

5. Methode

5.1 Anlage und Durchführung der Untersuchung

Zur Beantwortung der Fragestellungen wurde auf die Stichprobe des Modellversuchs *Verbund Selbstwirksamer Schulen* zurückgegriffen, die im Rahmen der wissenschaftlichen Begleitung erhoben wurde. An dem Modellversuch nahmen zehn Schulen aus dem gesamten Bundesgebiet teil. Das Ziel des Modellversuchs bestand in der Förderung schulinterner Innovationsprojekte. Die Modellschulen bauten zudem ein Netzwerk von interessierten Nachbarschulen auf, um ihre Erfahrungen weiterzugeben. Die erste Förderungsphase umfaßte drei Jahre und sah eine wissenschaftliche Begleitung des Modellversuchs mit drei Meßzeitpunkten im Abstand von jeweils einem Jahr vor. Im folgenden wird der Modellversuch in bezug auf die Fragestellung dieser Abhandlung beschrieben. Eine Gesamtdarstellung des Modellversuchs und der teilnehmenden Schulen findet sich bei Brockmeyer und Edelstein (1997), Edelstein (1995) sowie bei Jerusalem und Schwarzer (1999b). Der Modellversuch wurde gefördert durch die Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung, die Johann Jacobs Stiftung sowie die Freudenberg Stiftung.

Die wissenschaftliche Begleitung des Modellversuchs *Verbund Selbstwirksamer Schulen* unter Leitung von Prof. Dr. Wolfgang Edelstein, Prof. Dr. Diether Hopf, Prof. Dr. Matthias Jerusalem und Prof. Dr. Ralf Schwarzer verfolgte drei Hauptanliegen. Das erste bestand darin, Daten zur Selbstwirksamkeitsentwicklung und zum Klassenklima in der Sekundarstufe I zu erheben sowie neue Meßinstrumente zu entwickeln. Parallel zur Schülerbefragung wurden u.a. die Selbstwirksamkeitserwartungen bei Lehrern erhoben. Ein weiteres Hauptanliegen der wissenschaftlichen Begleitung bestand darüber hinaus in der Evaluation von Maßnahmen, die im Rahmen des Modellversuchs von bestimmten Lehrer- und Schülergruppen umgesetzt wurden. Die externe Evaluation ergänzte dabei die interne Evaluation, die von den Schulen selbst durchgeführt wurde. Die Beratung der Schulen bei psychologisch-pädagogischen Fragestellungen war ein weiterer Aufgabenbereich der wissenschaftlichen Begleitung.

Die von den Schulen initiierten Innovationsprojekte zielten im wesentlichen auf eine Förderung der Selbständigkeit der Schüler (z.B. Mitarbeit in einem Schulrestaurant), auf eine Weiterbildung im Umgang mit neuen Medien (z.B. EDV-Bibliothekssystem) und auf eine Verbesserung der Unterrichtsqualität

(z.B. Tutoren-Systeme, Bildung von Lehrer-Teams) ab. Die Maßnahmen wurden von den Schulen unter eigener Regie durchgeführt. Es gab *keine* standardisierten Treatments. Zumeist wurden bereits bestehende Projekte intensiviert und mit den Mitteln des Modellversuchs unterstützt. Die Innovationsbereitschaft der Schulen und bereits initiierte Innovationsprojekte waren ausschlaggebend bei der Bewerbung um die Teilnahme am Modellversuch gewesen.

Es sei an dieser Stelle ausdrücklich darauf hingewiesen, daß die Innovationsprojekte der Schulen nicht den Gegenstand der vorliegenden Fragestellung bilden. Eine solche Evaluation einzelner Projekte kann nur in enger Zusammenarbeit mit den Schulen selbst erfolgen. Da die Projekte z.T. nur mit sehr kleinen Schülergruppen und *nicht* im Sinne standardisierter Treatments durchgeführt wurden, könnte eine wissenschaftliche Abhandlung wie die vorliegende diesen Projekten nicht gerecht werden. Festzuhalten bleibt, daß mit den Schülern der Modellschulen keine Trainings oder andere Maßnahmen durchgeführt wurden, die direkt auf eine Veränderung von Kontroll- oder Selbstwirksamkeitsüberzeugungen abzielten, auch wenn der Name des Modellversuchs dies nahelegen könnte.

Aus der Tatsache, daß sich die Schulen um eine Teilnahme am Modellversuch selbst beworben haben und die Innovationsbereitschaft der Schulen ausschlaggebend bei der Bewerbung war, folgt, daß die Schulen als *nicht* repräsentativ für die Schulen des Bundesgebietes anzusehen sind. Es handelte sich vielmehr um sehr engagierte, innovationsfreudige Schulen, die bereit waren, ihre Projekte und Bemühungen zur Verbesserung von Schule öffentlich zu machen und sich einer externen Evaluation zu „stellen“.

5.1.1 Beschreibung der Modellschulen

Die Modellschulen wurden über eine bundesweite Ausschreibung geworben. Aus den 35 Bewerberschulen wurden zehn Schulen vom Lenkungsausschuß des Modellversuchs ausgewählt. Dem Lenkungsausschuß gehörten sowohl Wissenschaftler als auch Mitarbeiter der zuständigen Landes- und Bundesbehörden an. Die Auswahl der Schulen erfolgte unter mehreren Gesichtspunkten. Zum einen sollten sowohl Schulen in den neuen als auch in den alten Bundesländern Berücksichtigung finden. Zum anderen sollten möglichst viele unterschiedliche Schulformen vertreten sein, wobei auf einen Ausgleich zwischen alten und neuen Bundesländern sowie zwischen ländlichen und städtischen Einzugsgebieten geachtet wurde. Die Schulen erklärten sich unter Zustimmung der Kollegien bereit, sowohl an der Schüler- als auch an der Lehrerbefragung durch die wissenschaftliche Begleitung teilzunehmen.

Es wurde ferner erwartet, daß die Schulen zur Umsetzung innovativer Maßnahmen bereit waren, daß sie die Erfahrungen mit der Umsetzung protokollierten sowie ein Netzwerk von Schulen aufbauten, um die Erfahrungen an andere Schulen weiterzugeben. Die Entscheidung über die Art und Weise der Innovationsprojekte lag dabei allein in den Händen der Modellschulen selbst.

Tabelle 5.1 gibt eine kurze Charakterisierung der Schulen nach Bundesländern wieder. Es wurde jeweils nur eine Schule pro Bundesland ausgewählt. Von den zehn Pilotschulen kamen vier aus den neuen Bundesländern. Die Hälfte der Schulen war durch ein überwiegend städtisches Einzugsgebiet gekennzeichnet. Bei der Schule mit der größten Schüleranzahl ($N = 1230$) handelte es sich um eine Gesamtschule aus Hessen. Die Schule mit der geringsten Schülerstärke ($N = 340$) war in Berlin gelegen. Im Durchschnitt gab es zwischen den städtischen und ländlichen Schulen keine wesentlichen Unterschiede in der Schülerstärke.

Tabelle 5.1
Die Modellschulen nach Bundesländern

Bundesland	Schulform	Lehrer ^{a)} unter 40 Jahren (96/97)	Lehrer ^{a)} über 40 Jahren (96/97)	Schüler- stärke (96/97)	Einzugsgebiet (längster Schulweg in Minuten)
Berlin	Hauptschule	13	34	340	städtisch (60)
Brandenburg	Gesamtschule	23	14	440	ländlich (40)
Hamburg	Gewerbeschule	15	60	830	städtisch (90)
Hessen	Kooperative Ge- samtschule mit Oberstufe	18	63	1230	ländlich (35)
Mecklenburg- Vorpommern	Haupt- u. Real- schule mit Grund- schule	13	14	370	ländlich (40)
Niedersachsen	Gymnasium	4	63	740	städtisch (90)
Rheinland-Pfalz	Realschule	6	25	560	ländlich (45)
Sachsen	Haupt- u. Real- schule	23	10	520	städtisch (40)
Sachsen-Anhalt	Haupt- u. Real- schule	7	14	370	ländlich (15)
Schleswig- Holstein	Integrierte Ge- samtschule	21	25	500	städtisch (90)

Anmerkung.

^{a)} vollbeschäftigte Lehrer

Die Tabelle 5.2 stellt die Schwerpunktprojekte der Schulen dar. Diese kurze Beschreibung der vielfältigen Innovationsprojekte wird den Schulen dabei in keiner Weise gerecht. Es sei daher nochmals auf die Dokumentation des Modellversuchs bei Brockmeyer und Edelstein (1997) und auf den Abschlußbericht (Jerusalem & Schwarzer, 1999b) verwiesen.

Tabelle 5.2
Schwerpunkt-Projekte der Modellschulen

Bundesland	Schulform	Schwerpunkt der Innovationsprojekte
Berlin	Hauptschule	Herausbildung einer demokratischen Leitungsstruktur, Bildung eines erweiterten Schulleitungsgremiums, Einbindung eines möglichst großen Teils des Kollegiums in einzelne Bereiche und Projekte
Brandenburg	Gesamtschule	Bildung von Lehrergruppen, Öffnung des Unterrichts für externe Projekte
Hamburg	Gewerbeschule	Projekte unter dem Motto: „Betroffene zu Beteiligten machen!“, selbstverwaltetes Schulrestaurant
Hessen	Kooperative Gesamtschule mit Oberstufe	Entwicklung einer Corporate Identity, Weiterbildung der Schüler im kooperativen Konfliktlösen, Schüler werden als Mentoren bei der Betreuung jüngere Schüler eingesetzt
Mecklenburg-Vorpommern	Haupt- u. Realschule mit Grundschule	Optimierung der Organisationsstruktur der Schule, sozial-pädagogische Förderstunden, „Team Teaching“, Schüler bieten Kurse für Mitschüler an
Niedersachsen	Gymnasium	Teammodell: Klassenlehrer eines Jahrganges kooperieren eng
Rheinland-Pfalz	Realschule	Anpassung an veränderte Lebensbedingungen der Jugendlichen: u.a. Schulbibliothek als ein vernetztes Selbstlernzentrum
Sachsen	Haupt- u. Realschule	Sozial-pädagogische Förderangebote für die Schüler auch in der Freizeit, Öffnung der Schule als Freizeitstätte, Kooperation mit Sozialarbeitern
Sachsen-Anhalt	Haupt- u. Realschule	Schülerzentrierter Unterricht nach dem Dalton- Plan
Schleswig-Holstein	Integrierte Gesamtschule	Orientierung zur Demokratisierung im ‚Dream Team‘ aus u.a. Schulleitung, Jahrgangssprechern, Sozialpädagogen und Verbindungslehrern

Aus der kurzen Beschreibung der vielfältigen Maßnahmen wird ersichtlich, daß der Schwerpunkt vieler Bemühungen auf eine bessere Organisation der Lehrarbeit abzielte. So wurden an vielen Schulen erweiterte Lehrer-Teams gebil-

dete, die Verantwortung für bestimmte Jahrgänge, Schülergruppen oder Projekte übernahmen. Das Bestreben vieler Schulen bestand darin, Schüler aktiv in den Unterricht einzubeziehen und lebensnahe Projekte mit Schülern durchzuführen, wie etwa das selbstverwaltete Schulrestaurant der Gewerbeschule in Hamburg. Die Schulen bemühten sich, den Schülern demokratische Umgangsformen näherzubringen, Gewalt und Rassismus in der Schule mit den Schülern zu thematisieren. In diesem Kontext forcierte die Gesamtschule aus Hessen das Unterrichtsfach ‚Soziales Lernen‘: Schüler wurden in der Gesprächsführung geschult und zu Konfliktlösern ausgebildet. Einige Schulen führten Freiarbeit ein oder bauten diese aus. So verfolgte die Schule aus Sachsen-Anhalt den Unterricht nach dem Dalton-Plan. Dieser umfaßte Pensen und Freiarbeitsstunden, die von den Schülern zu vielfältigen Aktivitäten mit den Lehrern ihrer Wahl genutzt werden konnten.

Solche und ähnliche Projekte gibt es mittlerweile an vielen bundesdeutschen Schulen. Doch nur selten waren Schulen bisher bereit, sich über einen längeren Zeitraum hinweg bei ihren Bemühungen in die ‚Karten sehen zu lassen‘. Dafür gilt unser Dank den Schülern und Lehrern des Modellversuchs.

Tabelle 5.3
Anlage der Mehrkohortenlängsschnittstudie

<i>Kohorte</i>	<i>Jahrgangsstufen</i>		
	MZP1 1995 / 1996	MZP2 1996 / 1997	MZP3 1997 / 1998
8			7 (707)
7		7 (771)	8 (716)
6	7 (685)	8 (757)	9 (763)
5	8 (737)	9 (739)	10 (687)
4	9 (657)	10 (618)	11 (128)
3	10 (500)	11 (171)	12 (80)
2	11 (165)	12 (99)	13 (91)
1	12 (79)		

Anmerkung. Stichprobengröße in Klammern.

5.1.2 Anlage der Feldstudie

Entsprechend der Fragestellung der wissenschaftlichen Begleitung wurde die Feldstudie als Längsschnittuntersuchung mit drei Meßzeitpunkten im Abstand

von jeweils einem Jahr konzipiert. Die Erhebungswellen wurden in den Monaten von Dezember bis Februar der Jahre 1995 bis 1997 durchgeführt (Tabelle 5.3). Zu jeder Welle nahmen Schüler ab der siebten Jahrgangsstufe an der Befragung teil. Insgesamt konnten so acht Jahrgangsstufen-Kohorten berücksichtigt werden, wobei *nur* die Kohorten 2 bis 6 an allen drei Erhebungswellen beteiligt waren. Für die Kohorten 5 und 6 liegt ein vollständiger Längsschnitt innerhalb der Sekundarstufe I vor.

5.1.3 Schülerstichprobe

Die Gesamtstichprobe setzt sich aus den Schülern der zehn Pilotschulen ab der siebten Jahrgangsstufe zusammen. Zum ersten Meßzeitpunkt nahmen 3089 Schüler der Geburtsjahrgänge 1972 bis 1983 an der Untersuchung teil. Zum zweiten Meßzeitpunkt waren es 3458 Schüler. Zum dritten Meßzeitpunkt erhöhte sich die Anzahl der Untersuchungsteilnehmer nach Berücksichtigung der 13. Jahrgangsstufe noch einmal auf 3514 Schüler. Während zum ersten Meßzeitpunkt 84.5% der Schüler die Sekundarstufe I besuchten, verringerte sich ihr Anteil dadurch auf 83.4% zum zweiten bzw. 81.8% zum dritten Meßzeitpunkt (Tabelle 5.4).

Tabelle 5.4
Gesamtstichprobe der Schüler

	MZP1	MZP2	MZP3
<i>N</i>	3089	3458	3514
männlich	1595 (51.6%)	1741 (50.3%)	1765 (50.2%)
weiblich	1482 (48.0%)	1693 (49.0%)	1699 (48.3%)
ohne Angabe	12 (0.4%)	24 (0.7%)	50 (1.4%)
<i>Jahrgangsstufe</i>			
Sek. I	2609 (84.5%)	2885 (83.4%)	2873 (81.8%)
Sek. II	244 (7.8%)	270 (7.8%)	299 (8.5%)
ohne Angabe	236 (7.6%)	303 (8.8%)	342 (9.7%)
<i>Alter</i>			
Geburtsjahrgänge	1972-1983	1974-1984	1975-1986
stärkster Geburtsjahrgang	1980	1981	1982

Insgesamt liegen für 1091 Schüler der Kohorten 2 bis 6 Daten zu allen Meßzeitpunkten vor. Hierunter sind keine Schüler der Gewerbeschule in Hamburg, da die Schüler an dieser Schule i.R. nur für ein Jahr verbleiben. Zudem ist die Klassenstruktur der Gewerbeschule nicht mit den übrigen Schulen vergleichbar, so daß die Gewerbeschule in Hamburg auch aus diesem Grund für diese Abhandlung nicht berücksichtigt werden konnte.

Für die *Kohorten 5 und 6* liegt ein vollständiger Längsschnitt mit drei Meßzeitpunkten innerhalb der *Sekundarstufe I* vor. Diese Kohorten sind für die Fragestellung nach *Klassenklima-* und *Selbstwirksamkeitsveränderungen* von besonderem Interesse, da sich in diesen Jahrgangsstufen die *Klassenstruktur* nicht veränderte. Sie bilden daher die *Untersuchungsstichprobe*. Auf die Kohorten 2 und 3 kann hingegen nicht zurückgegriffen werden, da mit Eintritt in die Oberstufe das Kurssystem eingeführt wurde, was eine teilweise Auflösung der Klassenstruktur impliziert. Die Stichprobe reduzierte sich durch die Auswahl der *Kohorten 5 und 6* als Untersuchungsstichprobe auf 921 Schüler (Tabelle 5.5). Abweichungen von dieser Stichprobengröße ergaben sich für die einzelnen statistischen Analysen nach Ausschluß von Schülern mit fehlenden Daten für die jeweiligen Analysen.

Tabelle 5.5
Untersuchungsstichprobe

<i>N</i>	921
männlich	464 (50.4%)
weiblich	457 (49.6%)
ohne Angabe	0 (0%)
 <i>Alter</i>	
Geburtsjahrgänge	1979-1983
stärkster Geburtsjahrgang	1982
 <i>Kohortenzugehörigkeit</i>	
Kohorte 5	476 (51.7%)
Kohorte 6	445 (48.3%)
 <i>Besuchter Schulzweig</i>	
Hauptschule	60 (6.6%)
Realschule	367 (40.3%)
Gymnasium	267 (29.3%)
Gesamtschule	216 (23.7%)

Mädchen und Jungen sind in der Untersuchungsstichprobe in nahezu gleichen Anteilen vertreten. Die meisten Schüler wurden im Jahr 1982 geboren. Da nur das Geburtsjahr erhoben wurde, sind keine genaueren Angaben zum Alter der Stichprobe möglich. Der überwiegende Anteil der Schüler (40.3%) besuchte den Realschulzweig. Nur ein geringer Prozentsatz (6.6%) besuchte eine Hauptschule.

Zuordnung der Schüler zu Klassen

Für die Mehrebenenanalysen ist die Zuordnung der Schüler zu Klassen eine notwendige Voraussetzung. Eine Zuordnung zu einer bestimmten Klasse *mißlang*, wenn Schüler ungenaue oder fehlerhafte Angaben zur Klassenzugehörigkeit machten. Gerade in Hauptschulklassen führten fehlerhafte Angaben dazu, daß nur noch sehr wenige Schüler einer Klassen zugeordnet werden konnten, zumal die Schülerstärke der Hauptschulklassen deutlich kleiner war als in den übrigen Klassen. Wenn einer Klasse weniger als fünf Schüler zugeordnet werden konnten, wurde die gesamte Klasse für die *jeweilige Analyse* ausgeschlossen. Dieses Kriterium wurde so niedrig gewählt, um Hauptschulklassen mit zumeist geringer Klassengröße nicht zu benachteiligen und um die Vergleichbarkeit zu anderen Studien herzustellen (vgl. Gruehn, 1989).

Tabelle 5.6
Schüleranzahl pro Klasse

	MZP1	MZP2	MZP3
$N_{\text{Schüler}}$	883	888	878
N_{Klassen}	61	61	61
<i>Schüleranzahl pro Klasse</i>			
minimale Schülerstärke	5	5	5
maximale Schülerstärke	25	27	27
Mittelwert	14.5	14.6	14.4
Modalwert	15	15	15
Median	15	15	15

Anmerkungen. Schüleranzahl von Klassen mit einer Schülerstärke von mindestens fünf Schülern zu jedem Meßzeitpunkt.

Die Schüler der Untersuchungsstichprobe konnten 61 Klassen mit *durchgehend* mehr als fünf Schülern pro Klasse zugeordnet werden (Tabelle 5.6). Einige Mehrebenenanalysen bezogen sich nur auf einen bestimmten Meßzeitpunkt. Für diese Analysen war die Klassenstichprobe größer, da auch Klassen in Frage kamen, die nicht zu allen Meßzeitpunkten mindestens fünf Schüler umfaßten. Zum ersten Meßzeitpunkt waren dies 62 Klassen, zum zweiten 63 und zum dritten wiederum 62 Klassen.

Da die Zuordnung der Schüler zu Klassen mit *durchgehend* mindestens fünf Schülern pro Klasse für einige Mehrebenenanalysen Voraussetzung war, reduzierte sich für diese Analysen die Untersuchungsstichprobe auf 883 bis 888 Schüler (Tabelle 5.6). Die durchschnittliche Klassengröße der 61 Klassen mit mindestens fünf Schülern zu allen Meßzeitpunkten variierte leicht zwischen $M = 14.4$ und $M = 14.6$ Schülern. Der Modalwert betrug zu allen Meßzeitpunkten $Mo = 15$. Eine weitere Reduktion der Stichprobe ergab sich auch für diese Analysen durch Ausschluß von Schülern mit für die jeweiligen Analysen fehlenden Daten. In diesem Zusammenhang ist von Interesse, daß der Klassenlehrer während des Untersuchungszeitraumes nur bei fünf Klassen wechselte.

Drop-Out-Analyse

Um der Frage nachzugehen, inwieweit sich die Schüler der Untersuchungsstichprobe von den Schülern der Kohorte 5 und 6 unterschieden, die nicht berücksichtigt werden konnten, da für sie *nur* zum ersten Meßzeitpunkt Daten vorlagen, wurden Vergleiche auf zentralen Variablen herangezogen (Tabelle 5.7).

In der Untersuchungsstichprobe war das Geschlechtsverhältnis nahezu ausgeglichen, während sich in der Drop-Out-Stichprobe mehr Jungen als Mädchen befanden. Als *grober* Indikator des sozio-ökonomischen Status wurde die Frage nach einem eigenen Zimmer im Elternhaus eingesetzt. Die meisten Schüler (84.7%) der Untersuchungsstichprobe berichteten, daß ihnen ein eigenes Zimmer zur Verfügung stand. Im Gegensatz dazu verfügten nur 68.5% der Schüler der Drop-Out-Stichprobe über ein eigenes Zimmer. Hauptschüler waren in der Drop-Out-Stichprobe stärker vertreten, Gymnasiasten hingegen unterrepräsentiert. Darüber hinaus befanden sich in der Untersuchungsstichprobe mehr Schüler mit guten oder sehr guten Schulleistungen.

Diese Unterschiede deuten darauf hin, daß insbesondere schwächere Schüler, Hauptschüler und Schüler mit vermutlich geringerem sozio-ökonomischen Status aus der Untersuchungsstichprobe herausfielen. Hierfür erscheinen zwei Hauptursachen plausibel.

Die höhere Anteil von Hauptschülern in der Drop-Out-Stichprobe erklärt sich aus der Tatsache, daß diese Schüler die Schule i.R. nach dem neunten Schuljahr verlassen. Hauptschüler der Kohorte 5, die zum ersten Meßzeitpunkt die achte Jahrgangsstufe besuchten, hatten demnach die Schule zum dritten Meßzeitpunkt z.T. bereits verlassen. Da Hauptschüler zumeist aus Familien mit einem schwächeren sozio-ökonomischen Status kommen, erklärt sich hieraus auch dieser Unterschied.

Tabelle 5.7
Drop-Out-Analyse

	Untersuchungs- stichprobe	Unvollständiger Längsschnitt
<i>N</i>	921	510
männlich	50.4%	59.8%
weiblich	49.6%	40.2%
stärkster Geburtsjahrgang	1982	1981
<i>Sozio-ökonomischer Status</i>		
Schüler mit eigenem Zimmer im Elternhaus	84.7%	68.5%
<i>Besuchter Schulzweig</i>		
Hauptschüler	6.6%	21.2%
Realschüler	40.3%	36.5%
Gymnasiasten	29.3%	17.9%
Gesamtschüler	23.7%	24.3%
<i>Leistung</i>		
Schüler mit guten oder sehr guten Leistungen in...		
Mathematik	30.4%	26.1%
Deutsch	32.3%	25.1%
Englisch	30.2%	23.2%

Ferner ist anzunehmen, daß Schüler mit guten Schulleistungen weniger Probleme mit dem Ausfüllen des Fragebogens hatten. Eine längsschnittliche Zuordnung der Schüler gelang nur, wenn ein Kennwort nach vorgegebenen Regeln von den Schülern korrekt generiert wurde. Das Kennwort umfaßte z.B. die Anfangsbuchstaben des Vornamens des Vaters. Es ist wahrscheinlich, daß lei-

stungsschwächere Schüler hiermit mehr Probleme hatten, so daß für sie kein vollständiger Längsschnitt erzeugt werden konnte. Gleiches gilt für die Angabe zur Klasse.

Zusammenfassend bestand die Untersuchungsstichprobe aus Schülern der siebten bis zehnten Jahrgangsstufen aus besonders offenen, innovationsfreudigen Schulen des gesamten Bundesgebiets. Im Längsschnitt fielen insbesondere Hauptschüler und Schüler mit geringerem sozio-ökonomischen Status sowie Schüler mit weniger guten Leistungen heraus. Diese systematische Selektion schränkt die Aussagekraft der Ergebnisse ein, auch wenn diese Abhandlung nicht auf generalisierbare Befunde, sondern auf die Überprüfung psychologischer Hypothesen hinzielte.

5.2 Instrumentarium

Die Befragung der Schüler erfolgte mittels eines standardisierten Fragebogens. Dieser wurde zu allen Meßzeitpunkten *unverändert* eingesetzt. Der erste Teil des Fragebogens umfaßte 35 Items zu allgemeinen und spezifischen Selbstwirksamkeitserwartungen sowie zehn Items zur Prüfungsangst. Der zweite Teil enthielt acht Items zur subjektiven Relevanz von Unterrichtsfächern und Zeugnisnoten. Ein weiterer Teil thematisierte mit sieben Items Streßeinschätzungen sowie mit zehn Items Streßbewältigungsstrategien. Der vierte Teil des Fragebogens bezog sich mit 24 Items auf das Unterrichtsklima. Im fünften Teil wurden mit 14 Items körperliche Symptome und die Zufriedenheit in unterschiedlichen Lebensbereichen erfragt. Der letzte Teil umfaßte 18 Items zu Optimismus, Depression und schulischer Hilflosigkeit. Die Bearbeitung des Fragebogens wurde im Klassenverband unter Aufsicht eines Lehrers vorgenommen. Sie nahm etwa eine Unterrichtsstunde in Anspruch.

Für das Instrumentarium wurde in erster Linie auf bewährte Verfahren zurückgegriffen. Alle Instrumente wurden in einer Eichstichprobe¹ in Hinblick auf ihre Verteilung, Reliabilität, faktorielle Validität und kriteriumsorientierte Validität überprüft. Die Instrumente mit Skalen- und Itemkennwerten sind für die Gesamtstichprobe detailliert an anderer Stelle dokumentiert (Schwarzer & Jerusalem, 1999; im Internet: www.fu-berlin.de/gesund). Im folgenden werden die Konstruktionsprinzipien sowie ausgewählte Kennwerte der Reliabilität dargestellt. Besondere Aufmerksamkeit wird dabei der Überprüfung der faktoriellen Struktur der Selbstwirksamkeits- und Klimaitems zuteil. Die Dar-

¹ Die Eichstichprobe setzte sich aus den Schülern ($N = 463$) eines Gymnasiums in Nordrhein-Westfalen zusammen. Das Gymnasium hatte sich freiwillig zur Teilnahme an der Untersuchung bereit erklärt.

stellung dieser Kennwerte folgt jeweils im Anschluß an die Beschreibung der Instrumente eines Bereiches. *Deskriptive Statistiken der zentralen Tendenz sowie Geschlechtseffekte werden im Ergebnisteil abgehandelt.*

Gemäß der Hypothesen (Abschnitt 4.2) wurden drei Verfahren zur Messung von schulischen, sozialen und allgemeinen Selbstwirksamkeitserwartungen eingesetzt. Vier Skalen (Lehrerfürsorglichkeit, Lehrerbezugsnorm, supportive Schüler-Schüler-Beziehungen und Unterrichtszufriedenheit) wurden als Indikatoren des Mastery-Klimas herangezogen, zusätzlich wurde das Anforderungsniveau des Unterrichts aus Schülersicht gemessen. Als weitere Variablen wurden die Schulleistung, die Prüfungsangst, der Optimismus, die Depressivität, die Lebenszufriedenheit und die körperlichen Beschwerden der Schüler berücksichtigt.

5.2.1 Messung der Selbstwirksamkeitserwartungen

Nach einer Empfehlung Banduras (1997) sollten Items in Skalen zur Messung von Selbstwirksamkeitserwartungen unterschiedliche Schwierigkeitsgrade thematisieren, wobei sich die Breite aus der Fragestellung ergibt. Die Gewißheit oder Stärke der Überzeugungen wird über eine Ratingskala zum Ausdruck gebracht.

Spezifische Selbstwirksamkeitserwartungen sagen spezifisches Verhalten sehr genau vorher. Es ist jedoch fraglich, inwieweit Menschen bei der Bewältigung allgemeiner Lebensanforderungen von sehr spezifischen Überzeugungen profitieren. Für diese Untersuchung wurden daher Selbstwirksamkeitsmaße eingesetzt und z.T. neu entwickelt, die ein breiteres Spektrum von Anforderungssituationen aus dem *schulischen* und *sozialen* Bereich erfassen.

So mißt die Skala zur *schulischen Selbstwirksamkeitserwartung* (WIRKSCHUL) die Überzeugungsstärke, wiederkehrende schulische Anforderungen aus eigener Kraft bewältigen zu können, wie z.B. das Lösen von schwierigen Aufgaben an der Tafel. Die Skala zur *sozialen Selbstwirksamkeitserwartung* (WIRKSOZ) erfaßt hingegen die Kompetenzerwartung im Umgang mit sozialen Anforderungs- und Konfliktsituationen, wie z.B. das Lösen von sozialen Konflikten ohne Gewaltanwendung. Zusätzlich wurde mit der Skala zur Messung der *allgemeinen Selbstwirksamkeitserwartung* (Jerusalem & Schwarzer, 1999a) ein sehr bewährtes Instrument eingesetzt, welches Selbstwirksamkeitserwartungen auf einem allgemeinen Niveau thematisiert.

Die Items dieser Skalen wurden durchmischte im ersten Teil des Fragebogens dargeboten. Das *forced-choice* Antwortformat war vierstufig: „Trifft nicht zu“ (1), „Trifft kaum zu“ (2), „Trifft eher zu“ (3) und „Trifft genau zu“ (4). Zur Bestimmung der Rohwerte wurden die Summenwerte durch die Anzahl der Items

der jeweiligen Skala geteilt. *In dieser Weise wurde bei allen eingesetzten Instrumenten verfahren.*

Allgemeine Selbstwirksamkeitserwartung

Bei der allgemeinen Selbstwirksamkeitserwartung handelt es sich um eine Erwartungshaltung, welche „die subjektive Überzeugung zum Ausdruck bringt, aufgrund eigenen Handelns schwierige Anforderungen bewältigen zu können“ (Schwarzer, 1993b, S. 188). Im Gegensatz zu den spezifischen Selbstwirksamkeitserwartungen (Bandura, 1977, 1995, 1997) wird die allgemeine Selbstwirksamkeitserwartung als ein Aggregat über eine Vielzahl von Selbstwirksamkeitserwartungen unterschiedlicher Bereiche verstanden. Die allgemeine Selbstwirksamkeitserwartung kann einen persönlichen Schutzfaktor darstellen: „Die Zuversicht in die eigenen Fähigkeiten und Fertigkeiten ist im Sinne einer stabilen personalen Coping-Ressource zu verstehen“ (Schwarzer, 1993b, S. 189).

Tabelle 5.8
Allgemeine Selbstwirksamkeitserwartung (WIRKALL)
(Jerusalem & Schwarzer, 1999a)

-
1. Wenn mir jemand Widerstand leistet, finde ich Mittel und Wege, mich durchzusetzen.
 2. Die Lösung schwieriger Probleme gelingt mir immer, wenn ich mich darum bemühe.
 3. Es bereitet mir keine Schwierigkeiten, meine Absichten und Ziele zu verwirklichen.
 4. In unerwarteten Situationen weiß ich immer, wie ich mich verhalten soll.
 5. Auch bei überraschenden Ereignissen glaube ich, daß ich gut damit zurechtkommen werde.
 6. Schwierigkeiten sehe ich gelassen entgegen, weil ich mich immer auf meine Fähigkeiten verlassen kann.
 7. Was auch immer passiert, ich werde schon klarkommen.
 8. Für jedes Problem finde ich eine Lösung.
 9. Wenn eine neue Sache auf mich zukommt, weiß ich, wie ich damit umgehen kann.
 10. Wenn ein Problem auf mich zukommt, habe ich meist mehrere Ideen, wie ich damit fertig werde.
-

Die Skala zur Messung der allgemeinen Selbstwirksamkeitserwartung (WIRKALL) wurde Anfang der 80er Jahre von Jerusalem und Schwarzer entwickelt, wobei das Konstruktionsprinzip darin bestand, „eine Anforderung zu beschreiben, der das eigene Handlungspotential entgegen gestellt wird“

(Schwarzer, 1993b, S. 189). Die ursprüngliche Skala umfaßte 20 Items. Sie konnte ohne Qualitätsverlust auf zehn Items gekürzt werden (Jerusalem & Schwarzer, 1986). Seitdem ist die verkürzte Skala als Instrument zur Messung der allgemeinen Selbstwirksamkeitserwartung in zahlreichen Studien eingesetzt und in über 27 Sprachen² übersetzt worden. Die Skala erreichte in unterschiedlichen Stichproben eine Reliabilität zwischen $\alpha = .70$ und $\alpha = .90$. Konfirmatorische Faktorenanalysen konnten die Eindimensionalität in zahlreichen Stichproben bestätigen (Schwarzer, 1993a). Die Skala ist u. a. bedeutsam korreliert mit dem Optimismus, dem Selbstwertgefühl sowie mit internalen Kontrollüberzeugungen (Jerusalem & Mittag, 1995; Jerusalem & Schwarzer, 1992; Schwarzer & Hahn, 1994). Für die vorliegende Untersuchung wurde die Skala in einer leicht revidierten Fassung (Tabelle 5.8) eingesetzt.

Schulbezogene Selbstwirksamkeitserwartung

Für das Forschungsprojekt wurden acht Items zur Erfassung der schulbezogenen Selbstwirksamkeitserwartung entwickelt, die unterschiedliche schulische Anforderungssituationen ansprechen.

Tabelle 5.9
Schulische Selbstwirksamkeitserwartung (WIRKSCHUL)
(Jerusalem & Satow, 1999)

-
1. Ich kann auch die schwierigen Aufgaben im Unterricht lösen, wenn ich mich anstreuge.
 2. Es fällt mir leicht, neuen Unterrichtsstoff zu verstehen.
 3. Wenn ich eine schwierige Aufgabe an der Tafel lösen soll, glaube ich, daß ich das schaffen werde.
 4. Selbst wenn ich mal längere Zeit krank sein sollte, kann ich immer noch gute Leistungen erzielen.
 5. Wenn der Lehrer / die Lehrerin das Tempo noch mehr anzieht, werde ich die geforderten Leistungen kaum noch schaffen können. (-)
 6. Auch wenn der Lehrer / die Lehrerin an meinen Fähigkeiten zweifelt, bin ich mir sicher, daß ich gute Leistungen erzielen kann.
 7. Ich bin mir sicher, daß ich auch dann noch meine gewünschten Leistungen erreichen kann, wenn ich mal eine schlechte Note bekommen habe.
-

² Die unterschiedlichen Sprachversionen der Skala *Allgemeine Selbstwirksamkeitserwartung* sind im Internet unter www.fu-berlin.de/gesund dokumentiert.

Die Items wurden zunächst in der Eichstichprobe überprüft. Ein Item des Ausgangspools („Wenn meine Freunde mich besuchen, schaffe ich es trotzdem, meine Hausaufgaben zu erledigen“) wies in der Eichstichprobe eine sehr geringe Trennschärfe auf. Die Formulierung des Items erschien auch inhaltlich problematisch, da es sich bei der Erledigung von Hausaufgaben in Anwesenheit von Freunden in erster Linie um eine soziale Anforderungssituation handelt, denn hierbei spielen verstärkt soziale Kompetenzen eine Rolle. Das Item wurde eliminiert. Die endgültige Skala zur Messung schulischer Selbstwirksamkeitserwartung (Jerusalem & Satow, 1999) umfaßt somit sieben Items (Tabelle 5.9).

Im Unterschied zu fachspezifischen Skalen, mißt diese Skala schulische Kompetenzerwartungen auf einem mittleren Allgemeinheitsgrad. Es werden Anforderungssituationen angesprochen, die nicht spezifisch für ein bestimmtes Unterrichtsfach sind, sondern als typisch für den Schulalltag gelten können. Die Items benennen schwierigen Situation, wie z.B. krankheitsbedingte Lernrückstände oder das Lösen von Aufgaben an der Tafel. Durch die Antwortskala wird die Gewißheit zum Ausdruck gebracht, diese Anforderungen bewältigen zu können.

Soziale Selbstwirksamkeitserwartung

Bisher gab es kaum Instrumente, welche die Überzeugung von Jugendlichen zum Inhalt haben, soziale Konflikt- und Anforderungssituationen adäquat meistern zu können. Zur Messung der sozialen Selbstwirksamkeitserwartung wurden daher zwölf Items neu konstruiert. Jedes Item des Ausgangspools beschrieb eine schwierige soziale Anforderungssituation, die sich auf das Zusammenleben mit anderen Schülern bezog, nicht jedoch fach- oder unterrichtsgebunden war.

Aufgrund der Itemanalyse in der Eichstichprobe reduzierte sich die Skala auf acht Items, die ein breites Spektrum sozialer Anforderungs- und Konfliktsituationen thematisieren (Tabelle 5.10). Die Items sprechen unterschiedliche Themenbereiche sozialen Handelns an, z.B. adäquates Durchsetzungsvermögen in Gruppen („Ich traue mich zu sagen, was ich denke, auch wenn die anderen nicht meiner Meinung sind“), Kontrolle von Emotionen („Wenn mich jemand ärgert, schaffe ich es trotzdem, ruhig zu bleiben“) sowie kommunikative Kompetenzen („Wenn ich mich ganz traurig und mies fühle, schaffe ich es, mit den anderen darüber zu sprechen“).

Tabelle 5.10
Soziale Selbstwirksamkeitserwartung (WIRKSOZ)
(Satow & Mittag, 1999)

-
1. Ich traue mich zu sagen, was ich denke, auch wenn die anderen nicht meiner Meinung sind.
 2. Auch in einer ganz neuen Klasse kann ich schnell neue Freunde finden.
 3. Wenn mich jemand ungerecht behandelt, kann ich mich dagegen wehren.
 4. Wenn mich jemand ärgert, kann ich mich wehren, ohne Gewalt anzuwenden.
 5. Wenn ich etwas Falsches getan habe, schaffe ich es, mich zu entschuldigen.
 6. Wenn ich mich ganz traurig und mies fühle, schaffe ich es, mit den anderen darüber zu sprechen.
 7. Wenn mich jemand ärgert, schaffe ich es trotzdem, ruhig zu bleiben.
 8. Auch wenn mir alles zu viel wird, schaffe ich es, meine schlechte Laune nicht an anderen auszulassen.
-

Reliabilität und Stabilität in der Untersuchungsstichprobe

Wie erwähnt sind die Kennwerte aller Skalen und Items ausführlich bei Schwarzer und Jerusalem (1999, im Internet: www.fu-berlin.de/gesund) für die *Gesamtstichprobe* dokumentiert. Der Tabelle 5.11 können die Kennwerte der Reliabilität und Stabilität für die Untersuchungsstichprobe entnommen werden.

Tabelle 5.11
Reliabilität und Stabilität der Selbstwirksamkeitsskalen

	Items	Interne Konsistenz (Cronbachs α), 1. MZP	Guttman Split-half (1. MZP)	Stabilität zwischen 1. und 2. MZP (r_{t1t2})	Stabilität zwischen 1. und 3. MZP (r_{t1t3})
WIRKALL	10	.78	.75	.53	.46
WIRKSCHUL	7	.70	.68	.57	.48
WIRKSOZ	8	.56	.64	.50	.46

Die Skala zur Messung der sozialen Selbstwirksamkeitserwartung (WIRKSOZ) war mit $\alpha = .56$ weniger homogen als die beiden übrigen Selbstwirksamkeitsmaße. Hierin kommt zum Ausdruck, daß vielfältige soziale Anforderungs- und Konfliktsituationen abgefragt wurden. Zudem handelt es sich um ein völlig neu konstruiertes Instrument. Die Split-Half-Reliabilität fiel demgegenüber nur wenig kleiner aus als für die Skalen WIRKALL und WIRKSCHUL.

Auch die Stabilität ist vergleichbar mit der Stabilität der anderen Selbstwirksamkeitsskalen. Die Skala WIRKSOZ erfaßt damit ein heterogenes Merkmal hinreichend reliabel.

Insgesamt deuten die Kennwerte auf eine befriedigende bis gute Reliabilität der Skalen zur Messung der Selbstwirksamkeitserwartungen hin. Die Stabilität der Skalen bewegte sich in einem mittleren Bereich, wobei die Stabilität der schulischen Selbstwirksamkeitserwartung am höchsten ausfiel. Damit fällt die korrelative Stabilität in dieser Untersuchung etwas geringer aus als bei Helmke (1997).

Tabelle 5.12
Faktorielle Struktur der Selbstwirksamkeitsitems

	WIRKALL	WIRKSCHUL	WIRKSOZ
WIRKALL1	.51		
WIRKALL2	.46		
WIRKALL3	.47		
WIRKALL4	.44		
WIRKALL5	.44		
WIRKALL6	.64		
WIRKALL7	.55		
WIRKALL8	.60		
WIRKALL9	.49		
WIRKALL10	.55		
WIRKSCHUL1		.51	
WIRKSCHUL2		.50	
WIRKSCHUL3		.50	
WIRKSCHUL4		.61	
WIRKSCHUL5		.33	
WIRKSCHUL6		.50	
WIRKSCHUL7		.57	
WIRKSOZ1			.51
WIRKSOZ2			.41
WIRKSOZ3			.45
WIRKSOZ4			.39
WIRKSOZ5			.40
WIRKSOZ6			.31
WIRKSOZ7			.22
WIRKSOZ8			.20

Anmerkungen. WIRKALL: allgemeine Selbstwirksamkeitserwartung; WIRKSCHUL: schulische Selbstwirksamkeitserwartung; WIRKSOZ: soziale Selbstwirksamkeitserwartung.

Faktorielle Struktur in der Untersuchungsstichprobe

Die faktorielle Struktur der Items wurde im Ansatz einer konfirmatorischen Faktorenanalyse mit LISREL untersucht. Da dieses auf Itemebene geschah, wurden *Polychoric Correlations* (Jöreskog & Sörbom, 1988, 1989, 1993) zugrunde gelegt, die dem ordinalen Skalenniveau der Ratingskalen gerecht werden. Hinter den 25 Items wurden drei Faktoren angenommen: *Allgemeine Selbstwirksamkeitserwartung*, *Schulische Selbstwirksamkeitserwartung* und *Soziale Selbstwirksamkeitserwartung*. Der Modellfit ($\chi^2 = 838.7$; $df = 272$; $p = 0.00$; $RMSEA = 0.050$; $p = 0.57$) bestätigte die intendierte Struktur mit drei latenten Faktoren.

Tabelle 5.12 gibt die standardisierte Λ_X -Matrix zum ersten Meßzeitpunkt wieder. Die Ladungen für die allgemeine Selbstwirksamkeitserwartung lagen zwischen $\lambda = .44$ und $\lambda = .64$. Das Item mit der höchsten Ladung lautete: „Schwierigkeiten sehe ich gelassen entgegen, weil ich mich immer auf meine Fähigkeiten verlassen kann“. In vergleichbarer Größenordnung bewegten sich die Ladungen der schulischen Kompetenzerwartung ($\lambda = .33$ bis $\lambda = .61$). Die geringste Ladung wies dabei das negativ formulierte Item Nr. 5 auf. Das Item mit der höchsten Ladung lautete: „Selbst wenn ich mal längere Zeit krank sein sollte, kann ich immer noch gute Leistungen erzielen“. Die Itemladungen der sozialen Selbstwirksamkeitserwartung fielen erwartungsgemäß aufgrund der größeren Heterogenität der Skala geringer aus ($\lambda = .20$ bis $\lambda = .51$). Das Item mit der höchsten Ladung lautete: „Ich traue mich zu sagen, was ich denke, auch wenn die anderen nicht meiner Meinung sind“. Für die beiden Items, die sich auf eine Regulation der Emotionen beziehen („Wenn mich jemand ärgert, schaffe ich es trotzdem, ruhig zu bleiben“ und „Auch wenn mir alles zu viel wird, schaffe ich es, meine schlechte Laune nicht an anderen auszulassen“), resultierten Ladungen, die im Hinblick auf die faktorielle Struktur nicht mehr zu akzeptieren sind. Um auch diesen Aspekt der sozialen Interaktion – Regulation von Emotionen im Sinne einer sozialen Verträglichkeit – abzudecken, wurden die Items dennoch beibehalten. Für den Einsatz der Skala in einer anderen Untersuchung empfiehlt sich eine Neuformulierung dieser Items.

Die querschnittlichen Korrelationen zwischen den drei Selbstwirksamkeitsfaktoren (Tabelle 5.13) belegten, daß der latente Faktor *Allgemeine Selbstwirksamkeitserwartung* hoch mit den beiden spezifischeren Faktoren korreliert war. Hierbei ist zu berücksichtigen, daß es sich um die Korrelationen der latenten *Faktoren* handelt und nicht um die Korrelation der *Skalen*. Die Korrelationen der Skalen werden im Ergebnisteil (Abschnitt 6.1.3) berichtet. Sie fielen deutlich niedriger aus.

Tabelle 5.13

Querschnittliche Korrelationen der Selbstwirksamkeitsfaktoren

	WIRKALL	WIRKSCHUL	WIRKSOZ
WIRKALL	1.00		
WIRKSCHUL	.75*	1.00	
WIRKSOZ	.81*	.59*	1.00

Anmerkungen. WIRKALL: allgemeine Selbstwirksamkeitserwartung; WIRKSCHUL: schulische Selbstwirksamkeitserwartung; WIRKSOZ: soziale Selbstwirksamkeitserwartung.

* $p < .05$.

5.2.2 Messung des Unterrichtsklimas

Zur Messung des Unterrichtsklimas wurden fünf psychometrische Instrumente eingesetzt. Als Indikatoren des *Mastery-Klimas* dienen die Skalen *Lehrerfürsorglichkeit*, *Lehrerbezugsnormorientierung*, *Schüler-Sozialklima* und *Unterrichtszufriedenheit*. Die Skalen *Lehrerfürsorglichkeit*, *Schüler-Sozialklima* und *Unterrichtszufriedenheit* beruhen auf den *Landauer Skalen zum Sozialklima* (LASSO; Saldern und Littig, 1987). Schwarzer, Lange und Jerusalem (1982a) entwickelten das Instrument zur Erfassung der Lehrerbezugsnorm aus Schülersicht. Zusätzlich wurde das Anforderungsniveau aus Schülersicht ebenfalls mit Items der Landauer Skalen zum Sozialklima (Saldern & Littig, 1987) operationalisiert.

Alle Items dieser Skalen wurden durchmischt im vierten Teil des Fragebogens dargeboten. Die Instruktion lautete: „Denke jetzt an Deine Klassenlehrerin oder Deinen Klassenlehrer. Es geht hier nicht um Dich selbst, sondern um die ganze Klasse. Was fällt Dir auf, wenn Du den Unterricht beobachtest“. Das Antwortformat war wiederum vierstufig: „Trifft nicht zu“ (1), „Trifft kaum zu“ (2), „Trifft eher zu“ (3), „Trifft genau zu“ (4). Die Summenwerte der Skalen wurden durch die Anzahl der Items der jeweiligen Skala dividiert, um die Rohwerte zu bestimmen.

Fürsorglichkeit des Lehrers

Als erster Mastery-Indikator wurde eine Skala angewendet, die Fürsorglichkeit und Engagement des Lehrers zum Inhalt hat. Die Skala wurde den Landauer Skalen zum Sozialklima (Saldern & Littig, 1987) entnommen. Sie mißt, in wel-

chem Umfang die Schüler den Lehrer als unterstützend, hilfsbereit und an persönlichen Problemen interessiert erleben. Dabei bezieht sich die Fürsorge nicht nur auf das Geschehen im Unterricht, sondern auch auf die persönlichen Interessen und Probleme der Schüler (Tabelle 5.14).

Saldern und Littig (1987) ordnen die Skala der Dimension *Lehrer-Schüler-Beziehungen* zu. Sie berichten für die Subskala eine interne Konsistenz von $\alpha = .81$.

Tabelle 5.14
Fürsorglichkeit des Lehrers (FUERS)
(Saldern & Littig, 1987)

-
1. Wenn wir mit unserem Lehrer etwas bereden wollen, dann findet er auch Zeit dazu.
 2. Der Lehrer hilft uns wie ein Freund.
 3. Unser Lehrer kümmert sich um unsere Probleme.
 4. Unser Lehrer achtet zuwenig auf unsere Gefühle. (-)
 5. Unser Lehrer hilft jedem von uns, der Schwierigkeiten mit seinen Aufgaben hat.
 6. Im Unterricht reden wir oft über Fragen, die uns persönlich angehen.
 7. Unser Lehrer bemüht sich, unsere Wünsche so weit wie möglich zu erfüllen.
 8. Unser Lehrer ist meistens bereit, mit uns zu reden, wenn uns etwas nicht gefällt.
-

Schülerperzipierte Lehrerbezugsnorm

Als zweiter Mastery-Indikator wurde eine Skala zur Messung der Lehrerbezugsnormorientierung ausgewählt. Die Skala wurde von Schwarzer, Lange und Jerusalem (1982a) im Rahmen eines Forschungsprojekts zur schulspezifischen Sozialisation entwickelt. Sie mißt die von den *Schülern perzipierte* Bezugsnormorientierung des Lehrers (vgl. Rheinberg, 1979, 1999). Für diese Untersuchung wurde eine Kurzversion konstruiert, die vier Items umfaßt (Tabelle 5.15).

Ein hoher Skalenwert repräsentiert eine aus Schülersicht *individuelle* Bezugsnormorientierung des Klassenlehrers: Der Lehrer würdigt individuelle Leistungsfortschritte insbesondere bei schwächeren Schülern. Eine ausgeprägte Lehrerfürsorglichkeit und eine individuelle Lehrerbezugsnormorientierung werden als zentrale Merkmale individualisierter Lehrer-Schüler-Beziehungen angesehen, zusammen mit dem Schüler-Sozialklima und der Unterrichtszufriedenheit bilden sie das *Mastery-Klima*.

Tabelle 5.15
Lehrerbezugsnormorientierung (SPLB)
(Schwarzer, Lange & Jerusalem, 1982a)

1. Unser Lehrer bemerkt immer sofort, wenn sich meine Leistungen verbessern oder verschlechtern.
 2. Wenn ein schwacher Schüler sich verbessert, bedeutet das bei unserem Lehrer eine „gute Leistung“, auch wenn der Schüler immer noch unter dem Klassendurchschnitt liegt.
 3. Wenn ein Schüler seine Leistungen gegenüber früher verbessert, so wird er dafür vom Lehrer besonders gelobt.
 4. Unser Lehrer lobt auch die schlechtesten Schüler, wenn er merkt, daß sie sich verbessert haben.
-

Schüler-Sozialklima

Neben den individualisierten Lehrer-Schüler-Beziehungen, welche mit den Skalen Lehrfürsorglichkeit und Lehrerbezugsnorm erfaßt wurden, berücksichtigt das Mastery-Klima auch das soziale Klima unter den Schülern. Die vier Items zum Schüler-Sozialklima wurden der Acht-Item-Skala *Hilfsbereitschaft* der Landauer Skalen zum Sozialklima (Saldern & Littig, 1987) entnommen. Sie sprechen neben der sozialen Unterstützung durch Mitschüler auch das gegenseitige Vertrauen und die Rücksichtnahme in der Klasse an (Tabelle 5.16). Die Skala dient als weiterer Indikator des Mastery-Klimas. Sie nimmt dabei nicht wie die *Lehrerfürsorglichkeit* oder die *Lehrerbezugsnormorientierung* auf die Lehrer-Schüler-Beziehungen, sondern auf die Schüler-Schüler-Beziehungen Bezug.

Tabelle 5.16
Schüler-Sozialklima (SOZK)
(Saldern & Littig, 1987)

1. Diejenigen, die ihre Aufgaben verstanden haben, warten, bis die anderen auch so weit sind.
 2. Wenn jemand in der Klemme ist, kann er sich auf seine Mitschüler verlassen.
 3. Wenn jemand in einer Klassenarbeit schlecht abschneidet, finden sich meistens Mitschüler, die ihn trösten.
 4. Wenn jemand Schwierigkeiten hat, helfen ihm die Mitschüler.
-

Die Hilfsbereitschaft der Schüler kann nach Saldern und Littig (1987) als ein wesentliches Merkmal der subjektiven Lernumwelt angesehen werden. Die ursprüngliche Skala wies in der Untersuchung von Saldern und Littig eine interne Konsistenz von $\alpha = .82$ auf.

Unterrichtszufriedenheit

Als letzter Mastery-Indikator wurden Items ausgewählt, die das Interesse und die Freude der Schüler am Unterricht thematisieren (Tabelle 5.17). Die Skala erfaßt somit auch die affektiv-emotionale Bewertung des Unterrichtsgeschehens. Die drei Items wurden aus der Sechs-Item-Skala *Zufriedenheit mit dem Unterricht* der Landauer Skalen zum Sozialklima (Saldern & Littig, 1987) entnommen. Die ursprüngliche Subskala von Saldern und Littig (1987) wies eine interne Konsistenz von $\alpha = .86$ auf.

Tabelle 5.17
Unterrichtszufriedenheit (UZUF)
(Saldern & Littig, 1987)

1. Den meisten machen die im Unterricht besprochenen Themen Spaß.
2. Im allgemeinen haben wir viel vom Unterricht.
3. Wir finden den Unterricht meist interessant.

Anforderungsniveau

Das erlebte Anforderungsniveau (Tabelle 5.18) wurde mit fünf Items der Landauer Skalen zum Sozialklima (Saldern & Littig, 1987) operationalisiert. Das Anforderungsniveau ist gegenüber der objektiven Überforderung der Schüler, dem *Leistungsdruck*, abzugrenzen. Die Skala thematisiert entsprechend vor allem den Eindruck der Schüler, für gute Noten viel leisten zu müssen.

Nach dem Rahmenmodell wird erwartet, daß das Mastery-Klima das erlebte Anforderungsniveau nicht senkt. Die ursprüngliche Skala der Landauer Skalen zum Sozialklima bestand aus sieben Items. In der Untersuchung von Saldern und Littig (1987) erreichte sie eine Reliabilität von $\alpha = .76$. Die Autoren ordnen die Skala der Dimension *Merkmale des Unterrichts* zu.

Tabelle 5.18
Anforderungsniveau (AN)
(Saldern & Littig, 1987)

-
1. Wir müssen auch am Wochenende Aufgaben machen, wenn wir schaffen wollen, was von uns verlangt wird.
 2. Die meisten von uns kommen kaum nach mit den Hausaufgaben.
 3. Der Unterricht geht so schnell, daß wir oft nicht mitkommen.
 4. Wir müssen für eine gute Note in dieser Klasse sehr viel leisten.
 5. Wenn jemand einige Tage fehlt, muß er sich sehr anstrengen, um wieder Anschluß zu finden.
-

Reliabilität und Stabilität in der Untersuchungsstichprobe

Die Skala *Lehrerfürsorglichkeit* wies bei acht Items die höchste interne Konsistenz ($\alpha = .84$) auf. Aber auch die übrigen Mastery-Indikatoren und die Skala *Anforderungsniveau* erreichten bei deutlich weniger Items Kennwerte, die auf eine befriedigende Reliabilität der Skalen hindeuten (Tabelle 5.19).

Tabelle 5.19
Reliabilität und Stabilität der Klimaskalen

	Items	Interne Konsistenz (Cronbachs α), 1. MZP	Guttman Split-half (1. MZP)	Stabilität zwischen 1. und 2. MZP (r_{t1t2})	Stabilität zwischen 1. und 3. MZP (r_{t1t3})
<i>Mastery-Indikatoren</i>					
FUERS	8	.84	.85	.43	.42
SPLB	4	.65	.62	.38	.35
SOZK	4	.69	.69	.53	.44
UZUF	3	.68	.64	.41	.35
<i>Anforderungsniveau</i>					
AN	5	.64	.59	.42	.38

Anmerkungen. FUERS: Lehrerfürsorglichkeit; SPLB: Lehrerbezugsnorm; SOZK: Schüler-Sozialklima; UZUF: Unterrichtszufriedenheit; AN: Anforderungsniveau.

Die Skala zur Messung des Schüler-Sozialklimas erzielte die höchste Stabilität mit $r_{t1t2} = .53$ und $r_{t1t3} = .44$. Die geringste Stabilität war für die Beurteilung

der Lehrerbezugsnorm zu verzeichnen ($r_{112} = .38$ und $r_{113} = .35$). Offenbar veränderte sich damit die Bewertung der Lehrerbezugsnorm stärker als die Einschätzung der Lehrerfürsorglichkeit.

Faktorielle Struktur in der Untersuchungsstichprobe

Die Überprüfung der faktoriellen Struktur erfolgte wiederum als konfirmatorische Faktorenanalyse mit LISREL 8.12a. Dazu wurden fünf Klimafaktoren (Lehrerfürsorglichkeit, Lehrerbezugsnorm, Schüler-Sozialklima, Unterrichtszufriedenheit, Anforderungsniveau) spezifiziert, welche die Ausprägung der dazugehörigen Items erklären sollten. Das Modell wies einen sehr guten Modellfit auf ($\chi^2 = 697.1$; $df = 242$; $p = 0.00$; $RMSEA = 0.048$; $p = 0.76$). Die faktorielle Struktur der Items konnte bestätigt werden.

Die Ladungen (Tabelle 5.20) bewegten sich insgesamt auf einem hohen Niveau. Die Items mit der höchsten Ladung ($\lambda = .75$) auf dem Faktor *Lehrerfürsorglichkeit* lauteten: „Der Lehrer hilft uns wie ein Freund“ und „Unser Lehrer kümmert sich um unsere Probleme“. Es ist damit offensichtlich, daß die Skala die freundschaftliche, unterstützende Zuwendung des Klassenlehrers zu den individuellen Problemen der Schüler aus Schülersicht erfaßt. Auch die Ladungen der Skala *Lehrerbezugsnorm* bestätigten die guten psychometrischen Skalenkennwerte ($\lambda = .48$ bis $\lambda = .73$). Kennzeichnend für die Skala war das Item „Unser Lehrer lobt auch die schlechtesten Schüler, wenn er merkt, daß sie sich verbessert haben“ ($\lambda = .73$). Zwei Items wiesen besonders hohe Ladungen ($\lambda = .74$ und $\lambda = .78$) auf dem Faktor *Schüler-Sozialklima* auf. Sie lauteten: „Wenn jemand in der Klemme ist, kann er sich auf seine Mitschüler verlassen“ und „Wenn jemand Schwierigkeiten hat, helfen ihm die Mitschüler“. Obwohl die Skala *Unterrichtszufriedenheit* nur drei Items umfaßte, erreichten alle Items eine Ladung über $\lambda = .60$. Charakteristisch war das Item: „Wir finden den Unterricht meist interessant“. Die Faktorladungen der Items zur Messung des Anforderungsniveaus bewegten sich zwischen $\lambda = .39$ und $\lambda = .64$, wobei das Item mit der höchsten Ladung „Wir müssen für eine gute Note in der Klasse sehr viel leisten“ und das Item mit der geringsten Ladung „Der Unterricht geht so schnell, daß wir oft nicht mitkommen“ lautete, was unterstreicht, daß die Skala *primär* das erlebte *Anforderungsniveau* erfaßt.

Tabelle 5.20
Faktorielle Struktur der Klimaitems
(vollständig standardisierte Lösung, 1. MZP)

	FUERS	SPLB	SOZK	UZUF	AN
FUERS1	.68				
FUERS2	.75				
FUERS3	.75				
FUERS4	.48				
FUERS5	.64				
FUERS6	.45				
FUERS7	.64				
FUERS8	.69				
SPLB1		.51			
SPLB2		.48			
SPLB3		.54			
SPLB4		.73			
SOZK1			.36		
SOZK2			.74		
SOZK3			.63		
SOZK4			.78		
UZUF1				.65	
UZUF2				.63	
UZUF3				.68	
AN1					.52
AN2					.43
AN3					.39
AN4					.64
AN5					.55

Anmerkungen. FUERS: Lehrerfürsorglichkeit; SPLB: Lehrerbezugsnorm; SOZK: Schüler-Sozialklima; UZUF: Unterrichtszufriedenheit; AN: Anforderungsniveau.

Die querschnittlichen Korrelationen der Faktoren (Tabelle 5.21) bestätigten, daß die Mastery-Indikatoren hoch korreliert sind. So fand sich insbesondere ein enger Zusammenhang zwischen der Lehrerfürsorglichkeit und der Lehrerbezugsnorm ($r = .74$), zwischen der Lehrerfürsorglichkeit und der Unterrichtszufriedenheit ($r = .72$) sowie zwischen der Lehrerbezugsnorm und der Unterrichtszufriedenheit ($r = .60$). Das Anforderungsniveau war hingegen nur sehr geringfügig mit den Indikatoren des Mastery-Klimas korreliert, worin sich bereits die *Unabhängigkeit* des Mastery-Klimas vom Anforderungsniveau

abzeichnete. Eine detaillierte Analyse zu dieser Frage findet sich im Ergebnisteil.

Tabelle 5.21
Querschnittliche Korrelationen der Klimafaktoren

	FUERS	SPLB	SOZK	UZUF	AN
FUERS	1.00				
SPLB	.74*	1.00			
SOZK	.39*	.48*	1.00		
UZUF	.72*	.60*	.43*	1.00	
AN	-.10*	.12*	-.06	-.06	1.00

Anmerkungen. FUERS: Lehrerfürsorglichkeit; SPLB: Lehrerbezugsnorm; SOZK: Schüler-Sozialklima; UZUF: Unterrichtszufriedenheit; AN: Anforderungsniveau.

* $p < .05$

5.2.3 Sonstige Instrumente

Weitere Instrumente wurden eingesetzt, um die Schulleistung, die Prüfungsangst und das Befinden der Schüler zu erheben. Insbesondere das Befinden wurde in vielen Untersuchungen zur Wirkung des Klimas kaum berücksichtigt. In der vorliegenden Untersuchung kamen Skalen zum Optimismus, zur Depressivität, zur Lebenszufriedenheit und zu körperlichen Beschwerden als Indikatoren des Befindens der Schüler zum Einsatz. Auch diese Skalen sind ausführlich bei Schwarzer und Jerusalem (1999) dokumentiert.

Schulnotenindex

Die Schulleistung war hauptsächlich für zwei Fragestellungen von Bedeutung. Es galt zum einen zu untersuchen, inwieweit sich die Veränderung des Mastery-Klimas auch auf die Schulleistung auswirkte, denn hier hätte sich ein Effekt auf die Selbstwirksamkeitserwartungen *über* die Schulleistung zeigen können. Zum anderen interessierte die Schulleistung als Variable in einem längsschnittlichen Mediator-Modell. Im Ergebnis sollte das Mastery-Klima die Selbstwirksamkeitserwartungen und das Befinden der Schüler *nicht* über eine wohlwollendere Leistungsbewertung fördern.

Zur Bestimmung der Schulleistung wurde ein Index aus den selbstberichteten Schulnoten in den Hauptfächern *Mathematik, Deutsch* und *Englisch* im letzten

Zeugnis gebildet. Dazu wurden die Schüler im Schülerfragebogen gebeten, durch Ankreuzen ihre Noten in dem jeweiligen Fach anzugeben. Die Antwortkategorien lauteten (Kodierung in Klammern): Sehr gut (5), gut (4), befriedigend (3), ausreichend (2), mangelhaft oder ungenügend (1). Die Zeugnisnoten 5 und 6 bildeten somit *eine* Kategorie, die mit *Eins* kodiert wurde. Sehr gute Leistungen wurden hingegen mit *Fünf* kodiert. Ein hoher Wert auf dem Notenindex steht damit für gute Schulleistungen. Ein solcher Index kann die tatsächliche Leistung nur annähernd erfassen. Er wird durch eine Vielzahl von Störfaktoren (z.B. Antworttendenzen) beeinflusst. Der Notenindex wurde jedoch vor allem berücksichtigt, um sicherzustellen, daß die Effekte des Mastery-Klimas nicht auf eine wohlwollendere Leistungsbewertung zurückgehen.

Prüfungsangst

Zur Messung der Prüfungsangst kam die deutsche Version des *Test Anxiety Inventory* (TAI-G; Hodapp, 1991; Hodapp, Laux und Spielberger, 1982; Musch & Bröder, 1999) zum Einsatz. Die vorangestellte Instruktion lautete: „Was fühlst Du, wenn Du an Deine nächste Klassenarbeit denkst?“ Das Antwortformat war wiederum vierstufig: „Trifft nicht zu“ (1), „Trifft kaum zu“ (2), „Trifft eher zu“ (3), „Trifft genau zu“ (4). Anhand der Instruktion läßt sich kaum beurteilen, inwieweit eine Zustandsangst oder eine Ängstlichkeit im Sinne einer stabilen Eigenschaft erfaßt wurde. Aufschluß hierzu gibt jedoch die Stabilität der Skala, die weiter unten berichtet wird. Offen bleibt auch, wie spezifisch die gemessene Angst ist: Bezieht sie sich allein auf die Klassenarbeit in einem bestimmten Fach oder auf Prüfungssituationen generell. Insgesamt muß die Messung der Prüfungsangst daher als unbefriedigend angesehen werden. Sie soll jedoch lediglich als ein globaler Indikator für die Angst in Prüfungssituationen eingesetzt werden.

Tabelle 5.22
Emotionality-Skala (EMO) des TAI-G
(Hodapp, 1991)

-
1. Ich spüre ein komisches Gefühl im Magen.
 2. Ich bin am ganzen Körper verkrampft.
 3. Das Herz schlägt mir bis zum Hals.
 4. Ich habe ein beklemmendes Gefühl.
 5. Ich bin aufgeregt.
-

Die fünf Items zur Aufgeregtheit bringen körperliche und gefühlsmäßige Symptome der Aufgeregtheit angesichts einer bevorstehenden Klassenarbeit zum Ausdruck (Tabelle 5.22). Diese Symptome ähneln sehr denjenigen, die in der Skala *Körperliche Beschwerden* thematisiert werden. Und tatsächlich zeigte sich, daß das allgemeine physische und psychische Befinden einen Beitrag zur Aufklärung der Aufgeregtheit angesichts einer bevorstehenden Klassenarbeit leistete (vgl. Abschnitt 0 im Ergebnisteil). Mit fünf weiteren Items des TAI-G wurde die Besorgnis angesichts einer bevorstehenden Klassenarbeit quantifiziert (Tabelle 5.23).

Tabelle 5.23
Worry-Skala (WORRY) des TAI-G
(Hodapp, 1991)

-
1. Ich mache mir Sorgen, ob ich auch alles schaffe.
 2. Ich frage mich, ob meine Leistung ausreicht.
 3. Ich denke daran, wie wichtig mir ein gutes Ergebnis ist.
 4. Ich bin besorgt, daß etwas schief laufen könnte.
 5. Ich denke daran, was passiert, wenn ich schlecht abschneide.
-

Optimismus

Das Konzept des Optimismus ist vielfältig (vgl. Schwarzer, 1993b; Schwarzer, 1994). Optimismus kann als eine Art von Bewältigungsstrategie angesichts widriger Umstände angesehen werden (vgl. Epstein & Meier, 1989) und als ein Schutzfaktor, der auf der Überzeugung beruht, daß sich die Dinge auch ohne eigenes Zutun positiv entwickeln (vgl. Scheier & Carver, 1992). Optimismus als verzerrte Wahrnehmung der Realität ist nicht zuletzt eine gesunde Selbsterhaltungsstrategie (Schwarzer, 1993b). Er kann jedoch auch negative Folgen zeitigen, wenn z.B. ein objektives Risiko unterschätzt wird und notwendige, präventive Handlungen unterlassen werden (Weinstein, 1988). Man unterscheidet daher auch zwischen einem gesunden, funktionalen Optimismus und einem wenig nützlichen, defensiven Optimismus (Taylor, 1989).

Tabelle 5.24
Optimismus (OPTIM)
(Satow & Schwarzer, 1999)

-
1. Ich glaube, daß mein Leben sich positiv entwickeln wird.
 2. Ich glaube, daß es in der Schule noch weiter aufwärts geht.
 3. Mein Leben macht Spaß.
 4. Es kommt immer wieder etwas, worauf ich mich freuen kann.
 5. Ich werde eines Tages in meinem Beruf erfolgreich sein.
 6. Ich glaube, daß ich in Zukunft mit meinen Freunden und Freundinnen viel Spaß haben werde.
 7. Was auch immer passiert, ich kann die gute Seite daran sehen.
 8. Meine Zukunft sieht gut aus.
-

Für die vorliegende Untersuchung wurden acht Items entwickelt, die Optimismus im Sinne einer generalisierten Ergebniserwartung ansprechen (Tabelle 5.24). Durch die Zustimmung zu den Items bringen die Schüler ihre Überzeugung zum Ausdruck, daß sich die Zukunft positiv entwickeln wird. Dieser Optimismus grenzt sich gegenüber der allgemeinen Selbstwirksamkeitserwartung dadurch ab, daß für den günstigen Ausgang der Dinge nicht notwendigerweise eigenes Handeln erforderlich ist. Es werden also nicht eigene Kompetenzen im Umgang mit zukünftigen Problemen angesprochen. Allgemeine Selbstwirksamkeitserwartung und Optimismus sind natürlicherweise auf mittlerem bis hohem Niveau korreliert, da sich ein ausgeprägter Optimismus in der optimistischen Einschätzung eigener Kompetenzen widerspiegelt. Die Korrelationen (s. Ergebnisteil) bestätigten dies.

Die Skala wurde neben den Skalen zur Depressivität, zur Lebenszufriedenheit und zu körperlichen Beschwerden als Indikator des physischen und psychischen Befindens eingesetzt.

Depressivität

Als weiterer Indikator des physischen und psychischen Befindens kam die Skala *Depressivität* zum Einsatz (Tabelle 5.25). Das Instrument dient der Erfassung depressiver Tendenzen bei Schülern. Die Items thematisieren depressive emotionale und motivationale Lagen. Hohe Korrelationen fanden sich mit Bedrohungs- und Verlusteinschätzungen, mit einem emotionsorientierten Coping, körperlichen Beschwerden, geringer Lebenszufriedenheit und geringem Optimismus (Schwarzer & Jerusalem, 1999). Die Items wurden durchmischt mit den Items der Optimismusskala und weiteren Items im letzten Teil des Schülerfrage-

bogens dargeboten. Vorgegeben wurde eine vierstufige Ratingskala: „Trifft nicht zu“ (1), „Trifft kaum zu“ (2), „Trifft eher zu“ (3), „Trifft genau zu“ (4).

Tabelle 5.25
Depressivität (DEPRES)
(Schwarzer & Bäßler, 1999)

-
1. Ich bin oft ohne Grund traurig.
 2. Ich habe nie Lust, mit Freunden wegzugehen.
 3. Kein Mensch versteht mich.
 4. Wenn die anderen Spaß haben, kann ich nicht mitlachen.
 5. Mir ist selten zum Lachen zumute.
 6. Ich sitze oft da und möchte nichts tun.
-

Lebenszufriedenheit

Die Zufriedenheit mit unterschiedlichen Lebensbereichen und dem Leben insgesamt war Thema der Skala *Lebenszufriedenheit* (Tabelle 5.26). Vorgegeben wurde eine sechsstufige Antwortskala mit den Verankerungspunkten „sehr unzufrieden“ (1) und „sehr zufrieden“ (6). Die Items wurden durchmisch mit anderen Items im letzten Teil des Fragebogens plaziert.

Tabelle 5.26
Lebenszufriedenheit (SATIS)
(Mittag, 1999)

-
- Wie zufrieden bist Du im Augenblick mit den folgenden Bereichen Deines Lebens?
1. ...mit Deiner Schule.
 2. ...mit Deiner Freizeit.
 3. ...mit Deiner Familie.
 4. ...mit Deinen Freunden.
 5. ...mit Dir selbst.
 6. ...mit Deinem Leben insgesamt.
-

Körperliche Beschwerden

Neben Optimismus, Depressivität und Lebenszufriedenheit, die eher das psychische Befinden der Schüler ansprechen, wurde die Skala *Körperliche Beschwerden* als Indikator physischer Symptome eingesetzt (Tabelle 5.27). Die Items sind dem Berner Fragebogen zum Wohlbefinden Jugendlicher (Grob, Lüthi, Kaiser, Flammer, Mackinnon & Wearing, 1991) entnommen. Die Antwortskala sah vier Kategorien vor: „nie“ (1), „selten“ (2), „oft“ (3), „sehr oft“ (4).

Die Verteilung der Skala war linkssteil, da viele körperliche Symptome nur selten auftraten. Die meisten Schüler gaben an, etwa vier der vorgegebenen körperlichen Beschwerden zumindest zeitweise (selten) an sich zu beobachten (Satow & Bäßler, 1998). Einschlafprobleme und Kopfschmerzen wurden am häufigsten genannt.

Tabelle 5.27
Körperliche Beschwerden (COMPL)
(Grob, Lüthi, Kaiser, Flammer, Mackinnon & Wearing, 1991)

Wie häufig kam es in den letzten Wochen vor, daß...

1. ... Du Magenschmerzen hattest?
 2. ... Du krank warst und nicht zur Schule gehen konntest?
 3. ... Du unter Appetitlosigkeit gelitten hast?
 4. ... Dir schwindlig war?
 5. ... Du nicht einschlafen konntest?
 6. ... Dir plötzlich ganz heiß wurde?
 7. ... Du starke Kopfschmerzen hattest?
 8. ... Dich starkes Herzklopfen oder Herzstiche plagten?
-

Reliabilität und Stabilität in der Untersuchungsstichprobe

Alle Skalen überzeugten in Hinblick auf ihre interne Konsistenz und ihrer Split-Half-Reliabilität (Tabelle 5.28). Insbesondere die Skala *Optimismus* erreichte sowohl eine ausgezeichnete Konsistenz ($\alpha = .83$) als auch eine ausgezeichnete Split-Half-Reliabilität ($REL = .84$). Die Skala *Depressivität* wies indes die schwächste interne Konsistenz ($\alpha = .65$) und Split-Half-Reliabilität auf ($REL = .66$).

Betrachtet man die Stabilität über den Untersuchungszeitraum, so zeigte sich die höchste Stabilität für die körperlichen Beschwerden ($r_{113} = .50$). Die

geringste Stabilität wies hingegen die Skala *Lebenszufriedenheit* $r_{t1t3} = .44$ auf. Bei den Skalen zur Prüfungsangst war aufgrund der Instruktion nicht entscheidbar, ob es sich bei der gemessenen Angst eher um eine Zustandsangst oder um Ängstlichkeit handelt. Der Vergleich der Stabilität dieser Skalen mit der Stabilität der übrigen Skalen deutet darauf hin, daß sich die Prüfungsangst etwa in dem Umfang veränderte wie der Schulnotenindex. Skalen, die eher ein Trait erfassen (Optimismus), waren vom ersten zum zweiten Meßzeitpunkt weniger stabil. Man kann daher annehmen, daß die Skalen zur Prüfungsangst eher eine stabile Ängstlichkeit der Schüler angesichts von Prüfungssituationen erfaßten.

Tabelle 5.28
Reliabilität und Stabilität der sonstigen Skalen

	Items	Interne Konsistenz (Cronbachs α), 1. MZP	Guttman Split-half (1. MZP)	Stabilität zwischen 1. und 2. MZP (r_{t1t2})	Stabilität zwischen 1. und 3. MZP (r_{t1t3})
<i>Leistung und Prüfungsangst</i>					
LEIST	Index	-	-	.53	.46
WORRY	5	.78	.77	.50	.45
EMO	5	.82	.79	.53	.46
<i>Befinden</i>					
OPTIM	8	.83	.84	.46	.49
DEPRES	6	.65	.66	.48	.45
SATIS	6	.72	.74	.49	.44
COMPL	8	.73	.69	.53	.50

Anmerkungen. LEIST: Schulnotenindex über die Fächer Mathematik, Deutsch und Englisch; WORRY: Besorgnis; EMO: Aufgeregtheit; OPTIM: Optimismus; DEPRES: Depressivität; SATIS: Lebenszufriedenheit; COMPL: körperliche Beschwerden.

5.2.4 Zusammenfassung

Alle Instrumente waren hinsichtlich ihrer Reliabilität und Konstruktvalidität in einer Eichstichprobe und in der Gesamtstichprobe sorgfältig überprüft worden (Schwarzer & Jerusalem, 1999). Daher verwundert es nicht, daß sie auch in der Untersuchungsstichprobe befriedigende bis ausgezeichnete Kennwerte aufwiesen.

Es wurden drei Skalen zur Messung von schulischen, sozialen und allgemeinen Selbstwirksamkeitserwartungen eingesetzt, die sich insgesamt als reliabel und im Vergleich mit den übrigen Instrumenten als eher stabil erwiesen. Die Skala zur Messung sozialer Kompetenzerwartungen war neu konstruiert worden; die konfirmatorische Faktorenanalyse ergab Hinweise auf eine sinnvolle Weiterentwicklung der Items. Die intendierte faktorielle Struktur der Selbstwirksamkeitsitems konnte insgesamt mittels der konfirmatorischen Faktorenanalyse in der Untersuchungsstichprobe bestätigt werden.

Als Indikatoren des Mastery-Klimas kamen vier Skalen zur Anwendung, die ebenfalls hinsichtlich ihrer Kennwerte überzeugten, obwohl sie z.T. nur wenige Items umfaßten. Die fünfte Klimaskala *Anforderungsniveau* wies im Vergleich zu den übrigen Klimaskalen die geringste interne Konsistenz auf. Die faktorielle Struktur der fünf Klimaskalen fand ebenfalls ihre Bestätigung in einer konfirmatorischen Faktorenanalyse.

Zur Leistungsmessung wurde ein Index aus den Schulnoten in den Fächern Mathematik, Deutsch und Englisch gebildet. Seine Stabilität über den Untersuchungszeitraum entsprach der Stabilität der Prüfungsangst. Das psychische und physische Befinden der Schüler wurde anhand der Skalen *Optimismus*, *Depressivität*, *Lebenszufriedenheit* und *Körperliche Beschwerden* quantifiziert, deren psychometrische Qualitäten ebenfalls als befriedigend bis gut anzusehen sind. Die Korrelationen dieser Skalen, Mittelwerte und Geschlechtseffekte werden im Ergebnisteil behandelt.

5.3 Die Analyse hierarchischer Daten

Es besteht weitgehende Einigkeit in der Frage, daß menschliches Erleben und Verhalten oft erst aus dem *sozialen* Kontext heraus zu verstehen ist. Es ist die Gemeinschaft (oder auch das Kollektiv), welche unser Erleben und Verhalten maßgeblich formt. In jüngster Zeit wurde verstärkt darauf hingewiesen, daß die psychologische Forschung den Einfluß von Merkmalen eines Kollektivs auf das Erleben und Verhalten des Einzelnen nur selten berücksichtigt (vgl. Camp, Saylor & Harer, 1997; Hox, 1993, 1995; Hox & Kreft, 1994; Lincoln & Zeitz, 1980). Als Beispiel für solche Effekte sei hier genannt, daß die Arbeitslosenquote eines Bundeslandes als kollektive Eigenschaft neben dem individuellen Beschäftigungsstatus die individuelle Beurteilung des Lebensstandards beeinflussen kann (Engel, 1998). In vergleichbarer Weise wird in der vorliegenden Abhandlung angenommen, daß das Klima als Eigenschaft der Klassengemeinschaft die Entwicklung von Selbstwirksamkeitserwartungen formt.

Individuen und der soziale Kontext, in dem sie leben, bilden ein *hierarchisches System* mit einer fast unendlichen Anzahl von Ebenen: Menschen fin-

den sich zu Gruppen zusammen, die wiederum Bestandteil größerer Gemeinschaften sind, welche nicht zuletzt Staaten und Staatengemeinschaften konstituieren. Ein solches hierarchisches System stellt insbesondere die Schule dar: Schüler sind Klassen zugeordnet, die wiederum in Schulen zusammengefaßt sind usw. Die Anzahl der geschachtelten Ebenen, die man hier berücksichtigen kann, ist nahezu beliebig. So lassen sich auch innerhalb der Klassen weitere soziale Ebenen ausmachen: Schüler bilden Cliques, mehrere Cliques können als Freundschaftskreise aufgefaßt werden. Die Frage, welche dieser sozialen Ebenen den maßgeblichsten Einfluß auf das Erleben und Verhalten des Einzelnen nimmt, kann kaum beantwortet werden.

Um der hierarchischen Struktur gerecht zu werden, bedarf es statistischer Analysen, die in der Lage sind, unterschiedliche Analyseebenen (z.B. Schüler- und Klassenebene) zu berücksichtigen. Solche Ansätze werden Mehrebenenanalysen genannt.

Ein wesentlicher Vorteil von Mehrebenenanalysen besteht darin, daß sie Aussagen für mehrere Analyseebenen und z.T. auch über Interaktionen zwischen den Ebenen zulassen. Aussagen über Ebenen, die nicht als Analyseebenen berücksichtigt wurden, sind weiterhin nicht zulässig. Werden solche Aussagen trotzdem getroffen, spricht man von einem *ökologischen Fehlschluß* (vgl. Hox, 1995; Hox & Kreft, 1994; Schneider, 1991).

Die Gefahr eines ökologischen Fehlschlusses liegt darin, daß Ergebnisse, die allein auf einer bestimmten Ebene gültig sind, auf eine andere Ebene generalisiert werden, obwohl sich die Sachverhalte auf den Ebenen in Wahrheit unterschiedlich darstellen. Beispiele hierfür wurden schon früh in der Literatur beschrieben (Robinson, 1950). In der vorliegenden Abhandlung zeigte sich in diesem Zusammenhang, daß auf Klassenebene das Anforderungsniveau negativ mit der allgemeinen Selbstwirksamkeitserwartung korreliert war, während auf Schülerebene *kein* Zusammenhang bestand.

Aus dieser Tatsache folgt natürlicherweise, daß psychologische Konstrukte, die oft als Faktoren hinter mehreren Variablen aufgefaßt werden, nicht notwendigerweise auf allen Ebenen eines sozialen Systems zu finden sind. So gibt es etwa Hinweise, daß auf der Ebene von Familien *ein* (vererbter) Intelligenzfaktor (g), auf Ebene der Kinder jedoch weitere (erworbene) Intelligenzfaktoren existieren (Hox, 1995). Für die vorliegende Abhandlung war es daher notwendig zu untersuchen, ob sich das *Mastery-Klima* als Faktor hinter den Indikatoren *Lehrerfürsorglichkeit*, *Lehrerbezugsnorm*, *Schüler-Sozialklima* und *Unterrichtszufriedenheit* sowohl auf Schüler- als auch auf Klassenebene bewährte.

Um Merkmale eines Kollektivs zu untersuchen, werden häufig Individualdaten aggregiert. Hierbei kann es jedoch zu einer Bedeutungsverschiebung kommen, was als *aggregation bias* diskutiert wird (Bryk & Raudenbush, 1992). So reflektiert die Schichtzugehörigkeit auf Schülerebene die intellektuellen und

ökonomischen Ressourcen eines Schülers. Die Schichtzugehörigkeit der Schüler könnte auf einer höheren Ebenen, z.B. auf der Ebene der Schule, aggregiert werden. Doch die Bedeutung der aggregierten Schichtzugehörigkeit wäre auf dieser Ebene zweifelhaft, weil Schulen keiner Schicht zugehören. Auf der Ebene von Schulen würden mit der aggregierten Schichtzugehörigkeit eher die Ressourcen der Schule und ihre Einbettung in ein Werte- und Normensystem erfaßt werden.

Ein weiterer Grund für die Berücksichtigung der sozialen Struktur resultiert aus der Tatsache, daß das Zusammenleben von Menschen in einer Gruppe, z.B. von Schülern in einer Klasse, zu einem gemeinsamen Erfahrungshintergrund, gemeinsamen Verhaltensmustern u.ä. führt, was die Beantwortung von typischen psychologischen Meßinstrumenten beeinflusst. Die Annahme unabhängiger Beobachtungen, wie sie für viele statistische Tests getroffen wird, wäre damit verletzt, was eine verzerrte Schätzung von Standardfehlern zur Folge haben kann. Ein Maß für die Verletzung dieser Annahme und damit für die Bedeutung sozialer Ebene stellen die *Intra Class Correlations* (ICC) dar.

Dieser Problematik wird im Ansatz der linearen hierarchischen Modellierung (HLM; Bryk & Raudenbush, 1992) durch Zufallsterme auf den untersuchten Ebenen begegnet. Die Variabilität der Zufallseffekte wird bei der Schätzung der Standardfehler berücksichtigt. Mit anderen Worten werden die Standardfehler unter Berücksichtigung der *Intra Class Correlations* ermittelt bzw. für diese adjustiert.

Im Vordergrund der Fragestellung steht die Entwicklung bzw. Veränderung der Selbstwirksamkeitserwartungen in Abhängigkeit vom Mastery-Klima. Die einfachste Art der Veränderungsmessung stellt die Bildung von Differenzwerten dar. Die Reliabilität von Differenzwerten wird kontrovers diskutiert (vgl. Renkl & Gruber, 1995). Zum einen können sie von der Regression zur Mitte betroffen sein, was in den – oft zu beobachtenden – negativen Korrelationen zwischen Ausgangswert und Veränderung zum Ausdruck kommt. Zum anderen gelten sie vor allem dann als besonders unreliabel, wenn die Meßfehler der wiederholten Messungen korreliert sind (Meßfehlerautokorrelation). Ein wesentlicher Nachteil der Differenzwertbildung ist zudem darin zu sehen, daß lediglich zwei Meßzeitpunkte in ein solches Veränderungsmaß Eingang finden können.

Ein Ansatz, der es erlaubt, mehrere Meßzeitpunkte zu berücksichtigen, stellt die Modellierung *individueller Wachstumskurven* im Ansatz von HLM dar (Bryk & Raudenbush, 1992). Auf der Meßwiederholungs- oder Unit-Ebene wird die abhängige Variable dabei als Funktion der Zeit aufgefaßt. Auf der zweiten Ebene können die Effekte von Personenmerkmalen auf die Veränderung spezifiziert werden und auf der dritten Ebene die Effekte von Merkmalen eines Kollektivs, wie z.B. einer Schulklasse.

Dieser Ansatz weist auch gegenüber dem varianzanalytischen Modell Vorteile auf. Varianzanalysen mit Meßwiederholung stellen letztendlich eine Umsetzung der Differenzwertlogik dar (vgl. Renkl & Gruber, 1995). Bei zwei Meßzeitpunkten entspricht die Interaktion in der Varianzanalyse zwischen Meßwiederholungsfaktor und einem unabhängigen Faktor dem Haupteffekt dieses Faktors für den einfachen Differenzwert. Mit anderen Worten macht es bei zwei Meßzeitpunkten keinen Unterschied für das Ergebnis, ob man eine Varianzanalyse mit Meßwiederholung durchführt oder eine einfache Varianzanalyse über Differenzwerte. Im varianzanalytischen Modell können zudem gegenläufige Entwicklungen für verschiedene Personengruppen (z.B. in Abhängigkeit vom Alter oder einer kontinuierlichen Variable) nur unzulänglich berücksichtigt werden.

Die Modellierung individueller Wachstumskurven weist ihrerseits Nachteile auf. Für diese Abhandlung ist von Bedeutung, daß sich eine wechselseitige Dynamik zwischen den Variablen nicht modellieren läßt. Weitere Nachteile ergeben sich daraus, daß die Anzahl zu schätzender Parameter überproportional zur Anzahl berücksichtigter Variablen wächst, was bereits bei wenigen Variablen große Stichproben erforderlich macht, um zu halbwegs replizierbaren Ergebnissen zu gelangen (vgl. Ditton, 1998). Zudem sind die Möglichkeiten der Modellspezifikation nahezu grenzenlos und damit wenig überschaubar und in ihrer Gesamtheit kaum darstellbar.

Um dieser Problematik zu begegnen, wurde die Anzahl der berücksichtigten Variablen für die HLM-Analysen möglichst gering gehalten, indem z.B. nicht jede einzelne Klimaskala sondern der Mastery-Faktor als Prädiktor eingesetzt wurde. Zudem wurden die Modelle nach den Empfehlungen von Hox (1995) und Engel (1998) spezifiziert. Es kann damit für die Modelle der folgenden Analysen das Prinzip der *Sparsamkeit* bzw. *Einfachheit* geltend gemacht werden.

Die zweite zentrale Fragestellung dieser Abhandlung, inwieweit das Mastery-Klima als Moderator auf Klassenebene die Selbstwirksamkeitsdynamik (Self-efficacy dynamics) beeinflusst (Cross Level Interaction II), wurde im Ansatz längsschnittlicher lineare Strukturgleichungsmodelle untersucht. Auch die dritte Fragestellung nach der Auswirkung von Klima und Selbstwirksamkeitserwartungen auf die Leistung und das Befinden (Mediator-Hypothese) wurde in diesem Ansatz ergründet.

Lineare Strukturgleichungsmodelle (Structural Equations Models, SEM) vereinigen als multivariate Methode mehrere Vorzüge. So können sowohl Meßfehler als auch Meßfehlerautokorrelationen beachtet werden. Fitindizes erlauben die Beurteilung der Modellgüte. Nicht zuletzt kann der Dynamik zwischen Variablen Rechnung getragen werden, was bei der Frage nach der wechselseitigen

Abhängigkeit von spezifischen und allgemeinen Selbstwirksamkeitserwartungen von besonderem Interesse ist.

Der Ansatz wurde für hierarchische Datenstrukturen weiterentwickelt (McDonald & Goldstein, 1989; Muthén, 1989, 1990, 1994), um Kovarianzstrukturen auf Individual- und Aggregatebene unabhängig voneinander untersuchen zu können. Maximum Likelihood Estimates sind jedoch nur unter kaum auffindbaren und herzustellenden Voraussetzungen möglich.

Beide Methoden, HLM und SEM für hierarchische Daten, werden im folgenden in bezug auf die Fragestellung eingehender dargestellt.

5.3.1 Modellierung von Wachstumskurven mit HLM

Mehrebenen-Regressionsmodelle

Mehrebenen-Regressionsmodelle, z.B. *hierarchical linear models* (HLM) von Bryk und Raudenbush (1987, 1992), stellen eine Erweiterung der multiplen Regression dar und erlauben die Analyse hierarchischer Daten. Mit HLM können insgesamt drei Ebenen berücksichtigt werden. Das vollständige Mehrebenen-Regressionsmodell geht von einer hierarchischen Datenstruktur bei einer abhängigen Variable auf der untersten Ebene (Unit-Level) aus. Mehrebenen-Regressionsmodelle lassen sich als hierarchisches System von Regressionsgleichungen verstehen, wobei die Parameter der untersten Ebene durch Gleichungen der nächsthöheren Ebene beschrieben werden.

Zur Bestimmung der Parameter verwendet HLM ein iteratives Verfahren mit Fehlerkorrektur nach dem Maximum-Likelihood-Prinzip (Bryk & Raudenbush, 1992; Ditton, 1998; Hox, 1995). HLM für drei Ebenen (HLM3) nutzt im speziellen einen Full-Maximum-Likelihood-Ansatz (FML), bei dem sowohl die fixen Parameter als auch die Varianzen nach dem Maximum-Likelihood-Prinzip ermittelt werden. Zudem können Modelle mittels eines χ^2 -Differenztests verglichen werden (Hox, 1995). *P*-Werte zur Bestimmung der Parametersignifikanz werden dabei nicht nach der üblichen Wald-Methode (*z*-Test; Wald, 1943) berechnet, sondern auf die *t*-Verteilung bezogen. Als besonderes Merkmal liefert HLM darüber hinaus eine Schätzung der Parameterreliabilität als Anteil wahrer Parametervarianz in jedem Parameter (Bryk & Raudenbush, 1992; Hox, 1995).

Es gibt nur wenige Simulationsstudien, die Aufschluß über die notwendige Stichprobengröße geben (vgl. Ditton, 1998; Hox, 1995). Sie zeigten, daß bei

Mehrebenenmodellen der Anzahl der Aggregateinheiten eine besondere Bedeutung zukommt. Wenige Aggregateinheiten sind nicht geeignet, um erwartungstreue *Varianzschätzungen* zu erzielen, wobei sich *Parameterschätzungen* robuster darstellen. Eine korrekte Schätzung der Varianzanteile war in den Studien, über die Kreft (1996, zitiert nach Ditton, 1998) berichtet, erst bei 300 Aggregateinheiten gegeben. Bei einer Gesamtstichprobe von $N = 800$ konnten fixe Effekte bereits zuverlässig bestimmt werden (Mok, 1996, zitiert nach Ditton, 1998).

Notation

Die wiederholten Merkmalsmessungen bilden im Ansatz von Wachstumskurven die unterste Ebene. Die zweite Ebene ist hier die Ebene der Schüler, auf der dritten Ebene werden die Aggregateinheiten (Schulklassen) untersucht. Es wird die Notation von Bryk und Raudenbush (1987, 1992) für Modelle mit drei Ebenen übernommen.

Auf der untersten Ebene (Unit-Level oder Level 1) wird eine abhängige Variable Y (z.B. Selbstwirksamkeitserwartung) als lineare Funktion von α_{pijk} sowie einer Zufallsgröße (e_{ijk}) beschrieben (Gleichung 5.1):

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk} \alpha_{1ijk} + \pi_{2jk} \alpha_{2ijk} + \dots + \pi_{pjk} \alpha_{pijk} + e_{ijk} \quad (5.1),$$

wobei

Y_{ijk}	das wiederholt gemessene Merkmal zu Meßzeitpunkt i für Schüler j in Klasse k ,
π_{0jk}	die Konstante (Intercept) für Schüler j in Klasse k ,
π_{pjk}	den Koeffizienten (Slope) für Variable p auf Meßwiederholungsebene für Schüler j in Klasse k ,
α_{pijk}	die Ausprägung für Variable p ($1, \dots, P$) auf Meßwiederholungsebene zu Meßzeitpunkt i für Schüler j in Klasse k und
e_{ijk}	das Residuum zu Meßzeitpunkt i für Schüler j in Klasse k darstellt.

Jeder Regressionskoeffizient der untersten Ebene kann auf der nächsthöheren Ebene als fixiert (nicht zufällig variierend), in Teilen zufällig variierend oder als gänzlich zufällig variierend angesehen werden. Entsprechend kann für jeden Koeffizienten ein eigenes Modell spezifiziert werden, welches in seiner allgemeinen Form lautet

$$\pi_{pjk} = \beta_{p0k} + \sum_{q=1}^{Q_p} \beta_{pqk} X_{qjk} + r_{pjk} \quad (5.2)$$

wobei

β_{p0k} die Konstante für Klasse k ,
 β_{pqk} den Koeffizienten (Slope) für Schülervariable q in Klasse k ,
 X_{qjk} die Ausprägung für Schülervariable q für Schüler j in Klasse k und
 r_{pjk} das Residuum für Schüler j in Klasse k darstellt.

Die Koeffizienten β_{pqk} der Schülerebene können auf der Klassenebene durch die Klassenmerkmale W_s (z.B. Klassenklima) vorhergesagt werden nach

$$\beta_{pqk} = \gamma_{pq0} + \sum_{s=1}^{S_{pq}} \gamma_{pqs} W_{sk} + u_{pqk} \quad (5.3)$$

wobei

γ_{pq0} die Konstante,
 γ_{pqs} den Koeffizienten (Slope) für Klassenvariable s ,
 W_{sk} die Ausprägung der Klassenvariable s für Klasse k und
 u_{pqk} das Residuum für Klasse k darstellt.

Das Prinzip besteht also darin, die Varianz auf einer Ebene in fixe und zufällig variierende Anteile zu zerlegen und die nichtzufällige Variation eines Koeffizienten auf der nächsthöheren Ebene durch Merkmale dieser Ebene aufzuklären. Man spricht deshalb auch davon, daß für jeden Koeffizienten der untersten Ebene ein eigenes Modell der nächsthöheren Ebene aufgestellt wird, dessen Koeffizienten wiederum durch Modelle auf der nächsthöheren Ebene beschrieben werden. Durch Einsetzen erhält man eine Regressionsgleichung mit Zufallseffekten auf allen Ebenen und Regressionskoeffizienten der letzten Ebene.

Für die vorliegende Fragestellung ist von Interesse, wie sich die Selbstwirksamkeitserwartungen in Abhängigkeit von der zeitlichen Variation des Mastery-Klimas und in Abhängigkeit vom Klima auf Klassenebene über den Untersuchungszeitraum hinweg entwickelten.

Um die reinen Zeiteffekte abzuschätzen, wurde auf der Meßwiederholungsebene die Dummy-Variable t eingeführt, wobei der erste Meßzeitpunkt mit $t = 0$, der zweite Meßzeitpunkt mit $t = 1$ und der dritte Meßzeitpunkt mit $t = 2$ kodiert wurde. Durch die zusätzliche Berücksichtigung von t^2 (Polynom 2. Grades) konnte eine Wurzelfunktion angenähert werden (z.B. ein anfänglich starker Anstieg, der sich mit der Zeit verlangsamt).

Neben der zeitbedingten Entwicklung galt es den Effekt einer zeitlichen Veränderung des Mastery-Klimas zu ermitteln, denn nach dem Rahmenmodell wird erwartet, daß eine zeitliche Variation des Mastery-Klimas die zeitliche Entwicklung der Selbstwirksamkeitserwartung formt. Diese Hypothese konnte überprüft werden, indem neben den Zeitvariablen das zeitlich variierende Mastery-Klima MC_{2ijk} auf der Unit-Ebene einbezogen wurde mit

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk} t_{ijk} + \pi_{2jk} MC_{2ijk} + \pi_{3jk} t^2_{ijk} + e_{ijk}, \quad (5.4)$$

wobei

Y_{ijk}	die Ausprägung der abhängigen Variable zum Meßzeitpunkt i für Schüler j in Klasse k ,
π_{0jk}	die Konstante für Schüler j in Klasse k ,
π_{1jk}	den linearen Zeiteffekt,
π_{3jk}	den quadratischen Zeiteffekt,
t	die Dummy-Variable zur Zeitkodierung,
π_{2jk}	den Effekt der Veränderung des Mastery-Klimas,
MC_{2ijk}	die Ausprägung des Mastery-Klimas zu Meßzeitpunkt i für Schüler j in Klasse k und
e_{ijk}	das Residuum darstellt.

Die Schülerkonstante π_{0jk} wurde auf Schülerebene in Zusammenhang mit der Geschlechts- und Kohortenzugehörigkeit gebracht, was der Hypothese entspricht, daß die Geschlechts- und Kohortenzugehörigkeit einen generellen Effekt auf die Selbstwirksamkeitserwartungen ausüben, daß etwa Mädchen höhere soziale Kompetenzerwartungen aufweisen.

Der lineare Zeiteffekt π_{1jk} wurde aufgrund der Voranalysen (Abschnitt 6.2) mit der Kohortenzugehörigkeit vorhergesagt. Es konnte so die Annahme überprüft werden, daß die Kohortenzugehörigkeit das Wachstum der jeweils abhängigen Variable Y_{ijk} (Selbstwirksamkeitserwartungen, Optimismus und Notenindex) bestimmt.

Von größerer Bedeutung als die Schülereffekte sind die Klasseneffekte, also die Effekte des Klimas auf die Koeffizienten der Meßwiederholungsebene. Nach dem Rahmenmodell wird erwartet, daß Schüler in Klassen mit ausgeprägtem

Mastery-Klima höhere Selbstwirksamkeitserwartungen aufweisen, was dem Haupteffekt des Mastery-Klimas auf die Konstante π_{0jk} entspricht. Zudem sollte der Effekt einer Veränderung des Mastery-Klimas π_{2jk} abhängig sein von der Ausprägung des Mastery-Klimas auf Klassenebene, was der Cross Level Interaction I entspricht.

Die Umsetzung der Hypothesen in ein hierarchisches Gleichungssystem wird im Ergebnisteil wieder aufgegriffen. Um die Schreibweise zu vereinfachen wird dabei im weiteren zur Beschreibung der Variabilität eines Koeffizienten um seinen Erwartungswert auf der nächsthöheren Ebene der Laufindex ausgelassen (z.B. π_0 statt π_{0jk}). Für die Modellschätzung wurde WHLM 4.01.01 (Bryk, Raudenbush & Congdon, 1996) eingesetzt.

5.3.2 SEM für hierarchische Daten

Stichprobe und Stichprobenmatrix

LISREL-Modelle sind Kovarianzstrukturen: „In principle, all LISREL models are covariance structures, where the variances of Σ as well as the covariances are functions of parameters“ (Jöreskog, 1990, S. 400). Als Stichprobenmatrix wird daher prinzipiell eine Varianz-Kovarianz-Matrix eingesetzt, wenn die Daten auf einer Intervallskala gemessen wurden: „The general rule is that the covariance matrix should be analyzed“ (Jöreskog, 1990, S. 400). Die Verwendung einer *Korrelationsmatrix* zieht indes zahlreiche Probleme nach sich, wie eine verzerrte Schätzung der Λ -Matrix, inkorrekte Fitindizes und verzerrte Standardfehler (Jöreskog, 1990, 1993).

In der vorliegenden Abhandlung wurde von dieser Regel nur einmal bei der Analyse der faktoriellen Struktur der Items abgewichen (Abschnitt 5.2.1), da die Antwortskalen der Items jeweils nur wenige Kategorien umfaßten und daher *nicht* von intervallskalierten Daten auszugehen war. Statt dessen wurde auf *Polychoric Correlations* zurückgegriffen, die unter der Annahme berechnet werden, daß die Variablen in der Population kontinuierlich und intervallskaliert sind (Jöreskog, 1990). In einer Monte Carlo Studie haben Jöreskog und Sörbom (1988) zeigen können, daß mit *Polychoric Correlations* sehr gute Ergebnisse erzielt werden, auch wenn Verteilungsannahmen nicht erfüllt sind.

Die multivariate Normalverteilung stellt eine zentrale Voraussetzung für die sinnvolle Anwendung der Maximum-Likelihood-Fitfunktion dar. Bei Verletzung der Verteilungsannahme kann nach Jöreskog (1990) alternativ der von Browne (1982, 1984) entwickelte GLS-Ansatz verwendet werden. Die Berech-

nung der notwendigen Gewichtsmatrix ist in der Praxis jedoch sehr problematisch und auf wenige Variablen beschränkt, weswegen dieser Ansatz in der vorliegenden Abhandlung keine Anwendung fand.

Diese Abhandlung beruft sich statt dessen auf Simulationsstudien, die belegen, daß bei moderater Verletzung der Verteilungsannahmen und hinreichend großer Stichprobe ($N > 400$) die Verwendung der Maximum-Likelihood-Fitfunktion nicht zu *systematisch* verzerrten Parameterschätzungen führt (Boomsma, 1987). Hox (1993) schlägt für hierarchische Datenstrukturen eine Stichprobe von mindestens $N = 200$ vor und führt zugleich an, daß auf der Aggregatebene deutlich weniger Analyseneinheiten notwendig sind, da die Kovarianzen auf Gruppenebene aufgrund der Aggregation wesentlich präziser ausfallen als auf Individualebene. Jöreskog und Sörbom (1989, S. 191) fassen die Ergebnisse zahlreicher Monte Carlo Studien zusammen: „If such variables are approximately normally distributed [...] the use of the ML [...] estimates and their associated standard errors and χ^2 goodness-of-fit measures may be justified.“

Grundgedanke von Mehrebenenanalysen im SEM-Ansatz

Der SEM-Ansatz wurde zur Modellierung von Kovarianzstrukturen entwickelt. Üblicherweise wird daher aus den Daten eine Varianz-Kovarianz-Matrix S_T erzeugt und für diese ein Modell geschätzt.

Liegt eine hierarchische Datenstruktur vor, ist jedoch bei diesem Vorgehen unklar, wie die Parameter zu beurteilen sind, da die Zusammenhänge verschiedener Ebene konfundiert werden. Dies sei an einem Beispiel verdeutlicht, bei dem die Korrelation zwischen Schulleistung und Intelligenz ermittelt werden soll. Die Schulleistung ist nicht nur abhängig von der Fähigkeit eines Schülers, sondern auch vom Bewertungsmaßstab des Klassenlehrers. Ein Lehrer bewertet streng, der andere äußerst fair. Denkbar ist zudem, daß der Klassenlehrer seinen Bewertungsmaßstab an das Intelligenzniveau der Klasse anpaßt. Die einfache Korrelation zwischen Schulleistung und Intelligenz sagt aus diesem Grunde wenig aus. Man kann nicht beurteilen, ob sie den Zusammenhang zwischen Leistung und Intelligenz oder zwischen dem Bewertungsmaßstab des Lehrers und dem Intelligenzniveau der Klasse widerspiegelt.

Um in einem solchen Fall Abhilfe zu schaffen, können die Schulleistung und die Intelligenz von den Klasseneffekten bereinigt werden, in dem man sie am Klassenmittelwert zentriert (vgl. Härnqvist, 1978). So wird z.B. von der Schulleistung eines Schülers j der Klasse k der Klassenmittelwert M_k der Klasse k subtrahiert. Die Varianz-Kovarianz-Matrix dieser bereinigten Individualwerte wird als Varianz-Kovarianz-Matrix *Within* (S_W) bezeichnet. Sie repräsentiert die Zusammenhänge auf Schülerebene. Analog kann eine Varianz-Kovarianz-Ma-

trix *Between* (S_B) für die Klassenebene aus den Klassenmittelwerten erzeugt werden. Diese Matrix repräsentiert die Zusammenhänge auf Klassenebene. Beide Matrizen sind orthogonal zueinander: Multipliziert man sie, ergibt sich eine Null-Matrix³.

Der Grundgedanke von Mehrebenenanalysen im SEM-Ansatz besteht nun darin, daß man die Varianz-Kovarianz-Matrizen *Within* S_W und *Between* S_B erzeugt, und ein Modell mit beiden Matrizen schätzt. Ein Problem ergibt sich nun wiederum aus der Tatsache, daß beide Matrizen keine erwartungstreuen Schätzer für Σ_W und Σ_B sind (vgl. Hox, 1993, 1995; Schnabel, 1993). So stellt S_B einen Schätzer für die gewichtete Summe von Σ_W und Σ_B dar. Muthén (1990) hat ein Verfahren entwickelt, um dennoch Maximum Likelihood Estimates zu erzielen. Diese Vorgehensweise sieht – vereinfacht ausgedrückt – vor, daß für S_B sowohl das Modell für Σ_W als auch das Modell für Σ_B geschätzt wird, was leider zu einem deutlichen Anstieg an Modellparametern führt und zu der damit verbundenden Problematik der Identifizierbarkeit. Zudem liefert auch diese Vorgehensweise nur bei einer hinreichend großen Stichprobe auf Aggregatebene sowie bei einem balancierten Design (gleiche Schüleranzahl pro Klasse) Maximum Likelihood Estimates.

Für die vorliegende Arbeit wurde deshalb ein Kompromiß gefunden. Auf Schülerebene (innerhalb der Klassen) wurde die $N-G$ (wobei N = Schülerstichprobe und G = Klassenstichprobe) gewichtete Matrix S_{PW} berechnet (P für pooled). Muthén (1990) konnte zeigen, daß S_{PW} ein erwartungstreuer Schätzer für Σ_W ist. Auf Schülerebene können damit effektive Parameterschätzungen erzielt werden, zumal die Schülerstichprobe als hinreichend groß angesehen werden kann. Auf Klassenebene war die Stichprobe eher klein (Abschnitt 5.1.3), zudem variierte die Schülerstärke der Klassen erheblich. Modelle für die Klassenebene wurden daher lediglich mit S_B geschätzt. Von den Parameterschätzungen und Standardfehlern auf Klassenebene ist somit nicht zu erwarten, daß sie besonders effektiv oder genau sind. Sie sind vielmehr mit äußerster Vorsicht zu interpretieren.

Alle Strukturgleichungsmodelle wurden mit LISREL 8.12a (Jöreskog, 1990, 1993; Jöreskog & Sörbom, 1989, 1993) geschätzt. LISREL berechnet zahlreiche globale Fitindizes, mit deren Hilfe ein Modell bewertet werden kann. Der χ^2 -Test ist dabei an eine Reihe von Voraussetzungen gebunden. Zudem benachteiligt er sparsame Modelle. Eine große Stichprobe führt dazu, daß Modelle bereits aufgrund kleiner Abweichungen verworfen werden müßten, weil der Test die Hypothese überprüft, daß ein Modell die Realität *exakt* repräsentiert. Bereits aus diesem Grunde wäre daher die Anwendung des χ^2 -Tests

3 Aus dieser Tatsache folgt, daß die Frage nach einem Einfluß der aggregierten Variablen auf die zentrierten Schülervariablen unsinnig ist.

verfehlt (vgl. Jöreskog, 1993). Als Alternative wird daher der χ^2 -Differenztest eingesetzt. Dieser Test überprüft, ob eine Modellerweiterung zu einer verbesserten Anpassungsgüte führt. Darüber hinaus werden neben den klassischen Indizes, wie z.B. dem *GFI* (Jöreskog & Sörbom, 1989, 1993) und dem *NFI* (Bentler & Bonett, 1980), auch Kennwerte herangezogen, die die Sparsamkeit von Modellen berücksichtigen, wie etwa der *Root Mean Square Error of Approximation (RMSEA)* von Browne und Cudeck (1993). Prinzipiell gilt, daß mit globalen Fitindizes immer nur sehr schlechte Modelle zurückgewiesen werden können. Es ist kaum möglich bei zwei guten Modellen zu entscheiden, welches das bessere ist (Jöreskog, 1993; Tanaka, 1993).

5.3.3 Zusammenfassung

Zur Untersuchung der zentralen Fragestellung dieser Abhandlung wurden Mehrebenenanalysen mit HLM und LISREL durchgeführt. HLM wurde herangezogen, um die Hypothese zu überprüfen, daß die Entwicklung der Selbstwirksamkeitserwartungen abhängig von der zeitlichen Variation des Mastery-Klimas ist. Darüber hinaus wurde in diesem Ansatz überprüft, ob das Mastery-Klima auf Klassenebene einen Haupteffekt auf die Selbstwirksamkeitserwartungen ausübt und ob der Effekt einer zeitlichen Veränderung des Mastery-Klimas auf der Meßwiederholungsebene abhängig ist von der Ausprägung des Mastery-Klimas auf Klassenebene (Cross Level Interaction I). Im SEM-Ansatz wurde mit LISREL überprüft, ob die Selbstwirksamkeitsdynamik vom Mastery-Klima auf Klassenebene (Cross Level Interaction II) abhängig ist. Darüber hinaus wurde in diesem Ansatz der Frage nachgegangen, inwieweit Klima und Selbstwirksamkeitserwartungen die Leistung und das Befinden der Schüler bestimmen. Diese Zusammenhänge wurden dabei sowohl auf Schüler- als auch auf Klassenebene untersucht.

Beide Ansätze stellen hohe Ansprüche an die Verteilung und die Qualität der Daten sowie an den Umfang der Stichprobe, die nur zum Teil erfüllt werden konnten. So ist die Stichprobengröße auf Schülerebene sicherlich als hinreichend anzusehen. Auf Klassenebene reduzierte sich die Stichprobe durch die Aggregation der Daten jedoch auf wenige Einheiten (Klassen), was im SEM-Ansatz zu sehr großen Standardfehlern führt. Die Verteilung der Variablen wurde visuell und statistisch überprüft (Schwarzer & Jerusalem, 1999). Auch wenn nicht von einer multivariaten Normalverteilung auszugehen ist, sind die meisten Variablen annähernd normal verteilt. Um robuste Parameterschätzungen zu erhalten, wurden alle Modelle zudem möglichst sparsam formuliert. Komplexere Modelle wurden nur dann akzeptiert, wenn sie die restriktiveren Modelle im Modellfit deutlich übertrafen.