16, 121-135.

- Wittchen, H.-U., Schuster, P. & Pfister, H. (1999). Depressionen in der Allgemeinbevölkerung – schlecht erkannt und selten behandelt. *Nervenheilkunde*, 18, 202-209.
- Wittchen, H.-U., Müller, N. & Storz, S. (1998a). Psychische Störungen: Häufigkeit, psychosoziale Beeinträchtigungen und Zusammenhänge mit körperlichen Erkrankungen. *Das Gesundheitswesen*, 60 (Sonderheft 2), 95-100.
- Wittchen, H.-U., Nelson, C. B. & Lachner, G. (1998). Prevalence of mental disorders and psychosocial impairments in adolescents and young adults. *Psychological Medicine*, 28, 109-126.
- Woerner, W., Becker, A., Friedrich, C., Klasen, H., Goodman, R. & Rothenberger, A. (2002). Normierung und Evaluation der deutschen Elternversion des Strengths and Difficulties Questionnaire (SDQ): Ergebnisse einer repräsentativen Felderhebung. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 30, 105-112.
- Worchel, F.F., Hughes, J.N., Hall, B.M., Stanton, S.B., Stanton, H. & Little, V.Z. (1990). Evaluation of subclinical depression in children using self-, peer-, and teacher-report measures. *Journal of Abnormal Child Psychology*, 18, 271-282.
- World Health Organisation (1992). *International Classification of Mental and Behavioral Disorders*. Geneva: World Health Organisation.
- World Health Organisation (2000). *Bericht über die gesundheitliche Situation der jungen Menschen in der Europäischen Union. Arbeitspapier der Kommissionsdienststellen*. [Online Dokument] URL http://ec.europa.eu/health/ph_information/reporting/ke01_de.pdf (02.09.2007).
- Zachariadis, N. (1991). Adlerian psychotherapeutic approach to depression. *Psychiatriki*, 2, 305-308.
- Zimbardo, P. G. (1992). *Psychologie* (5., neu übers. und bearb. Auflage). Berlin: Springer.

Tabellenverzeichnis

Tabelle

- 3.1 Klassifikation depressiver Störungen nach DSM-IV und ICD-10
- 3.2a Vier für diagnostische Tätigkeit unterscheidbare Kategorien
- 3.2b Vier für diagnostische Tätigkeit unterscheidbare Kategorien
- 3.3 Auflistung epidemiologischer Untersuchungen zum Auftreten depressiver Störungen bei Kindern und Jugendlichen. Tabelle in Anlehnung an Groen und Petermann, 2002, S. 38, ergänzt von C. Kampisiou
- 3.4 Diagnostische Interviewschemata für die Anwendung bei Kindern und Jugendlichen (Essau, 2002, S. 32)
- 3.5 Beispiele von Selbstbeurteilungs-Fragebögen zur Erfassung von Depressionen bei Jugendlichen (Essau, 2002, S. 35)
- 3.6 Methoden zur Erfassung von Depression (nach Essau, 2002, S. 42)
- 4.1 Anforderungsschema für die einzusetzenden Messinstrumente
- 4.2 Liste der für die Untersuchung mit griechischen Schulkindern ausgewählten Messinstrumente
- 4.3 Liste der für die ausgewählten Merkmale in der Stichprobe herzustellenden Relationen
- 4.4 Liste der ausgewählten Schulorte
- 5.1 Geschlechtsspezifische Verteilung der untersuchten Kinder auf unterschiedlichen Altersgruppen
- 5.2 Geschlechtsspezifische Verteilung der untersuchten Stichprobe auf die zwei besuchten Schulformen
- 5.3 Geschlechtsspezifische Verteilung der untersuchten Stichprobe auf Orte unterschiedlicher Größe
- 5.4 Geschlechtsspezifische Verteilung der untersuchten Stichprobe auf Orte unterschiedlicher Größe
- 5.5 Zahl der weiblichen und männlichen Lehrkräfte in Grundschulen und Gymnasien nach Untersuchungsort
- 5.6 Zahl der in Grundschulen und Gymnasien arbeiteten Lehrkräfte, Ihr Unterrichtsfach am Tag der Untersuchung und die gesamte Zahl ihrer Schüler
- 6.1 Kennwerte und Trennschärfe der 25 Items der DTK-Subskala „Dysphorie/Selbstwertprobleme“
- 6.2 Kennwerte und Trennschärfe der 16 Items der DTK-Subskal „Agitiertes Verhalten“
- 6.3 Kennwerte und Trennschärfe der 14 Items der DTK-Skala „Müdigkeit/Autonome Reaktionen“
- 6.4 Häufigkeitsauszahlungen für Items mit hohem Mittelwert
- 6.5 Trennschärfe der DTK-Items in drei Stichproben
- 6.6 Reliabilität der Skala 1. „Dysphorie/Selbstwertprobleme“
- 6.7 Reliabilität der Skala 2 „Agitiertes Verhalten“
- 6.8 Reliabilität der Skala 3 „Müdigkeit/ autonome Reaktionen“
- 6.9 Reliabilität der DTK-Skala „Depressive Belastung“
- 6.10 Interne Konsistenzen der DTK-Skalen in verschiedenen Subgruppen
- 6.11 Oblimin-Faktorladungen für die 55 Items des DTK in der griechischen Untersuchung 2034GR
- 6.12 CDI Faktorenladungen zweier Items im Vergleich einer normativen und einer klinischen Stichprobe (s. Kovacs, 2003, S. 53)
- 6.13 Mittelwerte, Standardabweichungen und Varianzen der griechischen CDI Items
- 6.14 Antworthäufigkeiten für Items mit einem niedrigen Mittelwert (griechischer CDI)
- 6.15 Trennschärfe des CDI in einer griechischen, einer nordamerikanischen und einer deutschen Stichprobe
- 6.16 Reliabilität der griechischen Version des CDI
- 6.17 Interne Konsistenz des CDI in der Gesamtstichprobe und in einzelnen Teilstichproben
- 6.18 Faktorenanalyse der CDI-Daten aus der griechischen Untersuchung 2034G
- 6.19 Korrelationen zwischen den Gesamtwerten dreier CDI Versionen (in Anlehnung an Kovacs, 2003, S. 57)
- 6.20 Kennwerte der Fragen 1 bis 16 des ESDM in seiner griechischen Fassung (Kampisiou)
- 6.21 Häufigkeitsauszahlungen für Items mit niedrigem Mittelwerten
- 6.22 Trennschärfe der ESDM-Items in zwei Untersuchungen
- 6.23 Reliabilität der ESDM-Skala
- 6.24 Reliabilitätskennwerte aus Studien mit dem ESDM (Informationen aus Casanovas, 2003)
- 6.25 Auf Items bezogene Ergebnisse aus der Faktorenanalyse des ESDM in seiner griechischen Fassung (Kampisiou, 2003)
- 6.26 Die Faktorenstruktur der griechischen ESDM
- 6.27 Faktorenlösungen des ESDM in den Studien von Domènech (19xx), Casanovas (2003) und Kampisiou (2003)
- 6.28 Kennwerte der griechischen SDQ-Items nach Subskalen geordnet (Kampisiou, 2003)
- 6.29 Häufigkeitsauszahlungen für Items mit niedrigem oder hohem Mittelwert
- 6.30 Trennschärfe der SDQ-Items
- 6.31 Reliabilität der gesamten SDQ-Skala „Total Difficulties“ (Summwert aus vier Skalen)
- 6.32 Alpha-Werte für die 5 Subskalen und die Gesamtskala des SDQ in der Gesamtstichprobe und nach Geschlecht der Kinder (Kampisiou, 2003)

Tabelle

-
- 6.33 Reliabilitätswerte des SDQ in verschiedenen griechischen Untersuchungen
- 6.34 Structure Ma5trix für den SDQ nach der Anwendung in der Untersuchung Kampisiou (2003)
- 6.35 Oblimin-Rotation ohne die Items der Conduct Problems Scale
- 6.36 Varimax-Rotation der SDQ Daten aus der Untersuchung Kampisiou (2003)
- 7.1 Häufigkeits- und Prozentwertverteilung der Skala1-Gesamtpunktwerte für die Stichprobe von 2016 griechischen Kindern und Jugendlichen
- 7.2 Auflistung der S1-GPW aus zwei Untersuchungen: Rossmann, deutschsprachig, (1993) und Kampisiou, griechisch, (2003)
- 7.3 Trennpunkte für drei Gruppen nach GPW aus Skala 1 DTK
- 7.4 Verteilung der Schulkinder der Stichprobe auf drei Gruppen mit unterschiedlichen GPW aus Skala 1 des DTK, n =2016
- 7.5 Durchschnittliche Selbstzuweisungen an Items der Skala 1 pro Kind in den Gruppen 1, 2 und 3
- 7.6 Teilbereich der Tabelle über die Verteilung der GPW aus Skala 1 des DTK in der Untersuchung 2034GR, Kampisiou (2003)
- 7.7 Teilbereich der Tabelle über die Verteilung der GPW aus Skala 1 des DTK in den Untersuchungen Rossmann, deutschsprachig, (1993) und Kampisiou, griechisch, (2003)
- 7.8 Trennwerte zur diskriminanten Validität von GPW aus Skala 1 in der Untersuchung von Pötz (2003)
- 7.9 Sensitivität und Spezifität der diskriminanten Validität über Trennwerte auf Skala 1 des DTK nach Pötz (2003)
- 7.10 Verteilung der Skala2-Gesamtpunktwerte (S2-GPW)
- 7.11 Besetzung der nach den Gesamtpunktwerten aus DTK-Skala 2 „Agitiertes Verhalten“ gebildeten Teilgruppen AGV1, AGV2 und AGV3
- 7.12 Durchschnittlich gewählte Items agitierten Verhaltens pro Kind in drei Gruppen
- 7.13 Verteilung der nach Trennwert >9 (Pötz, 2003) gebildeten Teilgruppen von Kindern mit und ohne Hinweise auf mutmaßlich bestehende depressive Störungen klinischer Ausprägung
- 7.14 Verteilung der Skala3-Gesamtpunktwerte (S3-GPW)
- 7.15 Statistische Kennwerte der voran stehenden Tabelle zur Häufigkeitsverteilung der S3-GPW im DTK
- 7.16 Wertebereiche dreier Teilgruppen von Teilnehmern mit unterschiedlichen Selbstzuweisungen von Mengen an Items der Skala 3 (HMaR, MMaR und NMaR)
- 7.17 Teilgruppen von Teilnehmern mit unterschiedlichen Selbstzuweisungen von Mengen an Items der Skala 3 (HMaR, MMaR und NMaR)
- 7.18 S3-GPW als Trennwert zwischen vermutlich mit klinischen Ausprägungen depressiver Störungen belasteten Kindern und anderen
- 7.19 Teilgruppen mit vermutlich unterschiedlichen Belastungen durch depressive Symptome (Cutoff GPW > 7 nach Pötz, s. Rossmann, 2005)
- 7.20 Verteilung der Skala1+2-Gesamtpunktwerte (S1+2-GPW)
- 7.21 DTK Gesamtpunktwerte als Cutoff-Werte zur Identifizierung diagnostizierter depressiver Kinder (aus Pötz, 2003, S. 74)
- 7.22 Kritische DTK-Trennpunktwerte für die Zuordnung der Kinder zu drei Belastungsgruppen
- 7.23 Verteilung von n=2005 Schulkinder auf nach DTK-GPW in Skala 1+2 unterschiedlichen Belastungsgruppen
- 8.1 Mittelwerte und Standardabweichung (SA) der DTK-Skala „Dysphorie/ Selbstwertprobleme“ in der Gesamtstichprobe und in geschlechtsspezifischen Altersgruppen
- 8.2 Mittelwerte und Standardabweichung (SA) der GPW auf der DTK-Skala „Agitiertes Verhalten“ in der Gesamtstichprobe und in geschlechtsspezifischen Altersstufen
- 8.3 Mittelwerte und Standardabweichung (SA) der DTK-Skala „Müdigkeit / autonome Reaktionen“ in der Gesamtstichprobe und in geschlechtsspezifischen Altersstufen nebst separaten t- Tests
- 8.4 Geschlechtsspezifische Verteilung der Kinder aus der griechischen Stichprobe 2034GR auf zwei Gruppen mit unterschiedlicher Wahrscheinlichkeit für das Vorliegen depressiver Belastungen
- 8.5 Altersspezifische Verteilung der untersuchten Kinder auf die Risikogruppen depBel 1 und depBel 2
- 8.6 Geschlechtsspezifische Verteilung der Kinder der Stichprobe 2034GR auf die eingeführten Risikogruppen in unterschiedlichen Altersstufen
- 8.7 Richtung der bei Jungen und Mädchen auf den Altersstufen vom statistischen Erwartungen abweichenden jeweiligen Frequenz mutmaßlich depressiv belasteter Kinder
- 8.8 Nach Schultyp unterschiedene geschlechtsspezifische Verteilung auf DTK Risikogruppen
- 8.9 Überalterte Kinder und Jugendliche der untersuchten Stichprobe in den verschiedenen Schulklassen
- 8.10 Verteilung der in gymnasialen Schulklassen überalterten und der demgegenüber altersgerecht befindlichen Schülerinnen und Schüler auf definierte depressionsspezifische Risikogruppen (nach DTK)
- 8.11 Verteilung der überalterten Mädchen und Jungen auf die nach DTK- Gesamtpunktwerten gebildeten depressionsspezifischen Risikogruppen
- 8.12 Verteilungsparameter für die Teilgruppen alterskonformer und überalterter Jungen und Mädchen in Gymnasialklassen
- 8.13 Mittelwerte und Standardabweichungen auf den drei DTK-Subskalen für überalterte Jungen und Mädchen (nebst t-Test für Geschlechtsunterschiede)
- 8.14 Von Lehrkräften gegebene Leistungsbewertungen für die in ihren Schulklassen alterskonforme oder überalterte Schülerinnen und Schüler
-

Tabelle

-
- 8.15 Mit den Untersuchungsorten variierende Verteilung der Mädchen und Jungen auf die Risikogruppen depBel 1 und depBel 2
- 8.16 Geschlechtsspezifische Verteilung der jüngeren Kinder der Stichprobe (Grundschul Kinder) auf die Risikogruppen depBel 1 und depBel 2
- 8.17 Übersicht über die an den Wohnorten beobachtete Über- und Unterschreitung der Erwartungswerte für Kinder aus Grundschulen in der Risikogruppe depBel2
- 8.18 Geschlechtsspezifische Verteilung der älteren Jungen und Mädchen der Stichprobe (Gymnasien) auf die Risikogruppen depBel 1 und depBel 2
- 8.19 Übersicht über die an den Wohnorten beobachtete Über- und Unterschreitung der Erwartungswerte für Jungen und Mädchen aus den Gymnasien der Stichprobe in der Risikogruppe depBel2
- 8.20 Zusammenfassende Darstellung der mit den Wohnorten variierenden Zuweisungen der Jungen und Mädchen aus der untersuchten Stichprobe auf nach DTK Gesamtpunkten definierten Risikogruppen
- 8.21 Nach drei über die Einwohnerzahl definierten Stadtgrößen geordnete Verteilung der Kinder und Jugendlichen auf die definierten Belastungsgruppen
- 8.22 Nach drei Stadtgrößen geordnete Verteilung der Kinder aus den Grundschulen auf die definierten Belastungsgruppen
- 8.23 Nach drei Stadtgrößen geordnete geschlechtsspezifische Verteilung der Kinder und Jugendlichen aus den Gymnasien auf die definierten Belastungsgruppen
- 9.1 Verteilung der CDI-Gesamtpunktwerte (CDI-GPW)
- 9.2 Frequenzen der auf die wählbaren 78 Wahlpositionen im CDI entfallenden Antworten
- 9.3a Prozentuale Verteilung der nach Punktwerten
- 9.3b Prozentuale Verteilung der nach Punktwerten pro Item gegebenen Antworten
- 9.4 Kritische CDI-Gesamtpunktwerte (nach Kovacs, 2002) und die durch sie zu identifizierenden Belastungsgruppen
- 9.5 Verteilung der Kinder aus Stichprobe 2034GR auf nach Schweregraden unterschiedenen Gruppen mit mutmaßlich unterschiedlicher Belastung durch depressive Symptome
- 9.6 Nach CDI-Gesamtpunktwerten geschätzte Quoten von Hinweisen auf das Auftreten depressiver Störungen in der untersuchten Stichprobe 2034GR
- 10.1 Mittelwerte (M) und Standardabweichung (SA) der CDI-Werte in der Gesamtstichprobe und in geschlechtsspezifischen Altersstufen nebst T-Test
- 10.2 Geschlechtsspezifische Verteilung der Kinder aus der griechischen Stichprobe 2034GR auf zwei Gruppen mit unterschiedlicher Wahrscheinlichkeit für das Vorliegen depressiver Belastungen nach CDI (Cutoff GPW= 19)
- 10.3 Altersspezifische Verteilung der untersuchten Kinder auf die Risikogruppen Dep1+2 und Dep3
- 10.4 Richtung der auf den Altersstufen von statistischen Erwartungen abweichenden Zuordnungen von Kindern zu den Risikogruppe Dep3 (CDI) und DepBel3 (DTK)
- 10.5 Geschlechtsspezifische Verteilung der Kinder auf die eingeführten Risikogruppen in unterschiedlichen Altersstufen
- 10.6 Richtung der bei Jungen und Mädchen auf den Altersstufen von statistischen Erwartungen abweichenden Zuweisungen mutmaßlich depressiv belasteter Kinder nach DTK (Gruppe DepBel3) und nach CDI (Gruppe Dep3)
- 10.7 Nach Schulstufen unterschiedene geschlechtsspezifische Verteilungen der CDI-GPW auf zwei CDI-Risikogruppen (dep1+2 und dep3)
- 10.8 Verteilung der in gymnasialen Schulklassen altersgerecht oder nicht altersgerecht befindlichen Schülerinnen und Schüler auf definierte depressionsspezifische Risikogruppen dep1+2 und dep3 (nach CDI-GPW)
- 10.9 Verteilung der überalterten Mädchen und Jungen aus Gymnasien auf die nach CDI Gesamtpunktwerten gebildeten depressionsspezifischen Risikogruppen
- 10.10 Getrennt nach Schulort (Untersuchungsort) unterschiedene Verteilung der Gesamtstichprobe auf die CDI-Risikogruppen Dep1+2 und Dep3
- 10.11 Geschlechtsspezifische Verteilung der jüngeren Kinder der Stichprobe (Grundschul Kinder) auf die Risikogruppen Dep1+2 und Dep3
- 10.12 Geschlechtsspezifische Verteilung der älteren Kinder der Stichprobe (Gymnasiasten) auf die Risikogruppen Dep1+2 und Dep3
- 10.13 Nach drei Stadtgrößen geordnete Verteilung der Gesamtstichprobe auf die nach CDI-GPW gebildeten Gruppen Dep1+2 und Dep3
- 10.14 Nach drei Stadtgrößen geordnete Verteilung der Kinder aus den Grundschulen auf die nach CDI-GPW gebildeten Gruppen Dep1+2 und Dep3
- 10.15 Nach drei Stadtgrößen geordnete Verteilung der Kinder aus den Gymnasien auf die nach CDI-GPW gebildeten Gruppen Dep1+2 und Dep3
- 11.1 Reliabilitätswerte der SDQ-Subskalen aus unterschiedlichen Studien
- 11.2 Effizienz des SDQ in zwei klinischen Stichprobe von Kindern aus London (N=101) und Kindern aus Dhaka (N=89)
- 11.3 Verteilung der Gesamtwerte aus der SDQ-Skala „Total Difficulties“ („Td“-GPW) (Angaben aus Goodman, Renfrew & Mullick, 2000, S. 133)
- 11.4 Interkorrelationen zwischen den 5 Subskalen des SDQ und der SDQ-Gesamtskala
- 11.5 Kreuztabelle der in 2034GR beobachteten Verteilung der Werte aus den SDQ Subskalen Prosoziales Verhalten und Hyperaktivität (jeweils in 3 Klassen zusammengefasst)

Tabelle

- 11.6 Kreuztabelle der in 2034GR beobachteten Verteilung der Werte aus den SDQ Subskalen Prosoziales Verhalten und „Conduct problems“ (in drei statistischen Klassen zusammengefasst)
- 11.7 Kreuztabelle der in 2034GR beobachteten Verteilung der Werte aus den SDQ Subskalen Prosoziales Verhalten und „Peer problems“ (in Klassen zusammengefasst)
- 11.8 Korrelationen zwischen den Items der SDQ-Skala „Emotional symptoms“
- 11.9 Korrelationen zwischen den Items der SDQ-Skala „Conduct problems“
- 11.10 Korrelationen zwischen den Items der SDQ-Skala „Hyperactivity“
- 11.11 Korrelationen zwischen den Items der SDQ-Skala „Peer problems“
- 11.12 Korrelationen zwischen den Items der SDQ-Skala „Prosocial behaviour“
- 11.13 Korrelationen der SDQ-Skalen mit denen des DTK und des CDI innerhalb der Untersuchung 2034GR
- 11.14 Kreuztabelle der dichotomisierten Gesamtpunkte aus SDQ und DTK (Skala 1+2)
- 11.15 Kreuztabelle der dichotomisierten Gesamtpunkte aus SDQ und CDI
- 11.16 Kritische SDQ-Gesamtpunktwerte (nach Goodman)
- 11.17 Verteilung der Kinder aus der Stichprobe 2034GR auf nach Schweregraden unterschiedenen Gruppen mit mutmaßlich unterschiedlicher Belastung durch „psychische Störungen“
- 11.18 Nach SDQ-Gesamtpunktwerten geschätzte Prävalenzraten für das Vorliegen psychischer Störungen in der Stichprobe der Untersuchung 2034GR
- 12.1 Deutschsprachige Entsprechungen der Items 1 bis 16 in der griechischen und der spanischen Version der Skala ESDM
- 12.2 Häufigkeits- und Prozentualverteilung der ESDM-Gesamtpunktwerte aus den Angaben der Lehrkräfte für eine Stichprobe von 1825 griechischen Kindern und Jugendlichen.
- 12.3 Frequenzen für die auf insgesamt 48 wählbaren Positionen des Verfahrens ESDM gegebenen Lehrerurteile
- 12.4 Frequenzen der von Lehrkräften im ESDM für die Antwort „sehr oft“ pro Item abgegebenen Antworten
- 12.5 In drei Itemgruppen angeordnete und auf das jeweilige Item bezogene Frequenzen der Lehrerurteile zur ESDM Kategorie „sehr oft“
- 12.6 Für die Berechnung des Chi-Quadrats der vorangegangenen Tabelle benutzten Werte
- 12.7 Lehrerurteile über Häufigkeiten des Auftretens der ESDM Items bei den Kindern in ihren Schulklassen.
- 12.8 Rangplätze pro Itemgruppe nach der dem Prozentsatz in der Spalte „1+2 Punkte“ folgenden Rangreihe
- 12.9 Korrelationen (r) zwischen Urteilen der Lehrkräfte (ESDM) und Selbstangaben der Kinder (CDI, DTK-Skalen und SDQ-Skalen) für die Gesamtstichprobe und für zwei Altersstufen (Kinder an Grundschulen versus Jugendliche von Gymnasien).
- 12.10 Geschlechtsspezifische Varianten der Beziehungen zwischen Urteilen von Lehrkräften und Selbsteinschätzungen der Schulkinder in ihren Klassen
- 12.11 Spearman's Korrelationen (ρ) zwischen Urteilen der Lehrkräfte (ESDM) und Selbstangaben der Schulkinder (im CDI) in jeder der in die Untersuchung 2034GR einbezogenen Schulklassen
- 12.12 Faktorenanalyse der mit ESDM abgegebenen Urteile der Lehrkräfte für drei Gruppen von Lehrkräften
- 12.13 Abweichende Itemmengen in der Faktorenstruktur der ESDM-Urteile dreier Gruppen von Lehrkräften
- 12.14 Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 1
- 12.15 Übersicht über die bei CDI Frage 1 durch die Lehrkräfte richtig zugeordneten Prozentsätze an Kindern (prozentuale Trefferliste)
- 12.16 Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 10
- 12.17 Übersicht über die bei CDI Frage 10 durch die Lehrkräfte richtig zugeordneten Prozentsätze an Kindern (prozentuale Trefferliste)
- 12.18 Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 17
- 12.19 Übersicht über die bei CDI Frage 17 durch die Lehrkräfte richtig zugeordneten Prozentsätze an Kindern (prozentuale Trefferliste)
- 12.20 Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 21
- 12.22 Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 24
- 12.23 Übersicht über die bei CDI Frage 24 durch die Lehrkräfte richtig zugeordneten Prozentsätze an Kindern (prozentuale Trefferliste)
- 12.24 Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 214.25: Übersicht über die bei CDI Frage 27 durch die Lehrkräfte richtig zugeordneten Prozentsätze an Kindern (prozentuale Trefferliste)
- 12.26 Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 8
- 12.27 Übersicht über die bei CDI Frage 8 durch die Lehrkräfte richtig zugeordneten Prozentsätze an Kindern (prozentuale Trefferliste)
- 12.28 Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 13
- 12.29 Übersicht über die bei CDI Frage 13 durch die Lehrkräfte richtig zugeordneten Prozentsätze an Kindern (prozentuale Trefferliste)
- 12.30 Kreuztabelle mit den Daten für 9 im CDI und im ESDM übereinstimmenden Items, Item 23
- 12.31 Übersicht über die bei CDI Frage 23 durch die Lehrkräfte richtig zugeordneten Prozentsätze an Kindern (prozentuale Trefferliste)

Tabelle

- 12.32 Prozentualer Trefferliste für alle im ESDM verwendeten 9 CDI-Items
- 12.33 Geschlechtsspezifisch unterschiedene ESDM Urteile der Lehrkräfte aus Grundschulen über die bei den Grundschulkindern (N= 923) zu beobachtenden Symptome (Mann-Whitney Test)
- 12.34 Gesamtbeurteilung der Gymnasiasten (N= 902) im ESDM getrennt nach dem Geschlecht der Lehrkräfte (mittels Mann-Whitney Test)
- 12.35 Beurteilung der Grundschulkindern und der Gymnasiasten in 16 ESDM-Fragen durch Lehrerinnen und Lehrer (Mann-Whitney Test)
- 12.36 Überblick auf die signifikante Abweichungen bei der nach Lehrkräften geschlechtsspezifisch unterschiedenen Beurteilung der Grundschulkindern und der Gymnasiasten in den einzelnen ESDM-Fragen
- 12.37 Nach Studienfächer der Lehrkräfte unterschiedene ESDM Urteile für n = 902 Jugendliche an griechischen Gymnasien (Kruskal Wallis Test)(Gymnasiasten)
- 12.38 Spearman's Korrelationen zwischen ESDM und CDI für die nach Studienfächer getrennten Gruppen von Lehrkräften an den Gymnasien der Untersuchung 2034GR
- 12.39 Nach Geschlecht der Grundschulkindern unterschiedene Gesamtpunktwerte (N= 923) im ESDM (Mann-Whitney Test)
- 12.40 Nach Geschlecht der Gymnasiastkindern unterschiedene Gesamtpunktwerte (N= 902) im ESDM (Mann-Whitney Test)
- 12.41 Für 16 ESDM-Items gegebene Urteile der Lehrkräfte zum Auftreten depressiver Symptome bei Jungen und Mädchen an Grundschulen und Gymnasien (Mann-Whitney Test)
- 12.42 Unterschiede in den nach 16 ESDM-Items gegebenen Urteile der Lehrkräfte zum Auftreten depressiver Symptome bei Jungen und Mädchen an Grundschulen und Gymnasien (Mann-Whitney Test)
- 12.43 ESDM Urteile der Lehrkräfte in doppelter geschlechtsspezifischer Perspektive: Lehrer, Jungen und Mädchen, Lehrerinnen, Jungen und Mädchen (Gesamtpunktwerte, Mann-Whitney Test)
- 12.44 Gesamtbeurteilung der Grundschulkindern (N= 923) im ESDM nach Klassenstufen
- 12.45 Gesamtbeurteilung der Gymnasiastkindern (N= 923) im ESDM durch Lehrer und durch Lehrerinnen auf schulische Klassenstufen bezogen (N=902)
- 12.46 ESDM Urteile und Leistungseinschätzungen durch Lehrer und Lehrerinnen für Grundschulkindern und Gymnasiasten (Kruskal-Wallis Tests)
- 12.47 Nach durch Lehrkräfte angegebenen speziellen Lebensumständen unterschiedene Teilgruppen an Kindern und deren ESDM Daten (Kruskal Wallis Test)
- 12.48 Rangplatzdifferenzen zwischen ESDM Urteilen von Lehrkräften an Grundschulen und Gymnasien für die nach Informationen über Lebensumstände durch die Lehrkräfte unterschiedenen Teilgruppen an Kindern
- 12.49 ESDM Urteile und Frequenzen in den Schulklassen
- 12.50 ESDM Urteile und Schulklasse
- 13.1 Größe der nach GPW in DTK und CDI durch homogene und inhomogene Zuweisungen zu Risikogruppen gebildeten Teilgruppen
- 13.2 Kreuztabelle der über DTK und CDI ermittelten Risikogruppen von Schulkindern mit und ohne mutmaßlich bestehender depressiver Störungen

Abbildungsverzeichnis

Abbildung

- 5 Lage der griechischen Untersuchungsorte
- 5.1 Verteilung nach Geschlecht
- 5.2 Verteilung nach Alter
- 5.3 Verteilung nach Schulform
- 5.4 Verteilung nach Schiedlungsgröße
- 5.5 Verteilung nach Ort
- 7.1 Verteilung der Gesamtpunktwerte in der DTK-Skala „Dysphorie/Selbstwert“
- 7.2 Häufigkeitsverteilung der Gesamtpunktwerte auf Skala 2 des DTK (n = 2016)
- 7.3 Die Häufigkeitsverteilung der DTK-Skala „Müdigkeit/autonome Reaktionen“
- 7.4 Verteilung von Gesamtpunktwerten nach Skala „Depressive Belastung“ (DTK-Skala 1+ 2)
- 8.1 Score in der DTK-Skala „Dysphorie / Selbstwertprobleme“ in Abhängigkeit von Geschlecht und Alter
- 8.2 Score in der DTK-Skala „Agitiertes Verhalten“ in Abhängigkeit von Geschlecht und Alter
- 8.3 Score in der DTK-Skala „Müdigkeit / autonome Reaktionen“ in Abhängigkeit von Geschlecht und Alter
- 9.1 Verteilung der CDI-Gesamtpunktwerte in der griechischen Stichprobe (n = 2013)
- 10.1 Scores in der CDI-Skala in Abhängigkeit von Geschlecht und Alter
- 10.2 Veranschaulichung der hypothetischen Annahmen über die Beziehungen zwischen beobachteten und erwarteten Werten für den auf die jeweilige Stichprobe bezogenen prozentualen Anteil an Kindern und Jugendlichen in Risikogruppen dep3 und DebBel3
- 11.1 Verteilung der SDQ Gesamtpunktwerte
- 12.1 Verteilung der für 1825 griechische Schulkinder erhobenen ESDM Gesamtpunktwerte
- 12.2 Auf 14 Werteklassen verteilte Korrelationen zwischen 104 über ESDM abgegebene Urteile von Lehrkräften und den über CDI erfassten Selbstauskünften der Schulkinder in 104 Schulklassen

Danksagung

Obwohl eine Dissertation die Leistung einer einzelnen Person darzustellen vermag, wäre die Anfertigung auch dieser Arbeit in der vorliegenden Form nicht möglich ohne die unmittelbare oder mittelbare Unterstützung mehrerer Personen. Für diese Unterstützung möchte ich mich an dieser Stelle bei sämtlichen Beteiligten ganz herzlich bedanken.

Aus der Bereitschaft von 2034 griechischen Schulkindern, auf unsere der persönlichen Sphäre nahe liegenden Fragen zu antworten, entstand diese Arbeit. Deswegen möchte ich mich bei ihnen zuallererst herzlich bedanken und ihnen Optimismus wünschen.

Ein besonderes Dankeschön gebührt meinem Doktorvater Herrn Prof. Dr. Erich Perlwitz. Prof. Perlwitz stand für mich seit mehreren Jahren, in guten und schlechten sowohl akademischen als auch persönlichen Zeiten, ein vertrauensvoller Begleiter. Ich danke ihm für die Freiheit und das Vertrauen, die mir bei der Bearbeitung des Themas geschenkt wurden. Seine konstruktiv-kritische Impulse waren Anlass für eine für mich wertvolle Auseinandersetzung mit verschiedenen psychologischen Themen. Prof. Perlwitz konnte und kann meine Bewunderung für ihn sowohl als ein erfahrener und kompetenter Wissenschaftler als auch als ein toleranter, geduldiger und mit allen warmherziger Mensch aufwecken.

Frau Prof. Dr. Auckenthaler danke ich wärmstens für ihre Bereitschaft, diese Arbeit zu begutachten sowie für ihre hilfreiche Anmerkungen in der ersten Phase dieser Arbeit.

Während des ersten Entwurfs eines Untersuchungsplans waren auch die Kontakte zu Frau Prof. Dr. Triliva und Herrn Prof. Dr. Mellon aus der Universität in Kreta, Herrn Prof. Dr. Kolaitis aus der Universität in Athen und besonders zu Frau Prof. Dr. Stogiannidou wichtig. Frau Prof. Stogiannidou hat mich während der gesamten Promotionszeit auch über die räumliche Distanz zwischen Thessaloniki und Berlin Entfernung fachlich begleitet,

Ganz besonders möchte ich an dieser Stelle Herrn Prof. Dr. Peter Rossmann aus der Karl-Franzens-Universität Graz sowohl für seine Unterstützung bei der Adaption einer griechischen Version des DTK als auch für seine kritische und fördernde Anmerkungen hinsichtlich der Bearbeitung spezieller Fragenstellungen der Arbeit danken.

Für ihren Hilfe bei der Bildung und Stratifizierung der zu untersuchenden Stichprobe bin ich auch Dipl. Ingenieur C. Papachristou und Herrn Prof. Dr. Kioseoglou aus der Aristoteles Universität zu Thessaloniki dankbar. Letzteren gilt ein besonderer Dank für seine Beratung und Unterstützung bei der Arbeit mit dem Datensatz sowie für seine ständige Beratung in statistischen Fragen. Zudem möchte ich dem DAAD danken, der durch seine finanzielle Unterstützung die Durchführung dieser Studie unterstützt hat.

Einen lieben Dank auch an allen Personen, die mir bei der Übersetzung der in dieser Arbeit zunächst fremdsprachigen Verfahren beraten haben, nämlich an meinen guten Freunden K. Gkantas, F. Pistofidou, S. Tassi und T. Segovia sowie an Frau Prof. Dr. Bibou-Nakou, Frau Prof. Stogiannidou, beide aus der Aristoteles Universität zu Thessaloniki, und Dr. M. Casanovas. Letztere danke ich auch dafür, dass sie mir den Zugang zur spanischsprachigen Literatur über den ESDM ermöglicht hat.

Ausdrücklicher Dank gebührt Personen, die mich jahrelang begleiten, mein Leben mit Freundschaft, Liebe, Unterstützung und schöne Überraschungen bereichern:

Meinen Eltern und meinem Bruder, weil sie bedingungslos für mich immer da sind. Meinen lieben Freunden, die für mich eine wertvolle und stabile Selbstwertquelle bedeuten. Jeff Wright, meinem allerliebsten Partner, weil er mich glücklich macht.

Berlin, April 2008

Erklärung

Hiermit versichere ich, dass ich die vorgelegte Arbeit selbständig verfasst habe. Andere als die angegebenen Hilfsmittel habe ich nicht verwendet. Die Arbeit ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden.

Ort / Datum

Unterschrift