

## 7 Datengrundlage und methodisches Vorgehen

### 7.1 Stichprobe

Für die quantitativen Analysen in der vorliegenden Arbeit wird ein Teildatensatz aus dem Forschungsprojekt AIDA (Adaptation in der Adoleszenz) benutzt. AIDA wurde von Prof. Dr. Renate Valtin geleitet und von der DFG finanziert (vgl. Valtin et al., 2000; 2002; 2003; Valtin & Wagner, 2004). Der Datensatz wurde mir für die vorliegende Arbeit dankenswerter Weise zur Verfügung gestellt. AIDA ist das Folgeprojekt von NOVARA (Noten oder Verbalbeurteilung – Akzeptanz, Realisierung, Auswirkungen) und SABA (Schulische Adaptation und Bildungsaspiration). In den drei Projekten wird die schulische Leistungs- und Persönlichkeitsentwicklung von der Grundschule (Klasse 2) bis in die Sekundarstufe I (bis Klasse 9) an Schulen im Land Berlin untersucht (vgl. Valtin, 2002). Die erste Erhebung fand im Schuljahr 1994/95, die letzte quantitative Befragung im Jahr 2002 statt.

#### 7.1.1 Die Gesamtstichprobe

AIDA umfasst insgesamt Daten von drei Erhebungen in Klasse 7, 8 und 9 (vgl. Valtin et al., 2003). Zu jedem Messzeitpunkt wurden jeweils etwa 3000 bzw. mehr als 3000 Jugendliche erfasst. Tabelle 7.1 gibt eine Übersicht über die Teilnahme an den jeweiligen Messzeitpunkten nach Geschlecht, nach Schulstandort (Ost- und West-Berlin) sowie nach Schulform. Die Schüler/innen wurden jeweils gegen Ende des Schuljahres befragt.

Tabelle 7.1: Gesamtstichprobe der von AIDA befragten Schüler/innen

		Klasse 7	Klasse 8	Klasse 9
N		2928	3260	3342
	männlich	1377 (47,0%)	1546 (47,4%)	1601 (47,9%)
	weiblich	1551 (53,0%)	1714 (52,6%)	1740 (52,1%)
Schulstandort				
	Berlin-Ost	1689 (57,7%)	1766 (54,2%)	1873 (56,0%)
	Berlin-West	1239 (42,3%)	1494 (45,8%)	1469 (44,0%)
Schulform				
	Gymnasium	1441 (49,2%)	1533 (47,0%)	1471 (44,0%)
	Realschule	710 (24,3%)	856 (26,3%)	964 (28,8%)
	Hauptschule	196 (6,7%)	245 (7,5%)	273 (8,2%)
	Gesamtschule	581 (19,8%)	626 (19,2%)	634 (19,0%)

### 7.1.2 Die Untersuchungsstichprobe

Für die Bearbeitung der Fragestellungen der vorliegenden Arbeit, welche sowohl längsschnittliche als auch mehrebenenanalytische Auswertungen erfordern (vgl. Kapitel 6), ist die Bildung einer geeigneten Untersuchungsstichprobe notwendig. Ein Teil der in Klasse 7 und 8 befragten Schüler/innen nahm an beiden Messzeitpunkten teil. Über die beiden Messzeitpunkte lässt sich eine Panelstichprobe von  $N = 1977$  Jugendlichen legen. Diese Panelstichprobe steht im Mittelpunkt des eigenen Vorhabens. Sie reduziert sich allerdings beträchtlich, da in der vorliegenden Arbeit nur Schüler/innen berücksichtigt werden sollen, von denen durchgehend gültige Werte in den Items der Klima- und Persönlichkeitskonstrukte zu beiden Messzeitpunkten vorliegen und die zusammen mit mindestens vier weiteren Schüler/inne/n ihrer Klasse sowie mit mindestens einer weiteren Klasse ihrer Schule befragt wurden.<sup>1</sup>

Der in dieser Arbeit verwendete Datensatz besteht somit ausschließlich aus Schulen, von denen mindestens zwei Klassen befragt wurden, und ausschließlich aus Klassen, aus denen mindestens fünf Schüler/innen an der Befragung teilnahmen. Von jeder Schule wurden folglich mindestens zehn Schüler/innen befragt. Damit reduziert sich die verwendete Stichprobe auf  $N = 816$  Schüler/innen. Diese Reduzierung ist maßgeblich darauf zurückzuführen, dass viele Schüler/innen mit höchstens drei Mitschüler/inne/n ihrer Klasse befragt wurden und häufig nur eine Klasse einer Schule in die Befragung einbezogen wurde. Zudem mussten die Hauptschüler/innen aufgrund ihrer geringen Fallzahl von der Untersuchungsstichprobe ausgeschlossen werden. Die Stichprobenreduzierung wird allerdings mit dem Vorteil erkaufte, dass in der Untersuchungsstichprobe die analytische Trennung von Schul- und Klasseneffekten möglich ist. Diese Auswertungsmöglichkeit stellt eine unverzichtbare Voraussetzung für die Beantwortung der Fragestellungen der vorliegenden Arbeit dar.

---

<sup>1</sup> Die Mindestgröße von fünf Schüler/inne/n mit vollständigen Datensätzen pro Klasse mag gering erscheinen, ist aber nicht problematisch, da die durchschnittliche Klassengröße der berücksichtigten Klassen 12,75 Schüler/inne/n beträgt (vgl. dazu Gruehn, 2000). Auch Köller und Baumert (2001) erachten für Analysen auf Schulebene die Anzahl von mindestens zehn Schüler/inne/n pro Schule als hinreichend.

Tabelle 7.2: Untersuchungsstichprobe

N		816
	männlich	363 (44,5%)
	weiblich	453 (55,5%)
Schulstandort		
	Berlin-Ost	494 (60,5%)
	Berlin-West	322 (39,5%)
Schulform		
	Gymnasium	552 (67,6%)
	Realschule	119 (14,6%)
	Gesamtschule	145 (17,8%)

Ferner wurde versucht, eine Untersuchungsstichprobe über alle drei AIDA-Messzeitpunkte (Klasse 7, 8 und 9) zu bilden. N = 1500 Schüler/innen nahmen an allen drei Erhebungszeitpunkten teil. Reduziert man jedoch diese Panelstichprobe nach den gleichen Kriterien wie die hier verwendete Untersuchungsstichprobe (d.h. Ausschluss von Fällen mit fehlenden Werten und Berücksichtigung der hierarchischen Datenstruktur), gelangt man zu dem Ergebnis, dass quasi nur noch Gymnasiast/inn/en der Stichprobe angehören, da der Anteil der Schüler/innen anderer Schulformen beinahe verschwindet. Schulformeffekte können dann nicht mehr untersucht werden. Zu einem ähnlichen Resultat gelangt man bei der Bildung einer Längsschnittstichprobe von N = 1776 Schüler/inne/n, die lediglich die beiden Messzeitpunkte in Klasse 7 und Klasse 9 beinhaltet, denn nach der Reduzierung anhand der genannten Kriterien bleiben auch hier vorwiegend nur noch Gymnasiast/inn/en Bestandteil der Stichprobe.

Alternativ zu der hier gewählten Untersuchungsstichprobe ist es möglich, eine Panelstichprobe von N = 2339 Schüler/inne/n über die beiden letzten Messzeitpunkte zu legen (Klasse 8 und 9), welche nach Reduzierung noch N = 1024 Schüler/innen umfasst. Da in dieser Arbeit jedoch die Entwicklung der Schüler/innen in der Sekundarstufe verfolgt werden soll, die zeitlich möglichst eng an den Übergang von der Primar- zur Sekundarstufe anschließt, wird die Untersuchung der beiden ersten Messzeitpunkte (Klasse 7 und 8) bevorzugt. Das Arbeiten mit zwei unterschiedlichen Stichproben würde über den Rahmen der vorliegenden Arbeit hinausgehen und wird daher nicht vorgenommen.

### 7.1.3 Schulorganisatorische Merkmale

Bei der Untersuchung von Schulformunterschieden in der subjektiv wahrgenommenen Lernumwelt und in den Merkmalen der Schüler/innen bildet die Schulform eine Daten- und Analyseebene, während die vom AIDA-Fragebogen erhobenen Schülervariablen auf Individualdatenebene vorliegen. Hierarchisch betrachtet liegen darüber hinaus zwischen der Schulformebene und der Individualdatenebene weitere Ebenen vor, die bei der Fragestellung, ob Sekundarschulen bzw. Schulformen differenzielle Entwicklungsmilieus bilden, von Bedeutung sind: die Klassenebene und die Ebene der einzelnen Schule (vgl. Kapitel 3). Die Klassenebene erlangt auch im Zusammenhang mit dem Klassenklima Bedeutung (vgl. Kapitel 4). Tabelle 7.3 zeigt die Zusammensetzung der Schüler/innen in der Untersuchungsstichprobe nach Klassen und Schulen, gegliedert nach Schulformen. Auf den Standort der Schulen wurde bereits mit der Unterteilung in Ost- und West-Berlin eingegangen; eine Gliederung nach Stadtteilen folgt im nachfolgenden Abschnitt. Die 816 Schüler/innen der Untersuchungsstichprobe kommen aus 64 Klassen. Mindestens fünf und höchstens 24 Schüler/innen einer Klasse wurden befragt. Im Durchschnitt wurden 12,75 Schüler/innen einer Klasse befragt. 14 Schulen mit zwei, acht Schulen mit drei und drei Schulen mit vier Klassen bilden auf Schul- und Klassenebene die Untersuchungseinheiten.

Tabelle 7.3: Anzahl der Schüler/innen pro Klasse und Schule

	Gymnasium	Realschule	Gesamtschule	Gesamt
Klassen	38	11	15	64
Schulen	14	5	6	25
<i>Anzahl der Schüler/innen pro Klasse</i>				
minimale Anzahl der Schüler/innen	5	5	5	5
maximale Anzahl der Schüler/innen	24	20	17	24
Mittelwert	14,53	10,82	9,67	12,75
Standardabweichung	5,12	4,75	4,13	5,26
<i>Anzahl der Schüler/innen pro Schule</i>				
minimale Anzahl der Schüler/innen	23	13	11	11
maximale Anzahl der Schüler/innen	58	41	34	58
Mittelwert	39,43	23,80	24,17	32,64
Streuung	11,80	12,79	8,04	13,32

### 7.1.4 Stadtbezirk

Da sich die Stadtteile Berlins durch unterschiedliche Bevölkerungsstrukturen auszeichnen, ist es sinnvoll, die Verteilung der Schüler/innen auf die einzelnen Stadtteile, in denen ihre Schule liegt, zu erwähnen. Tabelle 7.4 gibt einen Überblick. Knapp ein Drittel (31,9%) der befragten Schüler/innen besuchten eine Schule in Tempelhof, ein Stadtteil in West-Berlin, der zu den Außenstadtteilen zählt. Schüler/innen von drei Gymnasien, drei Realschulen und zwei Gesamtschulen wurden dort befragt. Die Innenstadtteile (Friedrichshain, Prenzlauer Berg, Neukölln, Mitte) stellen ein Fünftel (19,9%) der Untersuchungsstichprobe und sind mit Friedrichshain (12,1%) am stärksten vertreten. Erwähnenswert ist, dass nur zwei Schulen aus Berliner Stadtteilen mit hohem Bevölkerungsanteil nicht-deutscher Herkunft – eine Gesamtschule in Neukölln und ein Gymnasium in Mitte – Bestandteil der Untersuchungsstichprobe sind. Die Anzahl der Schüler/innen, die diese beiden Schulen besuchten, machen 4,5% der Stichprobe aus. Somit ist nicht auszuschließen, dass der Anteil von Schüler/innen mit Migrationshintergrund in dieser Stichprobe eher klein ist.<sup>2</sup>

Tabelle 7.4: Anzahl der Schulen und Schüler/innen aufgliedert nach Stadtteil

	Anzahl der Sekundarschulen			Gesamt	Anzahl der Schüler/innen
	Gymnasium	Realschule	Gesamtschule		
Berlin-Ost					
Friedrichshain	2			2	99 (12,1%)
Hellersdorf	1	1	1	3	104 (12,7%)
Hohenschönhausen			1	1	19 (2,3%)
Lichtenberg	1			1	48 (5,9%)
Marzahn		1		1	16 (2,0%)
Mitte	1			1	26 (3,2%)
Pankow	2			2	69 (8,5%)
Prenzlauer Berg			1	1	27 (3,3%)
Weißensee	2			2	86 (10,5%)
Berlin-West					
Neukölln			1	1	11 (1,3%)
Steglitz	2			2	51 (6,3%)
Tempelhof	3	3	2	8	260 (31,9%)
<b>Gesamt</b>	<b>14</b>	<b>5</b>	<b>6</b>	<b>25</b>	<b>816 (100,0%)</b>

<sup>2</sup> In AIDA wurde keine Variable zum Migrationshintergrund der Schüler/innen erhoben.

### 7.1.5 Drop-out-Analyse

Längsschnittliche Untersuchungsdesigns bringen, insbesondere wenn die einzelnen Messzeitpunkte längere Zeit – z. B. ein Jahr – auseinander liegen, Ausfallquoten der befragten Personen mit sich, was mit „Mortalität“ der Stichprobe bezeichnet wird. Auch im AIDA-Datensatz liegt eine Ausfallquote vor: Nicht alle Schüler/innen wurden sowohl in Klasse 7 als auch in Klasse 8 befragt. Hinzu kommt die für die Fragestellungen der vorliegenden Arbeit notwendige Stichprobenreduzierung (vgl. Abschnitt 7.1.2), die möglicherweise eine positive Auslese der Schüler/innen begünstigt. Somit stellt sich die Frage, inwieweit sich die Untersuchungsstichprobe von der Gesamtstichprobe in zentralen Merkmalen unterscheidet. Maßnahmen der internen Selektion im Schulsystem wie Wiederholung eines Schuljahres oder Wechsel zu einer anderen Schulform führen z. B. dazu, dass ein Teil der Schüler/innen aufgrund ihrer schwachen schulischen Leistung aus einer längsschnittlich angelegten Stichprobe herausfällt (vgl. Eder, 1988). Daher erscheint es sinnvoll, die Untersuchungsstichprobe mit der Gesamtstichprobe in einer Drop-out-Analyse zu vergleichen. Tabelle 7.5 zeigt die Ergebnisse des Vergleichs.

Tabelle 7.5: Drop-Out-Analyse der Untersuchungsstichprobe und der Gesamtstichprobe

	Gesamtstichprobe		Untersuchungsstichprobe
	Klasse 7	Klasse 8	
<i>N</i>	2928	3260	816
männlich	1377 (47,0%)	1546 (47,4%)	363 (44,5%)
weiblich	1551 (53,0%)	1714 (52,6%)	453 (55,5%)
<i>Bildungsgrad der Eltern</i>			
Abitur Mutter	856 (29,2%)	1063 (32,6%)	302 (37,0%)
Abitur Vater	790 (27,0%)	994 (30,5%)	289 (36,5%)
Studium Mutter	931 (31,8%)	1094 (33,6%)	313 (38,4%)
Studium Vater	869 (29,7%)	1039 (31,9%)	302 (37,0%)
<i>Schulform</i>			
Gymnasium	1441 (49,2%)	1533 (47,0%)	552 (67,6%)
Realschule	710 (24,3%)	856 (26,3%)	119 (14,6%)
Hauptschule	196 (6,7%)	245 (7,5%)	0 (0,0%)
Gesamtschule	581 (19,8%)	626 (19,2%)	145 (17,8%)
<i>Gute oder sehr gute Schulnoten<sup>a)</sup></i>			
Deutsch (mündlich)	751 (25,6 %)	837 (25,7%)	235/257 (28,8/31,5%)
Deutsch (schriftlich)	533 (18,2%)	683 (20,9%)	186/201 (22,8/24,6%)
Mathematik	698 (23,8%)	831 (25,5%)	251/260 (30,8/31,9%)

Anmerkungen:

<sup>a)</sup> Für die Untersuchungsstichprobe vor dem Schrägstrich für Klasse 7 und nach dem Schrägstrich für Klasse 8.

Erwartungsgemäß stellt die Untersuchungsstichprobe verglichen mit der Gesamtstichprobe in Klasse 7 und 8 eine positive Auslese dar: Der Bildungsgrad der Eltern ist in der Untersuchungsstichprobe höher als in der Gesamtstichprobe; Gymnasiast/inn/en sind in der Untersuchungsstichprobe stärker und Realschüler/innen sind schwächer vertreten als in der Gesamtstichprobe; der Anteil guter oder sehr guter Schüler/innen in den Fächern Deutsch und Mathematik ist in der Untersuchungsstichprobe größer als in der Gesamtstichprobe. Hinsichtlich des Bildungsgrads der Eltern muss allerdings beachtet werden, dass es sich dabei lediglich um eine Angabe der Schüler/innen handelt und viele Schüler/innen zu beiden Zeitpunkten keine Angabe zum Bildungsgrad ihrer Eltern gemacht haben bzw. machen konnten (vgl. auch Abschnitt 7.2.3). Gleichfalls können die Schulnoten nur eingeschränkt einen Anhaltspunkt für die Ausfallquoten liefern, da vor allem in der Gesamtschulstichprobe für einen beträchtlichen Anteil der Schüler/innen keine gültigen Werte vorliegen.

#### **7.1.6 Anmerkung zur Untersuchungsstichprobe**

Hinsichtlich der Untersuchungsstichprobe sind zwei Probleme zu nennen:

- (1) Die Untersuchungsstichprobe stellt aus mehreren Gründen (z. B. aufgrund ihrer Reduzierung) keine Zufallsstichprobe dar. Von einer lehrbuchgerechten Repräsentativität der Stichprobe kann also nicht ausgegangen werden.
- (2) Die Gymnasiast/inn/en sind in der Stichprobe wesentlich stärker als die Real- und Gesamtschüler/innen vertreten. Ein „balanciertes Design“, in dem die Schulformgruppen gleich groß sind, liegt daher nicht vor.

ad (1): Grundsätzlich lässt sich in empirischen Studien der pädagogischen Feldforschung die Gewinnung einer Zufallsstichprobe nur schwer gewährleisten (vgl. Lange et al., 1983; Bortz, 2005; Merckens, 1996). Denn in der Regel bedarf es großer Mühe und Überredungskunst, die Beteiligten (Schüler/innen, Lehrer/innen, Eltern, Schulleiter/innen) zur Teilnahme an einer Längsschnittsstudie zu gewinnen. Dieser Umstand konnte bei der Erhebung der AIDA-Daten z. B. dazu führen, dass verschiedene Schüler/innen nur zu einem Messzeitpunkt an der Befragung teilnahmen oder nicht mehr als eine Klasse von einer Schule berücksichtigt werden konnte. Mit der Stichprobenreduzierung mussten dann solche Schüler/innen und Klassen von der Untersuchungsstichprobe systematisch ausgeschlossen werden.

ad (2): Bestimmte Auswertungsverfahren (z. B. die Varianzanalyse) setzen streng genommen gleiche Zellenverteilung voraus, stellen allerdings bei größeren Stichproben ( $n > 100$ ) auch bei ungleicher Zellenverteilung recht robuste Verfahren dar (vgl. Bortz, 2005). So ist es fraglich, ob ein balanciertes Design für die Anwendung bestimmter Auswertungsverfahren

unbedingt erforderlich ist. Darüber hinaus zeigt ein Blick in die Berliner Schulstatistik (Senatsverwaltung Berlin, 2005), dass sich die Schüler/innen/schaften ohnehin nicht gleichmäßig auf die verschiedenen Schularten verteilen: In den Schuljahren 1999/2000 und 2000/20001 besuchten die Berliner Schüler/innen der Jahrgangsstufen 7 bis 10, dazu gehören auch die Schüler/innen der Untersuchungsstichprobe, vorrangig das Gymnasium (mehr als 45.000 Schüler/innen) oder die Gesamtschule (mehr als 40.000 Schüler/innen) und deutlich weniger die Realschule (etwa 30.000 Schüler/innen) sowie weitaus weniger die Hauptschule (etwa 15.000 Schüler/innen). Eine Balancierung der Stichprobe nach Schulformen, die für die Schüler/innen/schaften der einzelnen Schulformen gleiche Zellenverteilungen erzielen wollte, würde also der Frage nach Repräsentativität grundsätzlich entgegenstehen.

Letztlich muss davon ausgegangen werden, dass die Aussagekraft der Ergebnisse der vorliegenden Arbeit eingeschränkt ist. Dies betrifft sowohl die Generalisierbarkeit der Befunde als auch die Überprüfung von Hypothesen. Allein aufgrund der Stichprobenzusammensetzung sollten die im empirischen Teil dieser Arbeit referierten Ergebnisse mit Vorsicht interpretiert werden und vorwiegend als Grundlage für die Formulierung von Thesen genutzt werden, die möglicherweise Anhaltspunkte für zukünftige Forschungsvorhaben darstellen.

Ferner sei noch erwähnt, dass Merrens (1996) auf einen interessanten Umgang mit Stichproben verweist, denen es an Repräsentativität mangelt. Die „gelegentliche Stichprobe“ ist oft in der Praxis der empirischen Sozialforschung anzutreffen, da das Ziehen repräsentativer oder geschichteter Stichproben aus verschiedenen Gründen nur selten möglich ist. Unter Rekurs auf Poppers Ansatz (vgl. Popper, 2005) stellt Merrens (1996) den Vorzug dieses Stichprobentyps heraus, wenn die Stichprobe strategisch genutzt wird: Werden mit gelegentlichen Stichproben Replikationsstudien durchgeführt, kann dadurch die Wahrscheinlichkeit der Falsifikation einer Hypothese erhöht werden. Voraussetzung dafür ist allerdings, dass entsprechend geeignete und überprüfte Instrumente zum Einsatz kommen, die dem Anspruch der Replikation auch gerecht werden.

## 7.2 Erhebungsinstrumente

Neben den demografischen Daten zur Schulform-, Schul- und Klassenzugehörigkeit, welche von den die Befragung durchführenden Personen dokumentiert wurden, besteht der in der vorliegenden Arbeit verwendete Datensatz vor allem aus Selbstberichtsdaten der Schüler/innen, welche mit dem AIDA-Fragebogen „Schule aus der Sicht der Schülerinnen und Schüler“ (vgl. Valtin et al., 2000; 2002; 2003), einem standardisierten Fragebogen, jeweils am Ende des Schuljahres erhoben wurden. Um die Persönlichkeitsmerkmale zu erfassen, kamen aus der Literatur bekannte und bewährte Instrumente in zum Teil verkürzter Form zum Einsatz. Für die Erhebung des Klimas wurden Items aus unterschiedlichen Verfahren in zum Teil sprachlich veränderter Form benutzt. Items und Skalen, die in dieser Arbeit längsschnittlich ausgewertet werden, wurden zu den betreffenden Messzeitpunkten unverändert eingesetzt. Im Folgenden werden zunächst die Erhebungsinstrumente inhaltlich dargestellt und beschrieben. Im nachfolgenden Unterkapitel 7.3 werden dann die Reliabilitäten der Skalen und die Ergebnisse der Faktorenanalysen vorgestellt.

### 7.2.1 Klassenklima

Kennzeichnend für die Schul- und Klassenklimaforschung im deutschsprachigen Raum ist, dass in den vergangenen drei Jahrzehnten immer wieder neue und damit recht unterschiedliche Erhebungsinstrumente für die jeweiligen Untersuchungen entwickelt wurden (vgl. Gruehn, 2000 und Abschnitt 4.2.2 dieser Arbeit). Folglich stellt sich die Frage nach der möglichen Vergleichbarkeit von Befunden, die in den verschiedenen Untersuchungen durch den Einsatz unterschiedlicher Instrumente erzielt werden konnten. Beispielsweise geben Dreesmann et al. (1992) als bis dahin erzielte kumulative Befunde zur Auswirkung des Klimas auf Persönlichkeitsmerkmale der Schüler/innen lediglich die Variable Leistungsangst an, obwohl zu diesem Zeitpunkt fast alle der in Kapitel 4 dieser Arbeit dargestellten Klimainstrumente entwickelt und eingesetzt worden waren und somit eine größere Zusammenschau vergleichbarer Ergebnisse zu erwarten gewesen wäre.

Auch die vorliegende Arbeit steht vor dem Problem, verschiedene Items aus unterschiedlichen Instrumenten kombinieren zu müssen, da in AIDA zur Erfassung des Schul- und Klassenklimas Items aus ganz unterschiedlichen Klima-Verfahren eingesetzt wurden. Dementsprechend wurde ein datengeleitetes Vorgehen gewählt, indem mittels explorativer Faktorenanalysen die Items auf zugrunde liegende Dimensionen untersucht wurden. In der AIDA-Gesamtstichprobe konnten dadurch zwei Klimaskalen mit zufrieden stellender

Reliabilität gebildet werden, die den Dimensionen „Engagement der Lehrer/innen“ und „Konkurrenzorientierung“ zugeordnet werden können. Ferner macht eine Durchsicht der Items verschiedener Erhebungsverfahren deutlich, dass eine gewisse Übereinstimmung der in AIDA verwendeten Items mit jenen Items verschiedener Verfahren besteht, die sich ebenfalls diesen beiden Klimaaspekten zuordnen lassen. Im Folgenden werden nun zunächst die Items inhaltlich vorgestellt und die Verbindung zu den Instrumenten aus anderen Verfahren aufgezeigt, bevor im nachfolgenden Abschnitt 7.3 die Ergebnisse der Faktorenanalysen an der Untersuchungsstichprobe dargestellt werden.

### 7.2.1.1 Engagement der Lehrer/innen

Tabelle 7.6 zeigt die in AIDA verwendeten Items, welche die Skala „Engagement der Lehrer/innen“ in der vorliegenden Arbeit bilden. Der Fragebereich zur Erfassung des wahrgenommenen Verhaltens der Lehrer/innen wurde im AIDA-Fragebogen mit der Frage „Was denkst du über deine Lehrerinnen und Lehrer?“ eingeleitet und zur Beantwortung wurde ein vierstufiges Antwortformat vorgegeben („Diese Aussage trifft zu für alle – viele – wenige – keine(n)“). Tabelle 7.7 macht deutlich, dass sich die größte Übereinstimmung mit vergleichbaren Skalen, die in anderen Untersuchungen verwendet wurden, im zweiten und dritten Item zeigt. Das Engagement der Lehrer/innen lässt sich demnach wohl am besten mit der in PISA verwendeten Skalenbezeichnung umschreiben: Interesse der Lehrer/innen am Lernen und an der Person des Schülers/der Schülerin.

Tabelle 7.6: Items der Skala „Engagement der Lehrer/innen“

- 
1. Unsere Lehrerinnen und Lehrer verstehen Spaß.<sup>a)</sup>
  2. Unsere Lehrerinnen und Lehrer haben Verständnis für unsere persönlichen Probleme.<sup>b)</sup>
  3. Unsere Lehrerinnen und Lehrer interessiert, dass wir wirklich etwas lernen.<sup>c)</sup>
  4. Unsere Lehrerinnen und Lehrer behandeln uns gerecht.<sup>d)</sup>
- 

Anmerkungen:

<sup>a)</sup> Dieses Item heißt bei Pekrun (1985): „Unsere Lehrer können Spaß verstehen.“

<sup>b)</sup> Dieses Items heißt bei Fend (1977) und bei Helmke (1983): „Ich glaube, die meisten unserer Lehrer haben wenig Verständnis für unsere persönlichen Probleme.“ Bei Holtappels (2003b) heißt es: „Die meisten Lehrer/innen kümmern sich nicht um persönliche Probleme der Schüler/innen“, und bei von Saldern und Littig (1987) heißt ein Item der Skala „Fürsorglichkeit des Lehrers“: „Unser Lehrer kümmert sich um unsere Probleme“.

<sup>c)</sup> Dieses Item heißt bei Fend (1977): „Unsere Lehrer bemühen sich sehr, dass alle Schüler wirklich etwas lernen.“ Bei von Saldern und Littig (1987) heißt ein Item der Skala „Fürsorglichkeit des Lehrers“: „Unser Lehrer hilft jedem von uns, der Schwierigkeiten mit seinen Aufgaben hat.“

<sup>d)</sup> Dieses Item gehört im LFSK (Eder, 1996) der Dimension „Gerechtigkeit“ an, wird jedoch auch der Kategorie Lehrer/in-Schüler/in-Beziehung zugeordnet. Ein Item dieser Dimension heißt z. B.: „Einige Schüler werden immer wieder von den Lehrern bevorzugt“. Bei Holtappels (2003b) heißt es: „Die Lehrer/innen hier behandeln nicht alle Schüler/innen gerecht“.

Tabelle 7.7: Skalen zum wahrgenommenen pädagogischen Engagement der Lehrer/innen und Übereinstimmung der Items mit den in AIDA verwendeten Items

	Name der Skala	Mit AIDA übereinstimmende Items <sup>a)</sup>				Anzahl der Items in der jew. Skala
		1	2	3	4	
1.	„Lehrerengagement“ (Fend, 1977)	-	(x)	(x)	-	7
2.	„Zuwendung“ (Helmke, 1983)	-	(x)	-	-	10
3.	„Unterstützung“ (Pekrun, 1983)	(x)	-	-	-	10
4.	„Fürsorglichkeit“ (von Saldern & Littig, 1987)	-	(x)	(x)	-	8
5.	„Lehrer-Schüler-Beziehung“ (Holtappels, 2003b)	-	(x)	-	(x)	8
6.	„Interesse des Lehrers am Lernen und der Person“ (PISA 2000, Kunter et al., 2002)	-	x	x	-	6

Anmerkungen:

<sup>a)</sup> Die Ziffern beziehen sich auf die Item-Nummerierung in Tabelle 7.6

x = identisch mit den in AIDA verwendeten Items; (x) = Wortlaut sehr ähnlich (vgl. Anmerkungen bei Tabelle 7.6); - = keine Übereinstimmung.

### 7.2.1.2 Konkurrenzorientierung

Die Items zum wahrgenommenen Konkurrenzverhalten zwischen den Schüler/inne/n einer Klasse stammen von Fend et al. (1976) und wurden auch in der Erhebung von 1977 von Fend und Mitarbeitern (vgl. Helmke, 1983) sowie – zum Teil leicht verändert – von der Arbeitsgruppe um Schwarzer (vgl. Lange et al., 1983; Jerusalem, 1984) und von Pekrun (1983; 1985a) mit der Bezeichnung „Wettbewerb“ oder „Konkurrenz“ eingesetzt. In AIDA wurden drei Items der Fendschen Skala exakt übernommen und um ein weiteres Item ergänzt (Tabelle 7.8). Folglich herrscht eine weitgehende Übereinstimmung der in AIDA verwendeten Skala mit den in anderen Untersuchungen eingesetzten Skalen (vgl. Tabelle 7.9). Eingeleitet wurde der Fragebereich mit der Frage „Was denkst du über deine Schule und über deine Klasse?“ und zur Beantwortung wurde ebenfalls ein vierstufiges Antwortformat vorgegeben („ja, trifft voll und ganz zu“ – „trifft eher zu“ – „trifft eher nicht zu“ – „nein, trifft überhaupt nicht zu“).

Tabelle 7.8: Items der Skala „Konkurrenzorientierung“

1. In unserer Klasse sieht jeder nur auf seinen eigenen Vorteil, wenn es um die Noten geht.
2. Bei uns hat man manchmal das Gefühl, dass sich Schüler untereinander keine guten Noten gönnen.<sup>a)</sup>
3. In unserer Klasse muss man ein wenig vorsichtig sein, zuviel für die Schule zu tun, weil die anderen sich sonst leicht darüber lustig machen.
4. In unserer Klasse sieht jeder Schüler im anderen den Konkurrenten.<sup>b)</sup>

Anmerkungen:

<sup>a)</sup> Dieses Item heißt bei Holtappels (2003b): „Viele Schüler/innen sind hier manchmal neidisch, wenn ein anderer bessere Leistungen hat als sie“ und bei PISA 2000: „Viele Schüler/innen sind manchmal neidisch, wenn ein anderer bessere Leistungen hat als sie.“

<sup>b)</sup> Dieses Item heißt in der Skala „Wettbewerb“ von Pekrun (1983) und Lange et al. (1983): „In meiner Klasse steht jeder im Wettbewerb mit dem anderen“. Bei Holtappels (2003b) heißt es: „In unserer Klasse versucht unter den Schüler/inne/n jeder besser zu sein als der andere“. Bei Eder (1998) heißt ein Item der Skala „Rivalität“: „Bei uns arbeiten die einzelnen Schüler eher gegeneinander als miteinander“.

Tabelle 7.9: Skalen zur wahrgenommenen Konkurrenzorientierung und Übereinstimmung der Items mit den in AIDA verwendeten Items

	Name der Skala	Mit AIDA übereinstimmende Items				Anzahl der Items in der jew. Skala
		1	2	3	4	
1.	„Wettbewerb“ (Pekrun, 1983)	x	x	-	(x)	9
2.	„Konkurrenz“ (Fend et al., 1976; Helmke, 1983)	x	x	-	x	6
3.	„Wettbewerb“ (Lange et al., 1983)	x	x	-	(x)	8
4.	„Konkurrenz“ (Holtappels, 2003b)	x	(x)	-	(x)	3
5.	Konkurrenz und soziale Desintegration (PISA 2000, Kunter et al., 2002)	x	(x)	-	-	5
6.	„Rivalität“ (Eder, 1996)	-	-	-	(x)	6

Anmerkungen:

<sup>a)</sup> Die Ziffern beziehen sich auf die Item-Nummerierung in Tabelle 7.8

x = identisch mit den in AIDA verwendeten Items; (x) = Wortlaut sehr ähnlich (vgl. Anmerkungen zu Tabelle 7.8); - = keine Übereinstimmung.

## 7.2.2 Selbstkonzept und Attribuierung

Um das Selbstkonzept der Begabung und die Attribuierung schulischer Misserfolge zu erfassen, wurden aus der Literatur bekannte und bewährte Instrumente übernommen, die allerdings zum Teil in gekürzter Fassung zum Einsatz kamen. In AIDA wurde eine ganze Reihe von Selbstkonzept-Variablen erhoben (neben dem Selbstkonzept der Begabung der Selbstwert, das Selbstkonzept des Aussehens und die Selbstakzeptanz durch andere), die gemeinsam einen Fragebereich im Fragebogen bildeten und mit der Frage „Wie schätzt du dich selber ein?“ eingeleitet wurden.<sup>3</sup> Vorgegeben war ein vierstufiges Antwortformat („ja,

<sup>3</sup> Die Items der unterschiedlichen Selbstkonzeptkonstrukte waren in diesem Fragebereich in einer (intendiert) ungeordneten Reihenfolge aufgeführt.

trifft voll und ganz zu“ – „trifft eher zu“ – „trifft eher nicht zu“ – „nein, trifft überhaupt nicht zu“). Die Items zum Selbstkonzept der Begabung (vgl. Tabelle 7.10) stammen aus dem gleichnamigen Fragebogen von Meyer (1972). Es handelt sich um Items, die eine eher generelle Einschätzung der eigenen intellektuellen Fähigkeiten erfassen, wobei alle Aussagen explizit eine soziale Vergleichsperspektive beinhalten. Dies erscheint bemerkenswert, da sich die Items anderer Selbstkonzept-Skalen, z. B. jene der von Eder (1996) oder Pekrun (1983; 1985a) verwendeten Skala „Allgemeines Fähigkeits-Selbstkonzept“, ausschließlich auf die eigene Person beziehen und frei von unmittelbaren Aussagen zu einer sozialen Vergleichsperspektive sind.

Tabelle 7.10: Items der Skala „Selbstkonzept der Begabung“

- 
1. Manchmal fühle ich mich anderen überlegen und glaube, dass sie manches von mir lernen könnten.
  2. Sehr viele Dinge kann ich besser als andere in meinem Alter.
  3. Ich habe den Eindruck, dass ich klüger bin als andere in meinem Alter.
- 

Die Attribuierung schulischer Misserfolge wurde von AIDA mit Items erhoben, die den Skalen von Helmke (1985; 1992a) entstammen und sich den vier Gründen „Aufmerksamkeit“, „Anstrengung“, „Lehrer“ und „Fähigkeit“ zuordnen lassen. Zusätzlich wurden drei Items eingesetzt, die im Rahmen der BIJU-Studie ergänzend zu den Skalen von Helmke konstruiert worden waren und sich der Ursache „Zufall“ zuordnen lassen (vgl. Köller, 1998). Die fünfzehn Items zur Attribuierung schulischer Misserfolge bildeten einen Fragebereich im AIDA-Fragebogen und wurden mit der Frage „Woran liegt es, wenn du in einer Arbeit nicht gut bist?“ eingeleitet.<sup>4</sup> Auch hier war ein vierstufiges Antwortformat vorgegeben („ja, trifft voll und ganz zu“ – „trifft eher zu“ – „trifft eher nicht zu“ – „nein, trifft überhaupt nicht zu“). In der vorliegenden Arbeit werden lediglich die beiden Skala „Fähigkeit“ als stabil-internales und „Zufall“ als unkontrollierbar-externales Attribuierungsmuster berücksichtigt (vgl. Tabelle 7.11).

---

<sup>4</sup> Gleichfalls waren in diesem Fragebereich die Items nicht nach den fünf Dimensionen angeordnet, sondern in einer (intendiert) ungeordneten Reihenfolge aufgeführt

Tabelle 7.11: Attribuierung schulischer Misserfolge auf Fähigkeit und Zufall

Mangelnde eigene Fähigkeit (Helmke, 1985; 1992)
1. Ich kann vieles nicht behalten.
2. Für die Schule bin ich einfach nicht so begabt.
3. Manche Sachen wollen mir einfach nicht in den Kopf.
Zufall (BIJU, vgl. Köller, 1998)
1. Es hängt einfach zu sehr vom Zufall ab, ob man etwas versteht oder eine Arbeit gelingt.
2. Es ist meistens Glücksache, ob man etwas versteht oder eine Arbeit gelingt.
3. Ich habe einfach Pech gehabt.

### 7.2.3 Weitere Variablen

Neben der Schulform-, der Schul- und der Klassenzugehörigkeit sowie dem Geschlecht der Schüler/innen wird in der vorliegenden Arbeit der Bildungsgrad der Eltern verwendet. Dieser kann als Indikator für den sozio-ökonomischen Status der Schüler/innen interpretiert werden. Sowohl in Klasse 7 als auch in Klasse 8 wurden die Schüler/innen zum Bildungsgrad ihrer Eltern befragt, indem sie Angaben zum Schul- und Hochschulabschluss ihrer Eltern machen sollten. In der vorliegenden Arbeit werden lediglich die Angaben der Schüler/innen zum Hochschulabschluss ihrer Eltern verwendet (vgl. Abbildung 7.1).<sup>5</sup> Da anzunehmen ist, dass die Angaben von Jugendlichen zum Bildungsgrad ihrer Eltern unpräzise sein können, wurden die Angaben zu den unterschiedlichen Messzeitpunkten berücksichtigt.<sup>6</sup> Letztlich wurde eine dichotomisierte Variable gebildet, die angibt, dass entweder mindestens ein Elternteil studiert hat (kodiert als 1) oder sowohl Mutter als auch Vater nicht studiert haben bzw. die Schüler/innen keine Kenntnis darüber besitzen (kodiert als 0).

Abbildung 7.1: Erfassung des Hochschulabschlusses der Eltern in AIDA

<b>Haben deine Eltern studiert?</b>			
<i>Kreuze bitte an, was für deine Mutter und was für deinen Vater zutrifft.</i>			
Mutter		Vater	
<input type="checkbox"/>	ja	<input type="checkbox"/>	ja
<input type="checkbox"/>	nein	<input type="checkbox"/>	nein
<input type="checkbox"/>	weiß nicht	<input type="checkbox"/>	weiß nicht

<sup>5</sup> Zum Hochschulabschluss der Eltern konnten die Schüler/innen insgesamt genauere Angaben machen als zum Schulabschluss ihrer Eltern.

<sup>6</sup> Nur wenn die Schüler/innen zu beiden Zeitpunkten die Frage für das jeweilige Elternteil bejahten, wurde die Angabe zum entsprechenden Elternteil in der vorliegenden Arbeit verwendet.

## 7.3 Faktorenanalysen

### 7.3.1 Klima

In Voranalysen wurden mit allen Klima-Items des AIDA-Datensatzes explorative Faktorenanalysen an der Gesamtstichprobe durchgeführt, in denen die beiden Skalen „Engagement der Lehrer/innen“ und „Konkurrenzorientierung“ herausgestellt werden konnten. Um die Güte der beiden Klimaskalen für die vorliegende Arbeit zu überprüfen, wurden Item- sowie Faktorenanalysen an der Untersuchungsstichprobe durchgeführt. Auf die Darstellung der Ergebnisse der Voranalysen zur Exploration der AIDA-Items wird im Folgenden verzichtet, da sie keinen wesentlichen Informationsgehalt besitzen, den nicht auch die Analysen zur faktorenanalytischen Struktur an der Untersuchungsstichprobe bieten. Die Itemkennwerte (Mittelwerte und Trennschärfen) sind im Anhang dieser Arbeit (vgl. Anhang 1) aufgeführt. Sie sind für alle Items zu beiden Messzeitpunkten akzeptabel oder zufriedenstellend, so dass kein Item von den weiteren Analysen ausgeschlossen werden muss.

#### 7.3.1.1 Explorative Faktorenanalysen

Für die explorativen Faktorenanalysen zum Klassenklima wurden allein die im vorhergehenden Unterkapitel 7.2 dargestellten Items zum wahrgenommenen Engagement der Lehrer/innen und zur wahrgenommenen Konkurrenzorientierung berücksichtigt. Zur Festlegung der Hauptkomponentenzahl wurde das Kaiser-Kriterium festgelegt, so dass alle Faktoren mit einem Eigenwert größer als eins extrahiert wurden. Für die beiden Messzeitpunkte wurden getrennt Faktorenanalysen mit orthogonaler Rotation (Varimax) durchgeführt, und zwar zunächst in Gesamtanalysen mit den Rohdaten und daraufhin auf Individualebene mit gruppenzentrierten Daten (within-analysis) und auf Klassenebene mit den Mittelwerten auf Klassenebene (between-analysis).<sup>7</sup> Auf Individualebene ( $N = 816$ ) können mit den beiden Komponenten gut über 40% erklärt werden, auf Klassenebene ( $N_{\text{Klasse}} = 64$ ) ergeben sich Lösungen von über 60%. Tabelle 7.12 zeigt die Ladungen der berücksichtigten Items auf den Hauptkomponenten, wobei nur substantielle Ladungen aufgeführt sind ( $\alpha \geq .4$ , vgl. Backhaus et al., 1994), sowie die erklärte Gesamtvarianz der jeweiligen Faktorenanalyse.

---

<sup>7</sup> Analysen „innerhalb“ (within-analysis) sind Analysen auf Individualdatenebene, sie unterscheiden sich jedoch von der Gesamtanalyse darin, dass die Daten um den jeweiligen Klassenmittelwert zentriert sind. Dadurch werden Klassen-, Schul- und Schulformunterschiede konstant gehalten und die Berechnungen zwischen den Schüler/innen innerhalb der Klassen durchgeführt. Die Analysen „zwischen“ (between-analysis) sind dann Analysen auf Klassenebene mit den aggregierten Klassenmittelwerten. In Abschn. 7.4.3 wird darauf noch genauer eingegangen.

Tabelle 7.12: Ergebnisse der Hauptkomponentenanalysen zum Klassenklima

	Klasse 7		Klasse 8	
	Hauptkomponenten		Hauptkomponenten	
	1	2	1	2
LE01	.76/.74/.81		.74/.72/.81	
LE02	.72/.71/.80		.70/.68/.74	
LE03	.72/.69/.81		.75/.74/.84	
LE04	.66/.65/.70		.72/.70/.78	
KO01		.60/.56/.76		.71/.68/.81
KO02		.73/.72/.83		.78/.78/.81
KO03		.56/.56/.65		.60/.59/.67
KO04		.69/.70/.75		.71/.67/.83
Erklärte Gesamtvarianz (%)	48/45/61		52/49/67	

Anmerkungen: Ladungsmatrizen der Faktorenanalysen mit anschließender Varimax-Rotation mit den Rohdaten vor dem ersten, auf Individualebene (within-analysis) nach dem ersten und auf Klassenebene (between-analysis) nach dem zweiten Schrägstrich. Abkürzungen: LE – Engagement der Lehrer/innen, KO – Konkurrenzorientierung. Bzgl. der Itemtexte vgl. Tabellen 7.6 und 7.8.

### 7.3.1.2 Konfirmatorische Faktorenanalysen

Die Überprüfung der faktoriellen Struktur in der Untersuchungsstichprobe erfolgte als konfirmatorische Faktorenanalyse mit AMOS 5. Dazu wurden zwei Faktoren spezifiziert, welche die Ausprägung der dazugehörigen Items zu einem Messzeitpunkt erklären sollten: Die beiden Klimaaspekte „Engagement der Lehrer/innen“ und „Konkurrenzorientierung“. Das Modell wies zu beiden Zeitpunkten einen sehr guten Modellfit<sup>8</sup> auf (Klasse 7:  $\chi^2 = 26.7$ ;  $df = 19$ ;  $p = .112$ ; RMSEA = .022; TLI = .985; CFI = .990; Klasse 8:  $\chi^2 = 47.7$ ;  $df = 19$ ;  $p = .000$ ; RMSEA = .043; TLI = .96; CFI = .973), so dass die faktorielle Struktur der Items bestätigt werden konnte. Die Ladungen bewegen sich insgesamt auf einem zufriedenstellend hohen Niveau (vgl. Tabelle 7.13). Für den Faktor „Konkurrenzorientierung“ kann das zweite Item „Bei uns hat man manchmal das Gefühl, dass sich Schüler untereinander keine guten Noten gönnen“ eindeutig mit der höchsten Ladung identifiziert werden, das auch in zufriedenstellender Weise die inhaltliche Seite der zugehörigen Dimension kennzeichnet.

Tabelle 7.13: Faktorielle Struktur der Klimaitems in Klasse 7 und 8 (standardisierte Lösung).

	Klasse 7		Klasse 8				
	Engagement der Lehrer/innen	Konkurrenzorientierung	Engagement der Lehrer/innen	Konkurrenzorientierung			
LE01	.69	KO01	.40	LE01	.66	KO01	.51
LE02	.59	KO02	.68	LE02	.58	KO02	.77
LE03	.58	KO03	.40	LE03	.63	KO03	.50
LE04	.53	KO04	.53	LE04	.61	KO04	.52

<sup>8</sup> Vgl. das nachfolgende Unterkapitel 7.4 zur Interpretation der Güteindizes.

### 7.3.1.3 Reliabilität

Wie Tabelle 7.14 zeigt, ergeben sich für die beiden Skalen zu beiden Messzeitpunkte insgesamt Reliabilitäten in akzeptabler Höhe ( $\alpha > .5$ ).

Tabelle 7.14: Reliabilitäten (Cronbachs  $\alpha$ ) der Skalen zur Erfassung des Klassenklimas

	Klasse 7	Klasse 8
Engagement der Lehrer/innen	.69	.71
Konkurrenzorientierung	.55	.66

## 7.3.2 Persönlichkeitsmerkmale

### 7.3.2.1 Explorative Faktorenanalysen

Tabelle 7.15: Ergebnisse der Hauptkomponentenanalysen zu den Persönlichkeitsmerkmalen

	Klasse 7			Klasse 8		
	Hauptkomponenten			Hauptkomponenten		
	1	2	3	1	2	3
SK01	.74/.74/.82			.82/.81/.88		
SK02	.83/.83/.84			.82/.83/.81		
SK03	.85/.85/.86			.87/.87/.92		
MF01		.83/.82/.87			.85/.85/.83	
MF02		.71/.70/.78			.67/.67/.51	
MF03		.72/.72/.71			.80/.79/.86	
MZ01			.76/.76/.76			.80/.79/.83
MZ02			.83/.83/.86			.79/.78/.82
MZ03			.74/.72/.80			.73/.74/.78
Erklärte Gesamtvarianz (%)	63/62/75			67/67/72		

Anmerkungen: Ladungsmatrizen der Faktorenanalysen mit anschließender Varimax-Rotation mit unzentrierten Daten vor dem ersten, auf Individualebene (within-analysis) nach dem ersten und auf Klassenebene (between-analysis) nach dem zweiten Schrägstrich. Abkürzungen: SK – Selbstkonzept der Begabung, MF – Attribuierung schulischer Misserfolge auf Fähigkeit, MZ – Attribuierung schulischer Misserfolge auf Zufall. Bzgl. der Itemtexte vgl. Tabellen 7.10 und 7.11.

Für die explorativen Faktorenanalysen zu den Persönlichkeitsvariablen wurden wiederum nur die im vorhergehenden Unterkapitel 7.2 dargestellten Selbstkonzept- und Attributionsitems einbezogen. Ziel der Faktorenanalysen war es auch, die Konstrukte sowohl auf Individual- als auch auf Klassenebene eindeutig voneinander abzugrenzen. Zur Festlegung der Hauptkomponentenzahl wurde wiederum das Kaiser-Kriterium festgelegt und für die beiden

Messzeitpunkte wurden getrennt Faktorenanalysen mit orthogonaler Rotation (Varimax) durchgeführt, und zwar zunächst mit unzentrierten Daten auf Individualebene und daraufhin mit zentrierten Daten auf Individualebene (within-analysis) und den Klassenmittelwerten auf Klassenebene (between-analysis). Auf Individualebene können mit den beiden Komponenten gut 60% erklärt werden, auf Klassenebene ergeben sich Lösungen von über 70% (vgl. Tabelle 7.15). Insgesamt lassen sich die Konstrukte eindeutig voneinander trennen.

### 7.3.2.2 Konfirmatorische Faktorenanalysen

Die Persönlichkeitskonstrukte konnten auch in konfirmatorischen Faktorenanalysen entsprechend gut zu beiden Zeitpunkten voneinander abgegrenzt werden (Klasse 7:  $\chi^2 = 53.9$ ;  $df = 24$ ;  $p = .000$ ; RMSEA = .039; TLI = .970; CFI = .980; Klasse 8:  $\chi^2 = 50.5$ ;  $df = 24$ ;  $p = .001$ ; RMSEA = .037; TLI = .980; CFI = .987). Die standardisierten Schätzungen zeigen, dass die Items mittelmäßige bis gute Indikatoren darstellen (vgl. Tabelle 7.16).

Tabelle 7.16: Faktorielle Struktur der Items der Persönlichkeitsvariablen in Klasse 7 und 8 (standardisierte Lösung)

Selbstkonzept	Klasse 7			Klasse 8		
	Attribuierung auf Fähigkeit	Attribuierung auf Zufall		Selbstkonzept	Attribuierung auf Fähigkeit	Attribuierung auf Zufall
SK01	.61	MF01 .66	MZ01 .65	SK01 .68	MF01 .72	MZ01 .73
SK02	.70	MF02 .58	MZ02 .81	SK02 .72	MF02 .62	MZ02 .83
SK03	.80	MF03 .64	MZ03 .53	SK03 .87	MF03 .71	MZ03 .48

### 7.3.2.3 Reliabilität

Es ergeben sich für die drei Skalen zu beiden Messzeitpunkten insgesamt zufriedenstellend hohe Reliabilitäten (vgl. Tabelle 7.17).

Tabelle 7.17: Reliabilitäten (Cronbachs  $\alpha$ ) der Skalen zur Erfassung der Persönlichkeitsmerkmale

	Klasse 7	Klasse 8
Selbstkonzept der Begabung	.74	.80
Attribuierung auf Fähigkeit	.66	.72
Attribuierung auf Zufall	.69	.71

## 7.4 Auswertungsverfahren

Im Folgenden werden die in der vorliegenden Arbeit benutzten Auswertungsverfahren generell angesprochen. Das methodische Vorgehen wird dann genauer im Zusammenhang mit den einzelnen Auswertungsschritten beschrieben.

### 7.4.1 Deskriptive Statistik

Um Unterschiede zwischen Gruppen (z. B. zwischen Schüler/inne/n unterschiedlicher Schulformen) zu untersuchen, werden Varianzanalysen berechnet. Das Signifikanzniveau wird auf 5% festgesetzt. Neben dem Prüfwert und der Irrtumswahrscheinlichkeit wird auch Eta-Quadrat ( $\eta^2$ ) angegeben, das als Maß für die Größe eines Effekts verwendet werden kann.<sup>9</sup> Nach Gediga und Kuhnt (1998) wird dabei von der Einteilung in Effekten von kleiner, mittlerer und hoher Größe ausgegangen (vgl. Tab. 7.17). Um Einzelvergleiche zwischen verschiedenen Gruppen durchzuführen, wird der Scheffé-Test benutzt (vgl. Bortz, 2005). Um querschnittliche Zusammenhänge zwischen Variablen (z. B. Klima und Persönlichkeitsmerkmalen) zu untersuchen, werden bivariate Korrelationen berechnet (Pearson's  $r$ ).<sup>10</sup> Zusammenhänge werden als substantiell angesehen, wenn sie eine Höhe von  $|r| \geq .20$  aufweisen (vgl. Benninghaus, 1996). Die Korrelation  $r$  kann auch als Maß für die Größe eines Effekts verwendet werden. Sedlmeier (zitiert nach Gediga und Kuhnt, 1998) hat hier eine Einteilung in kleine, mittlere und hohe Effektgrößen vorgeschlagen (vgl. Tabelle 7.18). Ferner werden die Mittelwertveränderungen über zwei Zeitpunkte mit einem t-Test für abhängige Stichproben überprüft. Als Maß für die praktische Bedeutsamkeit von signifikanten Ergebnissen aus einem t-Test für abhängige Stichproben wird die Effektstärke  $d$  angegeben. Diese wird nach der Formel

$$d = t \left[ \sqrt{2(1 - r) / n} \right]$$

---

<sup>9</sup> Eta ( $\eta$ ) ist eine Maßzahl, mit der die Beziehung zwischen einer nominalen und einer metrischen Variable beschrieben werden kann. Sie kann Zahlenwerte von 0 bis 1 annehmen. Eta-Quadrat ( $\eta^2$ ) gibt den Anteil der durch die unabhängigen Variable aufgeklärten Varianz gemessen an der Gesamtvarianz in der Stichprobe an (vgl. Benninghaus, 1996). Häufig wird Eta-Quadrat ( $\eta^2$ ) mit 100 multipliziert, um die Angabe des Verhältnisses in Prozent zu erhalten. Die Prozentwerte werden dann als Maß für die praktische Bedeutsamkeit der statistisch signifikanten Unterschiede verwendet. Verschiedene Autoren (vgl. z. B. Bortz, 2005) weisen darauf hin, dass das Konzept der statistischen Signifikanz mit Kriterien der praktischen Bedeutsamkeit verbunden werden sollte, da mit der Größe der Stichprobe auch Ergebnisse eher signifikant werden, während ein Effektstärkemaß davon nicht betroffen ist.

<sup>10</sup> Der Pearsonsche Produkt-Moment-Korrelations-Koeffizient  $r$  beschreibt den Grad und die Richtung einer linearen Beziehung zwischen zwei Variablen mit Intervallskalenniveau und kann Zahlenwerte von  $-1$  bis  $+1$  annehmen. Bei dem Wert  $r = 0$  liegt kein linearer Zusammenhang, bei dem Wert  $|r| = 1$  liegt ein perfekter Zusammenhang vor (Benninghaus, 1996).

berechnet, wobei  $r$  die Korrelation zwischen erstem und zweitem Messzeitpunkt ist. Bortz und Döring (1995, nach Gediga & Kuhnt, 1998) geben für  $d$  eine Einteilung in kleine, mittlere und hohe Effektgrößen an, die ebenfalls in Tabelle 7.18 aufgeführt ist.

Tabelle 7.18: Kleine, mittlere und hohe Effektgröße für unterschiedliche Effektstärkemaße (vgl. Gediga & Kuhnt, 1998)

	klein	mittel	hoch
Korrelation $r$	$\geq .1$	$\geq .3$	$\geq .5$
aufgeklärte Varianz $\eta^2$	$\geq .01$	$\geq .059$	$\geq .139$
Effektstärke $d$	$\geq .2$	$\geq .5$	$\geq .8$

## 7.4.2 Strukturgleichungsmodelle

Mit linearen Strukturgleichungsmodellen werden anhand empirischer Daten zuvor formulierte Hypothesen geprüft, um Merkmalszusammenhänge zu erklären. Ein zentrales Merkmal von Strukturgleichungsmodellen ist zudem die Unterscheidung von latenten und beobachteten Variablen. Latente Variablen sind nicht direkt, sondern nur indirekt über verschiedene Indikatoren, d.h. über die beobachteten Variablen, erfassbar. Lineare Strukturgleichungsmodelle integrieren Aspekte der Faktorenanalyse, der Regressionsanalyse und der Pfadanalyse und ermöglichen somit die simultane Überprüfung von sowohl Hypothesen zwischen theoretischen Konstrukten als auch Hypothesen zwischen theoretischen Konstrukten und beobachteten Variablen (vgl. Pfeifer & Schmidt, 1987; Bortz, 2005).<sup>11</sup> In der vorliegenden Arbeit kommen lineare Strukturgleichungsmodelle vorwiegend in Modellen nach dem „Cross-Lagged-Panel-Design“ zum Einsatz. Als Schätzverfahren wird die

<sup>11</sup> Allgemein unterscheidet man das Strukturmodell vom Messmodell. Während ersteres die Beziehung zwischen den latenten Variablen abbildet, entspricht das Messmodell dem Grundmodell der Faktorenanalyse. Für die Darstellung eines Modells gilt als Konvention, dass beobachtete Variablen durch Quadrate und latente Variablen durch Kreise dargestellt werden. Als notwendige Bedingung für die Identifizierbarkeit der Parameter gilt, dass die Anzahl der zu schätzenden Parameter höchstens so groß wie die Anzahl der Elemente der empirischen Ausgangsmatrix ist. Die Anzahl der bekannten Korrelationen ergibt sich bei  $k$  Indikatorvariablen zu  $k \cdot (k+1) / 2$ . Im Regelfall sollten jedoch deutlich mehr bekannte Korrelationen vorhanden sein als die Anzahl der zu schätzenden Parameter. Computerprogramme wie LISREL oder AMOS bieten unterschiedliche Schätzverfahren an, wobei in der Praxis die Maximum-likelihood-Schätzung (ML) dominiert. Das Programm beginnt dabei mit der Festsetzung von ersten Näherungswerten für die zu schätzenden Parameter. Diese werden daraufhin iterativ so lange verändert, bis die aus den geschätzten Parametern errechneten Korrelationen den empirisch ermittelten Korrelationen möglichst gut entsprechen.

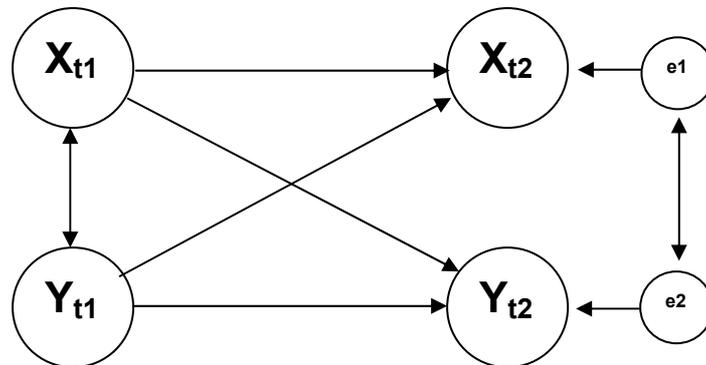
Maximum-likelihood-Schätzung verwendet. Es werden unterschiedliche Parameter genannt, mit denen die Güte des Modells beurteilt werden kann.<sup>12</sup>

Zur Untersuchung kausaler Effekte werden Strukturgleichungsmodelle nach dem „Cross-Lagged-Panel-Design“ berechnet. In diesem werden gegenseitige Wirkbeziehungen als Partialregressionskoeffizienten ermittelt (vgl. Abbildung 7.2). Die Methode der Analyse zeitverschobener Kreuzkorrelationen („Cross-Lagged Panel Analysis“, CLPA) setzt voraus, dass die beiden Variablen X und Y, deren ursächliche Beziehung untersucht werden soll, zu mindestens zwei Zeitpunkten t1 und t2 erhoben worden sind. Dabei fordert das Prinzip der „Synchronizität“, dass beide Variablen jeweils zum gleichen Zeitpunkt erhoben wurden, während das Prinzip der „Stationarität“ ein konstantes Verhältnis der Beziehungen zwischen den jeweils querschnittlich erhobenen Variablen erfordert: Diese sollen sich möglichst nicht in Höhe und Richtung der Korrelationen unterscheiden. Damit man von einem kausalen Einfluss sprechen kann, muss wenigstens einer der Cross-Lagged Pfade signifikant sein. Sind beide signifikant, kann dies als Wechselwirkung interpretiert werden (vgl. Calsyn, 1976; Pottebaum et al., 1986).

---

<sup>12</sup> Mit dem Chi-Quadrat-Wert wird die Nullhypothese überprüft, die davon ausgeht, dass die empirischen Daten und das theoretische Modell gut übereinstimmen. Ein Chi-Quadrat-Wert von Null signalisiert das Ergebnis einer perfekten Übereinstimmung und die Beibehaltung der Nullhypothese (vgl. Backhaus et al., 1994). Da jedoch der Chi-Quadrat-Wert nicht unabhängig von der Stichprobengröße ausfällt und bei großen Stichproben in der Regel signifikant wird, wird als alternative Prüfgröße der Quotient zwischen Chi-Quadrat-Wert und der Zahl der Freiheitsgrade eines Modells vorgeschlagen (vgl. Byrne, 1989). Akzeptable Werte liegen dann zwischen 1 und 5. Unabhängig von der Stichprobengröße und vergleichsweise robust ist der Tucker-Lewis-Index (TLI, vgl. Tucker & Lewis, 1973). TLI-Werte größer als .9 werden als Indikatoren akzeptabler Modellanpassung interpretiert. Der Root Mean Error of Approximation-Wert (RMSEA) ist ein Maß für die pro Freiheitsgrad gefundene Abweichung der empirischen Kovarianz- bzw. Korrelationsmatrix von der modellimpliziten Populationsmatrix. Nach Browne & Cudeck (1993) stellt ein RMSEA von unter .05 einen sehr guten Modellfit dar, während ein RMSEA von .08 noch als zufrieden stellende Anpassung vorgeschlagen wird. Der Comparative Fit Index (CFI) ist wie der TLI ein relativer Fit Index. Auch sein Wert liegt zwischen 0 und 1, wobei Werte ab .9 als gut interpretiert werden. Der Goodness of Fit Index (GFI) vergleicht die Passung des Modells mit dem so genannten Null-Modell. Mit einem Wert ab .95 kann man von einer guten Modellgüte ausgehen. Multipliziert man ihn mit 100, erhält man den prozentualen Anteil der gesamten Ausgangsvarianz, welche die der Berechnung zugrunde liegende Modellstruktur erklärt.

Abbildung 7.2: Cross-lagged-panel-Modell (ohne Messmodell) (Darstellung nach Schnabel, 1996)



### 7.4.3 Kausalbeziehungen und Regressionsanalysen

Wie im nachfolgenden Unterkapitel erläutert werden wird, werden in dieser Arbeit Regressionsanalysen im Rahmen des mehrbenenanalytischen Ansatzes berechnet. In ihnen werden Variablen aufeinander bezogen, von denen angenommen wird, dass sie aufeinander Einfluss nehmen. Um dies adäquat abzubilden, werden die in den Regressionsgleichungen angenommenen Kausalbeziehungen nicht innerhalb eines Messzeitpunktes spezifiziert, sondern die zeitliche Ordnung als wesentliches Merkmal der angenommenen kausalen Definition aufgefasst. Somit wird beispielsweise nicht versucht, das Selbstkonzept der Begabung in Klasse 7 durch die Konkurrenzorientierung in Klasse 7 zu erklären, sondern in diesem Fall müsste die Konkurrenzorientierung dem Selbstkonzept zeitlich vorausgehen, d.h. die Konkurrenzorientierung in Klasse 7 auf das Selbstkonzept in Klasse 8 bezogen werden. Ferner haben die Regressionsanalysen zum Ziel, Veränderungswirkungen abzubilden. Dies geschieht dadurch, dass die abhängige Variable zum vorhergehenden Zeitpunkt als unabhängige Variable in die Regressionsgleichung aufgenommen wird. Bleiben dann die unabhängigen Variablen, von denen ein Effekt auf die abhängige Variable erwartet wird, weiterhin signifikant, wird dies als Wirkung auf die Veränderung der abhängigen Variable interpretiert. Dieses Vorgehen findet man z. B. bei Baumert et al. (2000) und Köller (2004). Zudem wurde es in den Abschnitten 3.2.2 und 4.3.2.1 (vgl. Tabellen 3.2 und 4.5) dargestellt.

#### **7.4.4 Hierarchische Analysen**

In den Kapiteln 3 und 4 dieser Arbeit wurde auf die unterschiedlichen Ebenen eingegangen, auf denen die Qualität von Schule und Klassenklima lokalisiert werden kann: Schulqualität wird mit spezifischen Kriterien auf verschiedenen schulorganisatorischen Ebenen gesehen und bei Analysen zu Zusammenhängen von Klassenklimawahrnehmungen durch Schüler/innen und Persönlichkeitsmerkmalen der Schüler/innen müssen aufgrund der Frage nach dem Subjektbezug unterschiedliche Analyse- und Interpretationsebenen berücksichtigt werden. Aus dieser Perspektive stellen Stichproben von Schüler/inne/n immer geschachtelte Datensätze dar: Schüler/innen besuchen eine Klasse in einer Schule einer Schulform, die wiederum zu einem Schulsystem gehört. Diese hierarchische Schachtelung, die sich mit weiteren höheren hierarchischen Einheiten (Stadtteil, Bundesland etc.) sowie niedrigeren hierarchischen Einheiten (z. B. Schülereliquen) weiter fortsetzen ließe, macht die Notwendigkeit einer adäquaten statistischen Vorgehensweise deutlich. Das statistische Verfahren, das dieser Notwendigkeit Rechnung trägt und in dieser Arbeit zur Anwendung kommt, ist die Mehrebenenanalyse.

##### **7.4.4.1 Verwendung der Mehrebenenanalyse in der vorliegenden Arbeit**

Das mehrebenenanalytische Verfahren, wie es in der vorliegenden Arbeit zum Einsatz kommt, lehnt sich an die Arbeiten von Bryk und Raudenbush (1992) und Raudenbush et al. (2004) an, da die entsprechenden Analysen in dieser Arbeit auch mit dem von diesen Autoren entwickelten Programm HLM durchgeführt werden. Das Programm erlaubt Analysen wahlweise auf zwei (HLM/2L) oder drei (HLM/3L) Ebenen, so dass z. B. die Ebene der Schüler/innen, die Klassenebene und die Schulebene unterschieden werden können, und ermöglicht die simultane Berücksichtigung der verschiedenen Ebenen in einer Analyse. Die Analysestrategie ist damit Auswertungen überlegen, in denen Analysen getrennt für die verschiedenen Ebenen durchgeführt werden, denn die Modellierung unterschiedlicher Auswertungsebenen im HLM-Ansatz erlaubt, die verschiedenen Ebenen aufeinander zu beziehen. In der vorliegenden Arbeit werden hauptsächlich hierarchische lineare Modelle für kontinuierliche y Variablen (Zwei-Ebenen- und Drei-Ebenen-Modelle) berechnet, so dass im Folgenden nur auf diese Modelle eingegangen wird. Die Modellierung individueller Entwicklungsverläufe, die in jüngerer Zeit ein weiteres Anwendungsgebiet des HLM-Ansatzes darstellt (z. B. bei Köller, 1998; Gruehn, 2000), kommt in der vorliegenden Arbeit nicht zum Einsatz, da sie in der Regel mit einem längsschnittlichen Datensatz durchgeführt

wird, der mindestens drei Messzeitpunkte beinhaltet.<sup>13</sup> Die folgende kurze Erläuterung des Mehrebenenansatzes geschieht in Hinblick auf die in dieser Arbeit vorgenommenen Auswertungen und lehnt sich u.a. an die Darstellung hierarchisch linearer Modelle für kontinuierliche y Variablen von Engel (1998) an. Für eine ausführliche Darstellung sei auf Bryk und Raudenbush (1992), Engel (1998) und Hox (2002) verwiesen. Die Idee der einfachen Regression ist, eine abhängige Variable Y (Regressand) durch eine unabhängige Variable X (Regressor) linear zu erklären (z. B. Backhaus et al., 1994; Bortz, 2005). Grundlage der Regressionsanalyse bildet eine Gleichung, die z. B. das Attribuierungsverhalten  $y_i$  des Schülers  $i$  ( $i=1, \dots, n$ ) als Funktion des wahrgenommenen Verhaltens der Lehrer/innen und eines Zufallsterms  $e_i$  beschreiben kann:

$$y_i = \beta_0 + \beta_1 x_{1i} + e_i \quad (\text{Gleichung 7.1})$$

In dieser Gleichung werden dann die Parameter  $\beta_0$  als Regressionskonstante und  $\beta_1$  als Effekt des wahrgenommenen Verhaltens der Lehrer/innen geschätzt, wobei das Modell davon ausgeht, dass beide Effekte für alle  $i$  Personen der Stichprobe gleich sind. Das Besondere am HLM-Ansatz ist nun die Annahme, dass die Parameter  $\beta_0$  und  $\beta_1$  grundsätzlich variieren können – also nicht für alle  $i$  Personen gleich sind. Beispielsweise könnte sich sowohl das mittlere Attribuierungsverhalten als auch der Effekt des wahrgenommenen Verhaltens der Lehrer/innen auf das Attribuierungsverhalten von Klasse zu Klasse oder von Schule zu Schule unterscheiden. Während in der einfachen Regression solche Unterschiede konfundiert bleiben, können sie durch das mehrebenenanalytische Vorgehen aufgedeckt werden. In Zwei-Ebenen-Modellen wird  $\beta_0$  durch  $\beta_{0j}$  und  $\beta_1$  durch  $\beta_{1j}$  ersetzt, wobei  $j$  für die Einheiten der höheren Ebene – z. B. für  $j$  Klassen – steht:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + e_{ij} \quad (\text{Gleichung 7.2})$$

---

<sup>13</sup> In AIDA wurden Schüler/innen nicht nur in Klasse 7 und 8, sondern auch in Klasse 9 befragt. Über alle drei Messzeitpunkte lässt sich eine Panelstichprobe von 1500 Schüler/innen legen, so dass die Modellierung individueller Entwicklungsverläufe mit dem AIDA-Datensatz prinzipiell möglich ist. Für die vorliegende Arbeit ist es jedoch von zentraler Bedeutung, die Zugehörigkeit der Schüler/innen zu Klassen, Schulen und Schulformen zu berücksichtigen. Aufgrund dessen müsste diese Panelstichprobe (wie in Abschnitt 7.1.2 beschrieben) reduziert werden mit dem Ergebnis, dass quasi nur noch Gymnasiast/innen der Stichprobe angehören, da der Anteil der Schüler/innen anderer Schulformen dann so verschwindend gering ist, dass Schulformeffekte nicht mehr untersucht werden können.

Für die zweite Ebene werden weitere Regressionsgleichungen formuliert, da die Regressoren der unteren Ebene zu Regressanden der höheren Ebene werden:

$$\beta_{0j} = \gamma_0 + u_{0j} \quad (\text{Gleichung 7.3})$$

$$\beta_{1j} = \gamma_1 + u_{1j} \quad (\text{Gleichung 7.4})$$

Die beiden Gleichungen 7.3 und 7.4 werden in Gleichung 7.2 eingesetzt. Daraus resultiert eine Schätzgleichung mit mehreren Zufallskomponenten ( $e_{ij}$ ,  $u_{0j}$  und  $u_{1j}$ ). In HLM werden die Regressionskoeffizienten und Varianzen der Regressionsschätzungen auf allen Ebenen simultan geschätzt, wobei empirische Bayes-Schätzer und Maximum-Likelihood-Schätzungen angewendet werden (vgl. Bryk & Raudenbush, 1992). Variiert nun die höhere Ebene nicht bedeutsam um die Parameter der unteren Ebene, werden die Zufallskomponenten  $u_{0j}$  und  $u_{1j}$  der höheren Ebene Null oder nahe Null sein. Mit einem Chi-Quadrat-Test erhält man Auskunft darüber, ob eine Zufallsvariation existiert. Fällt der Test nicht signifikant aus, wird z. B. empfohlen, die entsprechende Zufallskomponente aus der Gleichung herauszunehmen, um die Stabilität der Schätzung zu erhöhen (vgl. Raudenbush et al., 2004). Variiert hingegen die höhere Ebene um die Parameter der unteren Ebene, geben die Zufallskomponenten zunächst an, welche Varianz auf der höheren Ebene erklärbar ist (dies wird mit dem sogenannten „Null-Modell“ vorgenommen). Beispielsweise wäre es möglich, dass das mittlere Attribuierungsverhalten der Schüler/innen einer Klasse von Klasse zu Klasse variiert. In diesem Fall könnte man dann durch das Einsetzen von Prädiktoren auf Klassenebene versuchen, den entsprechenden Fehlerterm zu reduzieren – z. B. durch das Einsetzen des Klassenmittelwerts des wahrgenommenen Verhaltens der Lehrer/innen als Prädiktor auf Klassenebene.

Mit der Modellierung von Prädiktoren auf Individual- und Klassenebene stellt sich die Frage nach ihrem Verhältnis zueinander, insbesondere dann, wenn es sich um Variablen handelt, die auf der Individualebene erhoben wurden und durch Aggregation (z. B. durch Bildung des Klassenmittelwerts) einer höheren Ebene zugewiesen werden. Dieses Vorgehen wird in der Forschung zum Klassenklima gewählt, um individuelles und kollektives Klima abzubilden (vgl. z. B. Eder, 1996). Um in einer Analyse eine eindeutige Trennung der Individual- und Klasseneffekte zu erzielen, wird auf Individualebene der Prädiktor am Klassenmittelwert zentriert, während auf Klassenebene der Klassenmittelwert als Prädiktor eingesetzt wird. Damit kann einerseits die Varianz zwischen den Schüler/inne/n innerhalb der einzelnen

Klassen und andererseits die Varianz zwischen den Klassen erklärt werden (vgl. Ditton, 1998).

Ferner können mit hierarchischen Modellen Interaktionseffekte zwischen den verschiedenen Ebenen modelliert werden (vgl. Hox, 2002). Beispielsweise könnte das wahrgenommene individuelle Verhalten der Lehrer/innen in Gymnasialklassen das Attribuierungsverhalten der Schüler/innen stärker beeinflussen als in Klassen der anderen Schulformen. Vorausgesetzt, die Zufallskomponente  $u_{1j}$  ist groß genug, um durch den Prädiktor „Gymnasium“  $W_1$  auf Klassenebene reduziert werden zu können, kann dieser der Gleichung 7.4 hinzugefügt werden:

$$\beta_{1j} = \gamma_{10} + \gamma_{10}W_1 + u_{1j} \quad (\text{Gleichung 7.5})$$

Bevor allerdings Interaktionseffekte dieser Art modelliert werden, wird empfohlen, die entsprechenden Prädiktoren zunächst in den Zufallsterm der Regressionskonstanten der unteren Ebene (Gleichung 7.3) einzusetzen (vgl. Raudenbush et al., 2004).

Die Ausdehnung des Zwei-Ebenen-Modells auf ein Drei-Ebenen-Modell ist prinzipiell nur dadurch gekennzeichnet, dass nun auch die Regressoren der zweiten Ebene zu Regressanden der dritten Ebene werden. Ob das gewählte Modell zwei oder drei Ebenen beinhaltet, hängt einerseits von der Fragestellung und andererseits von der Varianz ab, die auf den jeweiligen Ebenen überhaupt erklärbar ist. In der Regel wird dies zunächst mit einem „Null-Modell“ getan, das noch keine möglichen Prädiktoren, sondern nur die abhängige Variable beinhaltet, und mit dem die Varianz der abhängigen Variable in die Anteile der unterschiedlichen Ebenen zerlegt wird. Aus diesen Varianzanteilen lässt sich z. B. dann der „Intraclass Correlation Coefficient“ (ICC, vgl. Bryk & Raudenbush, 1992, S. 18) berechnen. Zeigt sich ein nur geringer Varianzanteil auf der dritten Ebene, z. B. auf der Schulebene, wird empfohlen, statt eines Drei-Ebenen-Modells ein Zwei-Ebenen-Modell zu spezifizieren (Bryk & Raudenbush, 1992).

Zur Darstellung der deskriptiven Befunde gehört die hierarchische Varianzzerlegung der verwendeten Skalen. An dieser Stelle soll bereits darauf hingewiesen werden, dass es sich dabei um die Angabe von Eta-Quadrat ( $\eta^2$ ) handelt. Dieses gibt das Verhältnis erklärter Varianz in der Stichprobe wieder. Dagegen stellt der „Intraclass Correlation Coefficient“ (ICC) eine Schätzung des Verhältnisses erklärter Varianz in der Population dar. Dies kann dazu führen, dass die beiden Koeffizienten unterschiedliche Maße erklärter Varianz anzeigen (vgl. Hox, 2002).

#### 7.4.4.2 Umgang mit hierarchischen Daten in Strukturgleichungsmodellen

Bei der Berechnung eines Zusammenhangs zweier Variablen in einer Stichprobe, die hierarchisch geschachtelt ist, sind zunächst die Zusammenhänge verschiedener Ebenen konfundiert. Beispielsweise kann die einfache Korrelation zwischen Selbstkonzept und Mathematiknote bei Schüler/inne/n keine Auskunft darüber geben, ob sie den auf Individualebene bestehenden Zusammenhang zwischen der Fachleistung der Schüler/innen und ihren Selbsteinschätzungen oder den auf Klassenebene vorzufindenden Zusammenhang zwischen dem Bewertungsmaßstab der Lehrer/innen und dem mittleren Klassenselbstkonzept widerspiegelt. Für die vorliegende Arbeit ist es daher wichtig, bei der Berechnung von Strukturgleichungsmodellen zumindest die Individualebene von den schulorganisatorisch höheren Ebenen (Klasse, Schule und Schulform) analytisch zu trennen.

Ditton (1998, S. 86) schlägt zur Ermittlung des korrekten Individualkoeffizienten zweier Variablen die „pooled-within-organisations Beziehung“ ( $\beta_w$ ) vor: „Um diese zu ermitteln, werden von den individuellen Messwerten die jeweiligen Gruppenmittelwerte subtrahiert, womit die Beziehung beider Variablen um Effekte der Gruppenzugehörigkeit bereinigt wird“. Durch dieses Verfahren entsteht eine Varianz-Kovarianz-Matrix von bereinigten Individualwerten, die auch als „pooled-within-Matrix“ ( $S_{PW}$ ) bezeichnet wird und die Zusammenhänge auf Ebene der Schüler/innen repräsentiert (vgl. Kaplan, 2000). Die Zentrierung um die Gruppenmittelwerte gewährleistet eine korrekte Schätzung von  $\beta_w$  (Ditton, 1998). Entsprechend kann eine „between-Matrix“ für die Klassenebene aus den Klassenmittelwerten erzeugt werden ( $S_B$ ). Diese Matrix repräsentiert dann die Zusammenhänge auf Klassenebene. Da in der vorliegenden Arbeit die Klassenstichprobe eher klein ist ( $N_{\text{Klassen}} = 64$ ), resultieren für die Schätzungen der Pfadgewichte auf Klassenebene möglicherweise große Standardfehler. Parameterschätzungen und Standardfehler auf Klassenebene sollten daher mit Vorsicht interpretiert werden.

#### 7.4.5 Verwendete Computerprogramme

Die deskriptiven statistischen Auswertungen werden mit SPSS 10 vorgenommen. Strukturgleichungsmodelle werden mit AMOS 5.0 berechnet. Hierarchische Analysen werden mit dem Programmpaket HLM 6.0 (HLM/2L und HLM/3L, Bryk & Raudenbush, 1992; Raudenbush et al., 2004) durchgeführt. Dieses steht für die vorliegende Arbeit jedoch nur als Studentenversion zur Verfügung und lässt daher lediglich die Berechnung von Modellen zu, in denen maximal vier Prädiktorvariablen pro Fehlerterm einbezogen werden können.