

Methodische Probleme und Möglichkeiten schulbezogener Längsschnittforschung

Wolfgang Schneider

1 Gründe für schulbezogene Längsschnittforschung

Für die meisten Bereiche der sozialwissenschaftlichen Forschung gilt nach wie vor, daß zur Überprüfung theoretischer Annahmen sogenannte Querschnittsuntersuchungen durchgeführt werden, die dadurch charakterisiert sind, daß empirische Erhebungen nur zu einem einzigen Testzeitpunkt stattfinden. Längsschnittuntersuchungen stellen demnach auch in der schulbezogenen Forschung eher die Ausnahme dar (vgl. Keeves 1988). Dies heißt jedoch nicht, daß sie völlig fehlen: in einer von Verdonik und Sherrod (1984) zusammengestellten Dokumentation (primär) amerikanischer Längsschnittstudien waren von insgesamt 116 erfaßten Studien etwa 20 Projekte im Bereich schulbezogener Forschung angesiedelt. Eine kürzlich für den europäischen Bereich vorgenommene Dokumentation sozialwissenschaftlicher und medizinischer Längsschnittstudien (Schneider u. Edelstein 1990) enthielt knapp 100 (von insgesamt annähernd 500 Längsschnittuntersuchungen), die dem Bereich pädagogisch-psychologischer Forschung zuzurechnen waren. Das Interesse an schulbezogenen Längsschnittstudien scheint demnach in jüngerer Zeit eher zuzunehmen.

Für diesen Trend lassen sich m.E. mehrere Gründe aufführen. Die Tatsache, daß die meisten wissenschaftlichen Erkenntnisse auf Studien basieren, in denen *Entwicklungsunterschiede* zwischen verschiedenen Altersgruppen und nicht *Entwicklungsveränderungen* von Schülern einer Alterskohorte über die Zeit im Mittelpunkt standen, limitiert unser Wissen um Gesetzmäßigkeiten des Entwicklungsverlaufs. Ein einmaliger „Schnappschuß“, wie in Querschnittsuntersuchungen üblich, gibt uns möglicherweise einen äußerst verzerrten Einblick in zugrundeliegende Wirkungszusammenhänge. Zuverlässige Aussagen über den Typus der „Entwicklungsfunktion“ im Sinne von Wohlwill (1980) können erst über sorgfältig geplante Längsschnittstudien erreicht werden. Als illustratives Beispiel mag die Forschung zur Entwicklung des schulischen Selbstkonzepts dienen. Während querschnittlich angelegte Untersuchungen für den Bereich der Grundschulzeit einen über die Schuljahre hinweg annähernd linearen Trend von einer Phase der krassen Selbstüberschätzung bis hin zum realistischen Selbstbild suggerieren, deuten Ergebnisse von Längsschnittstudien eher darauf hin, daß sich die Entwicklung sprunghaft und demnach nicht-linear vollzieht (vgl. Helmke in diesem Band).

Längsschnittstudien sind weiterhin aber beispielsweise auch dann unentbehrlich, wenn es um die Prognose von schulischen Leistungen geht. Hier interessiert weniger das absolute Ausmaß von Entwicklungsveränderungen in einem bestimmten schulischen Bereich über die Zeit als vielmehr die Frage, inwieweit diese Veränderungen durch zeitlich vorher erhobene Variablen (etwa die Intelligenz der Schüler) zuverlässig vorhergesagt werden können.

2 Definitionsprobleme

Sind auf der einen Seite die Vorzüge von Längsschnittstudien gegenüber Querschnitts-Designs relativ einfach zu benennen, fällt es andererseits nicht unbedingt leicht, die vielfältigen Probleme der Längsschnittforschung angemessen zu diskutieren. Die Schwierigkeiten beginnen schon beim Versuch, die konstituierenden Elemente von Längsschnittstudien präzise zu definieren. Der Begriff „längsschnittlich“ bezieht sich nicht auf einen einzigen methodischen Ansatz, sondern umfaßt die unterschiedlichsten methodischen Zugangsweisen. Das Spektrum reicht dabei von im Zeitreihen-Design durchgeführten Einzelfallstudien bis zu breit angelegten Längsschnittstudien, die mehrere tausend Personen und fast ebenso viele Merkmale umfassen. Der einzige gemeinsame Nenner dieser Studien besteht in der Variation von Zeit und dem Umstand, daß für eine Person mehrere Messungen vorliegen müssen (vgl. *Baltes u. Nesselroade 1979, Janson 1981*). Das Ausmaß dieser Definitionsprobleme wurde kürzlich in einer Diskussion deutlich, die zwischen Mitgliedern eines Komitees der European Science Foundation im Hinblick auf die Frage stattfand, welche Auswahlkriterien für die schon oben erwähnte Dokumentation von europäischen Längsschnittstudien (*Schneider u. Edelstein 1990*) festgelegt werden sollten. Obwohl alle Beteiligten über große Erfahrung mit Längsschnittstudien verfügten, gingen die Meinungen über das Definitionskriterium weit auseinander. Der letztendlich resultierende Kompromißvorschlag (mindestens 3 Meßzeitpunkte, mindestens 30 Versuchspersonen und eine Untersuchungszeit von mindestens 3 Jahren) wirkte nicht nur einigermaßen arbiträr, sondern mußte in der Folge auch durch zahlreiche Ausnahmeregelungen (etwa im Bereich der Kleinkindforschung) modifiziert werden.

3 Konzeptuelle Unterschiede zwischen Längsschnittstudien

Unabhängig von den Definitionsproblemen gilt es zu beachten, daß unterschiedliche Zielvorstellungen für Längsschnittstudien unterscheidbar sind. *Baltes und Nesselroade (1979)* zufolge gibt es drei Zielvorstellungen, die sich auf die *Beschreibung* von Entwicklungsvorgängen beziehen:

- a) die Identifizierung intraindividuelle Veränderungen;
- b) die Identifizierung interindividueller Unterschiede in individuellen Veränderungen, und
- c) die Identifizierung von Interrelationen zwischen unterschiedlichen Verhaltensklassen bzw. -kategorien in der Entwicklung.

Zwei weitere Ziele betreffen die *Erklärung* von Entwicklungsvorgängen:

- d) die Analyse von Ursachen intraindividuelle Veränderungen, und
- e) die Analyse der Ursachen für interindividuelle Unterschiede in den intraindividuellen Veränderungsdaten.

Diese Auflistung von Zielvorstellungen läßt sich m.E. noch durch eine weitere Kategorie ergänzen, die sich auf die Vorhersage individueller Unterschiede in einer bestimmten Domäne durch individuelle Unterschiede in einer anderen Domäne bezieht (vgl. *Schneider im Druck*). Ein Beispiel dafür stellen Längsschnittstudien dar, die den Einfluß der Sprachentwicklung im Vorschul- und Kindergartenalter (Domäne A) auf den Schriftspracherwerb in der Schule (Domäne B) erfassen. Diese Studien sind einerseits längsschnittlich konzipiert, registrieren andererseits jedoch nicht notwendigerweise

intraindividuelle Veränderungen in einer spezifischen Variablen über die Zeit hinweg.

Es lassen sich zwei verschiedene methodische Zugangsweisen („basic types of longitudinal inquiry“, vgl. *Appelbaum* u. *McCall* 1983) verwenden, um die aufgeführten Zielvorstellungen von Längsschnittstudien zu realisieren. Der eine betrifft die Analyse der „Entwicklungsfunktion“ sensu Wohlwill (1980), also die Analyse der absoluten Veränderung von Merkmalen über die Zeit hinweg. Typische Beispiele betreffen etwa Wachstumskurven für Körpergröße bzw. -gewicht oder Entwicklungsfunktionen für Kraft und Geschwindigkeit in der motorischen Entwicklung. Für die schulbezogene Längsschnittforschung lassen sich als sinnvolle Beispiele Wachstumskurven der Intelligenz oder Veränderungsverläufe im schulischen Selbstkonzept anführen.

Der zweite, für die schulbezogene Längsschnittforschung ungleich wichtigere methodische Zugang besteht in der Analyse individueller Unterschiede, wobei primär interessiert, ob bestimmte Personen zu unterschiedlichen Zeitpunkten in etwa die gleiche Rangposition innerhalb ihrer Gruppe beibehalten. Interessiert bei der Analyse von Entwicklungsfunktionen insbesondere die Frage der Kontinuität bzw. Diskontinuität von Entwicklungsvorgängen, so steht hier das Problem der Stabilität bzw. Instabilität von individuellen Unterschieden im Vordergrund der Betrachtung. Beide Analyseprozeduren mögen in der Empirie zu ähnlichen Ergebnissen führen, sind jedoch konzeptuell voneinander unabhängig. So ist es theoretisch beispielsweise möglich, daß wir bei einem bestimmten Untersuchungsmerkmal (etwa der Gedächtniskapazität von Kindergartenkindern) längsschnittlich einen monotonen, linearen Zuwachs in den Durchschnittswerten registrieren, auf der anderen Seite jedoch feststellen müssen, daß die individuellen Unterschiede in diesem Merkmal über die gleiche Zeitspanne hinweg höchst instabil sind. Es wird häufig übersehen, daß die Frage nach der Entwicklungsfunktion und die nach der Stabilität individueller Unterschiede zwei völlig unterschiedliche Aspekte eines gemeinsamen Problems ausmachen.

Auf die methodischen Implikationen wird in der Folge genauer eingegangen. Zuvor sollen jedoch ganz kurz einige Designprobleme erörtert werden, die bei schulbezogenen Längsschnittstudien häufig auftreten.

4 Allgemeine Designprobleme schulbezogener Längsschnittforschung

Welche Probleme längsschnittlicher Untersuchungsdesigns im Einzelfall besonders bedeutsam werden, hängt sicherlich von der konzeptuellen Ausrichtung, dem inhaltlichen Schwerpunkt und der geplanten Dauer einer Längsschnittstudie ab. Es geht im folgenden deshalb nur um einige wenige allgemeine Punkte, die in der einschlägigen Literatur häufiger diskutiert werden.

Ein sicherlich sehr ernstzunehmendes Problem betrifft dabei die Frage, wie Längsschnittstudien optimal „getaktet“ werden können, d.h. nach welchen Kriterien die zeitliche Anordnung von Meßzeitpunkten erfolgen soll. Wünschbar wäre beispielsweise eine Anordnung, die in Abhängigkeit vom theoretischen Wissen eine Häufung von Meßzeitpunkten in „kritischen Entwicklungsphasen“ vorsieht, in denen sich besonders dramatische Veränderungen abzeichnen sollten. Der von *Siegler* und *Jenkins* (1989) verwendete *mikrogenetische* Ansatz kommt dieser Vorstellung sehr nahe. In einer Längsschnittstudie mit 10 Vorschulkindern wurden über eine Periode von etwa drei Monaten die Vorgehensweisen der Kinder bei der Addition von Zahlen genauestens beobachtet, um die Phase des qualitativen Übergangs zu besseren Problemlösestrategien exakt zu registrieren. Über den gewählten Ansatz ließen sich die

Bedingungen für erfolgreiche Strategiewechsel in der Tat recht erfolgreich bestimmen. Es leuchtet ein, daß ein solcher Ansatz bei größeren Stichproben seine organisatorischen Grenzen hat und auch nur dann lohnend erscheint, wenn unser theoretisches Wissen eine einigermaßen exakte zeitliche Vorhersage solcher „kritischer Phasen“ gewährleistet. In vielen großangelegten schulbezogenen Längsschnittstudien hat es den Anschein, daß praktische Zwänge (z. B. Organisations- und Auswertungsaufwand, Anpassung an Ferienordnungen und Lehrerwünsche) die Planung von Meßzeitpunkten im Vergleich zu theoretischen Gesichtspunkten stärker bestimmen. Dies zeigt etwa die Organisationsstruktur der von *Schneider* u. *Edelstein* (1990) zusammengestellten Längsschnittstudien, deren Erhebungen in der Regel in gleichen Zeitabständen (meist jährliche Intervalle) durchgeführt wurden, ohne daß dafür triftige theoretische Gründe geltend gemacht werden könnten. Ein methodischer Vorteil einer solchen Erhebungsstrategie ist dann gegeben, wenn Auswertungsprozeduren wie etwa Kausalmodelle mit latenten Variablen (siehe unten) verwendet werden, deren Interpretation dann erschwert wird, wenn ungleiche Zeitintervalle zwischen den einzelnen Erhebungen vorliegen (zur Problematik und evtl. Lösungsmöglichkeiten vgl. *Gollob* u. *Reichardt* 1987; *Rudinger*, *Andres*, *Rietz* u. *Schneider* 1989).

Grundfragen der Stichprobenstruktur und -größe sind ebenfalls eng mit Ansprüchen an die Datenauswertung verknüpft. Ist man daran interessiert, multivariate Auswertungsprozeduren wie etwa Strukturgleichungsmodelle einzusetzen und dabei mehrere Analyse-Ebenen (Schulen, Schulklassen und Schüler) zu beachten, sind große Stichproben unerlässlich. Plant man weiterhin eine Untersuchungszeit von mehreren Jahren ein, so sollte die Stichprobe umfangreich genug sein, um Verluste an Klassen und Schülern, die etwa durch Lehrerwechsel bedingt sein können, ohne negative Konsequenzen für die Datenauswertung aufzufangen. Da in den meisten schulbezogenen Längsschnittuntersuchungen die großen Probandenzahlen mit vergleichsweise geringen personellen und materiellen Projektmöglichkeiten kovariieren, sind die Möglichkeiten der „Stichprobenpflege“ äußerst beschränkt. Es ist von daher ratsam, bei Untersuchungen im Schulbereich von vornherein mit höheren „drop-out“-Quoten zu kalkulieren.

5 Methodische Probleme von Längsschnittstudien

Die Art der methodischen Probleme von Längsschnittstudien hängt nicht zuletzt davon ab, welche der oben aufgeführten sechs Zielvorstellungen angegangen werden soll. Es spielt demnach eine wichtige Rolle, ob Aspekte der Kontinuität/Diskontinuität oder solche der Stabilität/Instabilität von Veränderungen im Vordergrund stehen. In vielen Fällen weisen Studien, die sich mit der Analyse von Entwicklungsfunktionen, also mit intraindividuellen Veränderungen in bestimmten Merkmalen über die Zeit befassen, methodisch vergleichsweise wenig Problem auf; dies insbesondere dann, wenn sie eher mikroanalytisch angelegt sind, innerhalb relativ kurzer Zeiteinheiten durchgeführt werden und sich in den einzelnen Untersuchungsbereichen immer wieder auf die gleichen Meßinstrumente stützen können. Wie noch zu zeigen sein wird, lassen sich hier Entwicklungsveränderungen vergleichsweise problemlos erfassen.

Ein gravierendes Problem besteht allerdings darin, daß nur in wenigen Fällen identische Meßinstrumente über lange Zeiträume hinweg verwendet werden können. Selbst wenn dies möglich sein sollte, stellt sich meist die Frage, ob über das gleiche Meßinstrument in unterschiedlichen Altersstufen wirklich Gleichartiges erfaßt wird (vgl. *Magnusson* 1981). Meist müssen für unterschiedliche Altersstufen parallele Verfahren verwendet werden, was demnach die Analyse *intraindividuell*er Veränderungen in den betref-

fenden Merkmalen unmöglich macht. Dieses fundamentale Problem läßt sich auch dann nicht umgehen, wenn – wie etwa von *Goldstein* (1979) vorgeschlagen – der Versuch gemacht wird, für unterschiedliche Instrumente über verschiedene Werte-Transformationen eine gemeinsame Skala zu konstruieren. Über diese Maßnahme kann man allenfalls *relative* Veränderungen erfassen, d.h. Vergleiche für Subgruppen einer bestimmten Population vornehmen. Die Analyse absoluter Merkmalsveränderungen läßt sich so nicht realisieren.

Dies impliziert, daß in den meisten Längsschnittstudien, in denen schulische Entwicklungsverläufe über eine Reihe von Jahren hinweg beschrieben und erklärt werden sollen, der Schwerpunkt in der Analyse individueller Unterschiede, also in der Untersuchung der Frage liegen muß, ob Unterschiede zwischen Schülern zeitstabil oder instabil sind. Da in den wenigsten Fällen gleiche Instrumente für unterschiedliche Altersgruppen verfügbar sind, ist die Möglichkeit zur Analyse des Entwicklungsverlaufs von Merkmalen sensu *Wohlwill* oftmals beschränkt.

Wenn wir uns nun den methodischen Problemen von Längsschnittstudien zuwenden, sollen zunächst dennoch kurz einige Probleme diskutiert werden, die im Zusammenhang mit der Erfassung *intraindividuelle* Veränderungen hervorgehoben worden sind (vgl. für ausführliche Darstellungen *Rogosa* 1988, *Schneider* 1989a). Im zweiten Schritt wird dann ausführlich auf methodische Probleme eingegangen, die im Zusammenhang mit der längsschnittlichen Analyse interindividueller Unterschiede stehen und demnach die schulbezogene Längsschnittforschung zentraler betreffen.

5.1 Probleme der Veränderungsmessung: Wie ernst sind sie wirklich?

Debatten über Schwierigkeiten der Erfassungen von Veränderungen in psychologischen Merkmalen sind nun schon seit mehreren Jahrzehnten geführt worden. Insbesondere die klassische Arbeit von *Cronbach* und *Furby* (1970) hat zur Überzeugung vieler Forscher beigetragen, daß individuelle Veränderungswerte generell unreliabel, irreführend und auch unfair sind und nach Möglichkeit umgangen werden sollten. Diese Auffassung wird auch noch in neueren Überblicksarbeiten (z. B. *Keeves* 1988) weitergegeben. Methodologische Überprüfungen dieser Annahme führen jedoch zum Ergebnis, daß es sich hier um einen Mythos handelt, dem es kritisch zu begegnen gilt (vgl. *Rogosa* 1988). Sie haben ergeben, daß Differenzwerte weder notorisch unreliabel sind noch durch Meßartefakte (Regression zur Mitte) übermäßig verzerrt werden. So ließ sich in den Untersuchungen von *Rogosa* und Mitarbeitern (1982) zeigen, daß Differenzwerte nur dann niedrige Reliabilitäten aufweisen, wenn die individuellen Wachstumskurven für alle Personen einer gegebenen Stichprobe relativ ähnlich sind. Reliabilität steht in diesem Fall für die Genauigkeit, mit der sich Probanden aufgrund ihrer Differenzwerte auf der Wachstumskurve anordnen lassen, unabhängig davon, ob die Schätzer für die Wachstumskurve präzise oder unpräzise sind. Daraus folgt als wichtige Implikation, daß niedrige Reliabilitäten keinesfalls mangelnde Präzision des Schätzverfahrens indizieren und damit eine inhaltlich sinnvolle Erfassung intraindividuelle Veränderungen absolut nicht ausschließen. Wenn für eine bestimmte Stichprobe nachgewiesen werden kann, daß sich ihre Merkmalswerte in etwa der gleichen Weise verändern, gibt es eben keine Möglichkeit, individuelle Unterschiede in den Veränderungsdaten zuverlässig zu bestimmen. Liegen andererseits jedoch deutliche interindividuelle Unterschiede in den Veränderungsdaten vor, kommt dies notwendigerweise auch in hohen Reliabilitätskennwerten zum Ausdruck.

In ähnlicher Weise ist auch die Bedeutung sog. „Regression-zur-Mitte“-Effekte überschätzt worden (vgl. *Nesselroade*, *Stigler* u. *Baltes* 1980, *Rogosa* 1988). Darunter wird

üblicherweise das Phänomen verstanden, daß gerade bei individuellen Meßwerten, die zum ersten Erhebungszeitpunkt vergleichsweise weit vom Stichprobenmittelwert entfernt sind, eine Tendenz besteht, bei einer Meßwiederholung weniger extrem auszufallen, also näher beim Gruppenmittelwert zu liegen. *Nesselroade* und *Mitarbeiter* (1980) haben empirisch belegen können, daß die Bedeutung von „Regression-zur-Mitte“-Effekten keinesfalls hoch zu veranschlagen ist, und daß sie sich gerade im Fall von mehr als zwei Erhebungszeitpunkten gut kontrollieren lassen.

Faßt man die Ergebnisse neuerer methodologischer Arbeiten zum Problem intraindividuellere Veränderungen zusammen, so zeigt sich eine klare Abkehr von den Positionen früherer Jahre. Die für Längsschnittforscher wichtigste Botschaft besteht darin, daß es wieder gute Gründe für die Verwendung des einfachen Differenzwertes gibt, daß demgegenüber Veränderungs-Residualwerte eher skeptisch bewertet werden, und daß die durch Regression zur Mitte verursachten Verzerrungen der Ergebnisse insbesondere dann vermieden werden können, wenn mehr als zwei Meßzeitpunkte vorliegen.

5.1.1 Neuere Ansätze zur Modellierung längsschnittlicher Veränderungen

Die Analyse intraindividuellere Veränderungen ist in neuerer Zeit dadurch erleichtert worden, daß komplexe statistische Methoden zur Analyse von Wachstumskurvenmodellen entwickelt und in Computerprogramme umgesetzt worden sind. Dabei dürfte das auf der Bayes-Statistik basierende Modell individuellen Wachstums von *Bryk* und *Raudenbush* (1987) inzwischen am weitesten gediehen sein. Es eignet sich für die Analyse von Längsschnittdaten, die auf mindestens drei Meßzeitpunkten basieren und annähernd multivariat normalverteilt sind. Die Logik dieses Ansatzes, der in der Literatur als „hierarchical linear modeling“ (HLM) bezeichnet wird, soll im folgenden kurz zusammengefaßt werden.

In einem ersten Schritt werden für jede Versuchsperson individuelle Wachstumsraten für ein gegebenes Merkmal über die Methode der kleinsten Quadrate geschätzt („within-subject“ model). Der nächste Schritt besteht darin, über ein gewichtetes Kleinstquadrat-Schätzverfahren das „between-subject“-Modell zu bestimmen: es geht hier darum, die beobachtete Varianz in den individuellen Wachstumskurven durch die Einführung geeigneter Hintergrundvariablen (z. B. Geschlecht oder soziale Schichtzugehörigkeit) zu reduzieren und damit zu erklären. In diesem Auswertungsschritt wird ein zweiter Schätzwert für die individuellen Wachstumsparameter generiert, der allein auf den Hintergrundvariablen basiert. Der letzte Verfahrensschritt besteht darin, daß

Tabelle 1 Wesentliche Vorzüge des HLM-Ansatzes (nach *Bryk* u. *Raudenbush* 1987)

- 1) Beschreibung der Struktur von mittleren (durchschnittlichen) Wachstumskurven
- 2) Schätzung des Ausmaßes und der Art von individuellen Abweichungen von der durchschnittlichen Wachstumskurve
- 3) Schätzung der Reliabilität von Maßen zur Erfassung des Eingangsstatus und von individuellen Veränderungen
- 4) Schätzung der Korrelation zwischen dem Eingangsstatus von Personen und den Wachstumsraten
- 5) Erfassung von Korrelaten für Eingangs- wie auch Veränderungswerte
- 6) Erfassung der Angemessenheit von „between-subject“-Modellen (Schätzung der Reduktion in unerklärter Parametervarianz)
- 7) Vorhersage zukünftiger individueller Wachstumsraten

über empirische Bayes-Verfahren ein zusammengesetzter Schätzwert für die individuelle Wachstumskurven ermittelt wird, in dem die in den beiden ersten Schritten gewonnenen Schätzer optimal gewichtet sind.

Zusammengefaßt liegen die wesentlichen Vorzüge dieses Modellansatzes darin, daß die Möglichkeit gegeben ist, die Struktur von Wachstumsraten in einer Stichprobe für mehrere Merkmale zu ermitteln, die Reliabilität von Eingangswerten und Veränderungsdaten zu bestimmen, und Hypothesen über die Effekte theoretisch relevanter Hintergrundvariablen auf die Veränderung von Wachstumsparametern zu überprüfen.

Das HLM-Verfahren ist multifunktional: es läßt sich nicht nur für die Schätzung von Wachstumskurven, sondern auch für statistische Meta-Analysen sowie für die Schätzung von Mehrebenen-Modellen verwenden. Letztere sind gerade für die schulbezogene (Längsschnitt-)Forschung relevant und werden noch genauer erörtert.

5.2 Methodische Probleme bei der längsschnittlichen Analyse von individuellen Unterschieden

Wie schon oben angedeutet, fokussieren Längsschnittstudien, die sich mit individuellen Unterschieden befassen, auf Fragen der Stabilität von Merkmalen bzw. auf Fragen der Vorhersagbarkeit. Dies trifft sicherlich auf die Majorität schulbezogener Längsschnittstudien zu. Längsschnittliche Datenauswertungen basieren dabei in der Regel auf Korrelations- bzw. Kovarianzmatrizen, über die sich die Konsistenz individueller Unterschiede über die Zeit hinweg messen läßt. Es sei an dieser Stelle explizit herausgestellt, daß hohe Retest-Korrelationen alleine noch keine Aussage darüber zulassen, ob zu unterschiedlichen Zeitpunkten jeweils das Gleiche gemessen wurde. Theoretisch ist es jedenfalls denkbar, daß die Rangordnung zwischen Individuen (im Hinblick auf ein gemeinsames Merkmal) auch dann gleichbleibt, wenn das Merkmal im Lauf der Zeit seine Bedeutung verändert. Die Frage, ob ein spezifisches Merkmal im Verlauf der Zeit seine psychologische Bedeutung verändert, läßt sich nur über zusätzliche Validierungsprozeduren (etwa über die zusätzliche Erfassung merkmalskorrelierter Referenzvariablen) abschätzen.

Neben diesem häufig anzutreffenden Mißverständnis findet man öfters auch irrige Meinungen über die Möglichkeiten statistischer Verfahren, z. B. daß sich aus Regressionsanalysen irgendwelche Informationen über Veränderungen in Variablen ableiten lassen. Wir wollen deshalb im folgenden versuchen, die Probleme und Möglichkeiten derjenigen statistischen Prozeduren kritisch zu beleuchten, die bei Längsschnittstudien zur Analyse individueller Unterschiede vergleichsweise häufig eingesetzt wurden. Die grundsätzliche Frage wird hier immer sein, inwieweit diese statistischen Werkzeuge den großen theoretischen Ansprüchen und Erwartungen gerecht werden können, die viele breitangelegte Längsschnittstudien begleiten. Alle im folgenden diskutierten statistischen Prozeduren basieren auf dem klassischen Modell der Korrelations- und Regressionsanalyse und unterscheiden sich im wesentlichen durch die Komplexität der Modellannahmen. Es wird zunächst auf vergleichsweise einfache und gebräuchliche Ansätze eingegangen, danach die Problematik neuerer Modellentwicklungen diskutiert.

5.2.1 Probleme multipler Regressionsanalysen

Die multiple Regressionsanalyse stellt die Standardprozedur für alle Fragestellungen dar, die sich auf die Analyse der Ursachen für interindividuelle Unterschiede in intrain-

dividuelle Veränderungsdaten bzw. auf die Vorhersage individueller Unterschiede in einer bestimmten Domäne durch individuelle Unterschiede in einer anderen Domäne (s.o.) beziehen. Sie ist demnach prinzipiell für schulbezogene Längsschnittstudien geeignet. Im klassischen Modell der Regressionsanalyse wird der Wert einer abhängigen Variablen (dem Kriterium) durch die gewichtete Summe von N unabhängigen Variablen (den Prädiktoren) so zu schätzen versucht, daß die Summe der quadrierten Residuen (Differenzen zwischen dem aktuellen Kriteriumswert und seinem Schätzwert) ein Minimum ergibt.

Ein erstes gravierendes Problem ergibt sich daraus, daß in den Modellannahmen lediglich das Kriterium als Zufallsvariable deklariert wird, während für die Prädiktoren fehlerfreie Messungen unterstellt werden (vgl. *Draper u. Smith* 1981; *Maxwell* 1975). Da in den Sozialwissenschaften diese Anforderung an die Daten in der Regel nicht erfüllt ist, fallen die Varianzschätzungen der Prädiktoren verzerrt aus, was sich in fehlerbehafteten multiplen Korrelationskoeffizienten niederschlägt („upward“ or „downward bias“, vgl. *Pedhazur* 1982). Ein zweites Problem multipler Regressionsanalysen besteht darin, daß die spezifischen Anteile von Prädiktoren an der Kriteriumsvarianz nur dann einigermaßen exakt bestimmbar sind, wenn keine Korrelationen zwischen den Prädiktoren existieren, sie also voneinander unabhängig sind. Auch diese Forderung ist in den meisten sozialwissenschaftlichen Anwendungen nicht erfüllt, so daß ein weiterer Grund für die Verzerrung von Regressions-Schätzparametern gegeben ist. Wenn man beispielsweise 10 Prädiktoren annimmt, gehen in die Regressionsgleichung immerhin 45 ($n(n-1)/2$) fehlerbehaftete Schätzungen ein (vgl. auch *Gaensslen u. Schubö* 1973). Sind die Interkorrelationen zwischen Prädiktoren insgesamt hoch (Multikollinearität), resultieren daraus ernsthafte Probleme für die Schätzung. Hohe Multikollinearität hat extrem negative Effekte für die Schätzung der Standardfehler der Regressionskoeffizienten (*Pedhazur* 1982). Allgemein gilt, daß die Verzerrung in den Schätzwerten der multiplen Regressionsanalyse um so größer ausfällt, je geringer die Reliabilität der Prädiktoren (d.h., je größer ihr Meßfehler) und je höher die Interkorrelation zwischen den Prädiktoren ausfällt. Beide Probleme tauchen in schulbezogenen Längsschnittstudien immer wieder auf. Nimmt man etwa die umfangreiche Literatur zur Prognose des Schulerfolgs, beispielsweise die Arbeiten zur frühen Vorhersage der Lesefertigkeit durch Prädiktoren, die im Kindergartenalter erhoben wurden, beeindruckt die hohe Interkorrelation der Prädiktor-Maße (s. *Schneider* 1983b, *Schneider u. Näslund* im Druck). In diesen Fällen ist es bei der Berechnung multipler Regressionsanalysen unerlässlich, die Statistiken zur Multikollinearität, die in einigen modernen Software-Paketen (z. B. SAS) ausgegeben werden können, sorgfältig zu analysieren.

Abschließend sei angemerkt, daß Meßfehler- und Multikollinearitätsprobleme in gleicher Weise für andere, sophistiziertere Varianten der Regressionsanalyse zutreffen. Sie gelten insbesondere auch für Varianzpartitionierungsverfahren, die unter dem Namen *Kommunalitätenanalyse* (*Cooley u. Lohnes* 1976) bzw. *hierarchische Regressionsanalyse* (*Cohen u. Cohen* 1983) bekanntgeworden und in der schulbezogenen Längsschnittforschung große Beachtung erfahren haben. Häufig zitierte Anwendungen finden sich in Reanalysen zu Daten des „Coleman Report“ (z. B. *Mayeske u. Beaton* 1975) oder in älteren Studien zur „Interactional Evaluation of Educational Achievement“ (IEA: *Purves* 1973). Im deutschsprachigen Bereich wurde das Verfahren ebenfalls in Studien zur Prognose des Schulerfolgs eingesetzt (*Quack* 1979; *Schneider* 1980; *Treiber* 1980). Bei diesen Verfahren geht es darum, Proportionen in der Varianz der Kriteriumsvariablen zu identifizieren, die eindeutig auf spezifische Prädiktoren zurückzuführen sind (unique variance), und davon Proportionen der Varianz zu trennen, die auf den kombinierten Effekt mehrerer Prädiktoren zurückzuführen sind (communality).

Zusätzlich zu den schon erwähnten Problemen stellt sich bei der Kommunalitätenanalyse die Frage, wie man die Kommunalitäten höherer Ordnung, die bei mehr als zwei Prädiktoren bzw. Prädiktor-Sets entstehen, inhaltlich interpretieren soll. Da sie auf der Korrelation von Prädiktoren basieren, lassen sie sich nicht als Interaktionen im üblichen (varianzanalytischen) Sinn auffassen. Das Problem der Interpretation wird besonders evident, wenn negative Vorzeichen bei den Kommunalitäten auftauchen. Dies ist besonders dann der Fall, wenn einige Prädiktorvariablen miteinander positiv, andere miteinander negativ korrelieren. Das logische Problem dieses Phänomens läßt sich auch nicht durch die Maßnahme lösen, negative Kommunalitäten auf Null zu setzen, wie sie etwa von *Mayeske* und *Beaton* (1975) praktiziert wurde. Angesichts der vielfältigen Schwierigkeiten mit der Kommunalitätenanalyse wird das Verfahren heute insbesondere dann als inadäquat eingestuft, wenn es darum gehen soll, die relative Bedeutung der Effekte unabhängiger Variablen für eine abhängige Variable zu erfassen (vgl. *Pedhazur* 1982). Es kann sich nach Auffassung von *Pedhazur* dann als sinnvoll erweisen, wenn im Rahmen eines Kausalmodells der Effekt bestimmter Variablen auf das Kriterium kontrolliert werden soll, um den Effekt anderer Prädiktoren „reiner“ zu erfassen. Zu diesem Problembereich liegen inzwischen jedoch weitaus sophisticatedere Prozeduren vor, auf deren Möglichkeiten und Probleme in der Folge eingegangen werden soll.

5.2.2 Probleme und Möglichkeiten von Strukturgleichungs-(Kausal-)Modellen

Zusätzlich zu den schon aufgeführten Problemen von multiplen Regressionsverfahren wird häufig der Kritikpunkt angeführt, daß sie allzu häufig theoriearme Forschung ermöglichen:

„The routine procedure today is to feed into a computer all the independent variates that are available and to hope for the best“ (*Maxwell* 1975, S. 53).

Neuere Längsschnittstudien zu schulbezogenen Forschungsthemen sind demgegenüber meist von genauer spezifizierbaren theoretischen Annahmen geleitet, die über das Verfahren der multiplen Regression nicht näher überprüft werden können. Längsschnittstudien, die sich beispielsweise dem Problem der gegenseitigen Abhängigkeit von schulischem Selbstkonzept und schulischen Leistungen und über Veränderung über die Zeit widmen, sind auf methodisch anspruchsvollere Analyseprozeduren angewiesen (vgl. *Pekrun* 1987). Sogenannte *Pfadanalysen mit manifesten Variablen*, über die sich Abhängigkeitsbeziehungen zwischen Prädiktorvariablen modellieren lassen, stellen konsequente Weiterentwicklungen des Regressionsansatzes dar. *Kausalmodelle mit latenten Variablen* bieten demgegenüber eine Reihe von weiteren Vorteilen, die zu einer enormen Verbreitung dieser Ansätze innerhalb recht kurzer Zeit geführt haben. Beide Varianten von Strukturgleichungsmodellen werden im folgenden näher skizziert.

5.2.2.1 Pfadanalysen mit manifesten Variablen

Der entscheidende Unterschied zwischen multiplen Regressionsmodellen und Pfadmodellen mit manifesten (beobachteten) Variablen besteht darin, daß der Forscher im Fall der multiplen Regression seine Prädiktoren allesamt als exogene Variablen behandelt, d.h. keine Erklärung für die Interkorrelation dieser Merkmale liefert, während in der Pfadanalyse hier weiterhin zwischen exogenen und endogenen Merkmalen differenziert wird. Die Pfadanalyse bietet also die Möglichkeit, postulierte Zusammenhänge zwi-

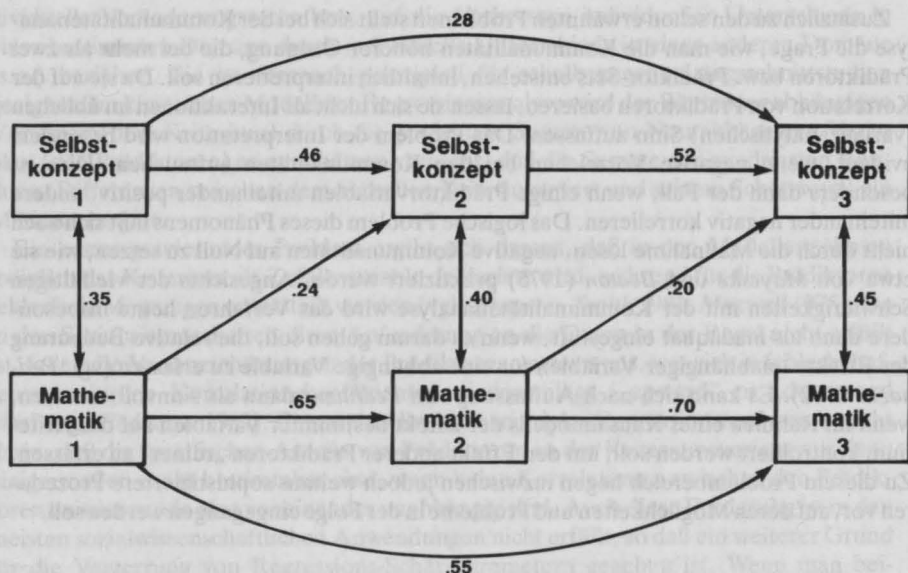


Abb. 1 Pfadmodell mit manifesten Variablen zur Beziehung zwischen schulischem Selbstkonzept und Mathematikleistungen

schen Prädiktorvariablen in ihrer Einwirkung auf ein gegebenes Kriterium genauer zu überprüfen. Dies erlaubt eine Zerlegung der Variableninterkorrelationen in direkte und indirekte Einflüsse. In Abbildung 1 sind die Vorzüge der Pfadanalyse mit manifesten Variablen gegenüber der multiplen Regression an einem fiktiven Beispiel illustriert.

Das Pfaddiagramm enthält ein Modell für drei aufeinanderfolgende Meßzeitpunkte (t_1 - t_3), für die jeweils Indikatoren für schulisches Selbstkonzept und Mathematikleistungen verfügbar sind. Lediglich die beiden zuerst erhobenen Variablen sind exogen (werden also nicht weiter durch vorhergehende Merkmale erklärt). Alle übrigen Merkmale sind endogen; für sie lassen sich direkte und indirekte Effekte der vorgeordneten Variablen ermöglichen. Doppelpfeile indizieren Korrelationen, einfache Pfeile Pfad- bzw. Strukturkoeffizienten. Das fiktive Modell in Abbildung 1 ist nicht vollständig spezifiziert: es fehlen Pfade von zeitlich vorgeordneten Selbstkonzept- zu nachgeordneten Mathematik-Variablen. Es liegt hier die theoretische Annahme zugrunde, daß die Schulleistung gegenüber dem Selbstkonzept „kausal prädominant“ ist, was meint, daß Schulleistungen das Selbstkonzept beeinflussen, nicht vice versa. Nur wenn Pfadmodelle solchermaßen restringiert sind (also Pfade aus theoretischen Gründen oder aufgrund von empirischer Bedeutungslosigkeit weggelassen werden können), lassen sie sich auch empirisch testen. Der Vergleich der Ausgangs-Korrelationsmatrix mit der über das Modell geschätzten (rückgerechneten) Korrelationsmatrix erlaubt Hinweise darauf, wie gut das theoretische Modell mit den Daten übereinstimmt. Auf diese Weise läßt sich beispielsweise das in Abbildung 1 dargestellte Modell mit einer Variante vergleichen, bei der davon ausgegangen wird, daß das Selbstkonzept die Schulleistungen beeinflusst, oder auch zu einer Version in Beziehung setzen, bei der keinerlei „reziproke Kausalität“ angenommen, also Unabhängigkeit der Entwicklungsverläufe für Selbstkonzept und Mathematikleistung postuliert wird (autoregressives Modell). Es besteht

hier also die Möglichkeit, konkurrierende Hypothesen zu testen und ggf. zu bestätigen bzw. zu falsifizieren.

Trotz dieses offensichtlichen Vorzugs von Pfadmodellen mit manifesten Variablen gegenüber den multiplen Regressionsanalysen sind ihre Anwendungsmöglichkeiten in der schulbezogenen Längsschnittforschung dann äußerst begrenzt, wenn es sich um großangelegte Projekte handelt. Beispiele dafür sind im deutschsprachigen Bereich das Konstanzer Forschungsprogramm zur Gesamtschule (Fend 1982), die Münchner und Heidelberger Studien zur Prognose von Mathematikleistungen (Helmke, Schneider u. Weinert 1986; Treiber u. Weinert 1985) sowie international gesehen die unterschiedlichen IEA-Studien (Anderson u. Mitarbeiter 1989; Travers u. Westbury 1989; Robitaille u. Garden 1989). Für diese und andere längsschnittliche Untersuchungen, in denen meist mehr als 100 Variablen zu unterschiedlichen Meßzeitpunkten erhoben wurden, lassen sich regressions- und pfadanalytische Verfahren auf der Basis manifester Variablen nicht mehr sinnvoll verwenden.

Pfadanalysen mit manifesten Variablen gehen zudem wie schon die multiplen Regressionsansätze von den unrealistischen Vorannahmen aus, daß die Variablen meßfehlerfrei erhoben wurden, und daß die Residuen von unterschiedlichen Gleichungen voneinander unabhängig sind. Gerade die letztgenannte Annahme läßt sich beispielsweise kaum für Längsschnittstudien aufrechterhalten, in denen die gleichen Probanden mehrfach mit den gleichen Instrumenten getestet werden. Für die Analyse längsschnittlicher Daten scheinen Strukturgleichungsmodelle mit latenten Variablen deshalb besser geeignet, weil sich mit ihnen einige – wenn auch nicht alle – der aufgeführten Problempunkte erfolgreich umgehen lassen.

5.2.2.2 Möglichkeiten von Kausalmodellen mit latenten Variablen

Strukturgleichungsmodelle der „zweiten Generation“ (Schneider 1986) unterscheiden sich von den oben beschriebenen pfadanalytischen Ansätzen grundlegend darin, daß zwischen latenten (nicht beobachteten) und gemessenen Variablen unterschieden wird. Diese Modelle stelle Kombinationen faktoren- und regressionsanalytischer Ansätze dar: während über ein (faktorenanalytisches) *Meßmodell* die Beziehung zwischen den gemessenen Indikatoren und den latenten Konstrukten definiert wird, dient das *Strukturgleichungsmodell* dazu, die Abhängigkeiten zwischen den latenten Variablen festzulegen. Das eigentliche Kausalmodell wird hier also auf der Ebene der latenten Konstrukte spezifiziert, die den Forscher in der Regel weitaus mehr interessieren als die zugrundeliegenden gemessenen Variablen.

In jüngerer Zeit sind mehrere Monographien bzw. Sammelbände erschienen, die brauchbare Einführungen in diesen Themenbereich bieten. Die meisten fokussieren auf die von Jöreskog und Sörbom (1984) entwickelte Pfadanalyse mit latenten Variablen (LISREL = Analysis of Linear Structural Relationships) und geben dabei z.T. detaillierte Anleitungen für die Arbeiten mit LISREL (s. Andres 1988; Byrne 1989; Loehlin 1987; Pfeifer u. Schmidt 1987). Methodisch anspruchsvollere Abhandlungen finden sich bei Hayduk (1987) oder im Sammelband von Möbus und Schneider (1986).

Die wesentlichen Vorzüge des LISREL-Ansatzes lassen sich im Hinblick auf die schulbezogene Längsschnittforschung wie folgt zusammenfassen:

1. Der Forscher wird dazu genötigt, seine impliziten Annahmen über Ursache-Wirkungszusammenhänge in ein Modell zu transferieren, das sich anhand von empirischen Daten schätzen läßt.

2. Die Unterscheidung zwischen einem Meß- und einem Strukturmodell erlaubt eine vergleichsweise ökonomische Spezifikation des Kausalmodells, da die hierfür relevanten theoretischen Konstrukte meist durch mehrere gemessene Variablen repräsentiert werden.
3. Die Unterscheidung zwischen einem Meßmodell und einem Strukturmodell ermöglicht weiterhin die getrennte Schätzung von *Meßfehlern* in den manifesten Variablen und *Spezifikations-* bzw. *Gleichungsfehlern* im Kausalmodell: große Spezifikationsfehler deuten meist darauf hin, daß das Strukturmodell nicht vollständig spezifiziert wurde und theoretisch wichtige Prädiktormerkmale fehlen.
4. Es kann zwischen der Reliabilität von gemessenen Variablen und der *Stabilität* struktureller Beziehungen eindeutig unterschieden werden. Dieses für die Längsschnittforschung besonders wichtige Problem läßt sich über Pfadanalysen mit manifesten Variablen prinzipiell nicht lösen.
5. Strukturmodelle lassen sich über LISREL nicht nur schätzen, sondern auch anhand verschiedener sog. „goodness-of-fit“-Tests auch ihre Anpassung an die verfügbaren Daten hin überprüfen. Existieren unterschiedliche Modellvorstellungen im Hinblick auf die längsschnittliche Struktur der Konstrukt-Relationen, so lassen sich diese unmittelbar testen und bezüglich ihrer Kompatibilität mit den zugrundeliegenden Daten direkt miteinander vergleichen.
6. Umgekehrt können identische Modelle für unterschiedliche (Alters-)Gruppen spezifiziert werden, um die Verallgemeinerungsmöglichkeiten des theoretischen Ansatzes zu erkunden (vgl. hierzu *Bentler 1980*).
7. Obwohl Strukturgleichungsmodelle im allgemeinen auf Korrelations- oder Kovarianzmatrizen basieren, lassen sich zusätzlich auch Mittelwertsstrukturen analysieren. Dies bedeutet in der Regel, daß für den Fall multipler Gruppenvergleiche (s.o.) *relative* Veränderungen in den geschätzten Mittelwerten von latenten Variablen über verschiedene Meßzeitpunkte hinweg erfaßt werden können.

Der Unterschied zu den in Abbildung 1 dargestellten Pfadmodellen mit manifesten Variablen läßt sich anhand eines ebenfalls fiktiven Beispiels im Zusammenhang zwischen dem Selbstkonzept für Mathematik und Mathematikleistungen in der Grundschule illustrieren. Dieses Beispiel verdeutlicht die zusätzlichen Möglichkeiten, die LISREL-Modelle zur Analyse von Längsschnittdaten gegenüber den herkömmlichen Pfadmodellen mit beobachteten Variablen bieten. Die theoretischen Konstrukte sind dabei als Kreise, die gemessenen Variablen als Vierecke dargestellt. Pfeile, die von außen auf die Vierecke treffen, bezeichnen Meßfehler in den Variablen, während die von außen auf die Kreise führenden Pfeile den Gleichungsfehler indizieren. Die Pfade von den Kreisen zu den Vierecken repräsentieren Faktorladungen (das Meßmodell), die einfachen Pfade zwischen Kreisen Struktur- bzw. Regressions-Koeffizienten für die latenten Variablen. Doppelpfeile indizieren Korrelationen zwischen den latenten Konstrukten.

Das fiktive Beispiel suggeriert, daß für beide theoretische Konstrukte mittelhohe bis hohe Stabilitäten über die Grundschulzeit hinweg angenommen werden und die Merkmale zu allen Zeitpunkten relativ zuverlässig gemessen werden können. Aus Abbildung 2 geht hervor, daß lediglich das Selbstkonzept der Kinder zu allen Zeitpunkten mit identischen Verfahren ermittelt wurde; aus naheliegenden Gründen wurden für jedes Schuljahr unterschiedliche (lehrplanvalide) Tests zur Erfassung mathematischer Fertigkeiten herangezogen.

Aus Abbildung 2 geht weiterhin hervor, daß die beiden Konstrukte zum letzten Erhebungszeitpunkt im Vergleich zu den beiden vorangehenden Wellen schwerpunktmäßig durch jeweils andere Variablen repräsentiert werden, was sich aus den Faktorladungen

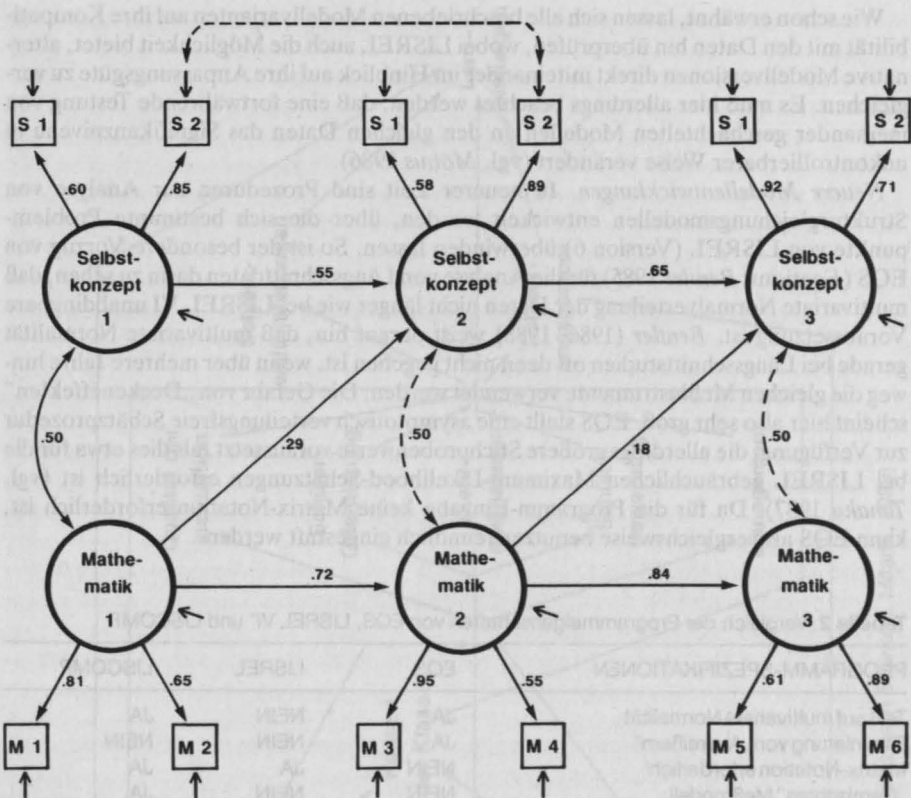


Abb. 2 Strukturgleichungsmodell mit latenten Variablen zur Beziehung zwischen schulischem Selbstkonzept und Mathematikleistungen

ablesen läßt. Bei der Interpretation der Befunde sollte deshalb immer berücksichtigt werden, daß die Bedeutung der Konstrukte sich mit der Zeit verändert hat.

Aufgrund der Flexibilität von LISREL gibt es verschiedene Möglichkeiten, das in Abbildung 2 wiedergegebene Modell mit anderen Modellversionen zu vergleichen. So könnte man beispielsweise annehmen, daß ein restriktiveres (autoregressives) Modell, bei dem keinerlei Querbeziehungen zwischen Mathematikleistung und Selbstkonzept angenommen werden, die Daten mindestens genauso gut beschreiben und erklären kann. Andererseits könnte man auch vermuten, daß das Selbstkonzept gegenüber der Mathematikleistung „kausal prädominant“ ist, die Querbeziehungen zwischen beiden Konstrukten also genau umgekehrt zu spezifizieren sind. Weitere Modifikationsmöglichkeiten sind in Abbildung 2 durch die gestrichelten Doppelpfeile (Korrelationen) angedeutet: eine zusätzliche Restriktion läßt sich in das Modell etwa dadurch einbauen, daß die Korrelation zwischen den beiden Konstrukten für alle Testzeitpunkte als invariant angenommen wird. Diese Restriktion macht das Modell natürlich leichter falsifizierbar. Der zweite gestrichelte Doppelpfeil in Abbildung 2 bezeichnet eine Korrelation zwischen den Meßfehlern von identischen Variablen, die zu unterschiedlichen Zeitpunkten vorgegeben wurden. Durch diesen bei Längsschnittdaten durchaus sinnvollen Schritt wird eine Modellrestriktion (Unabhängigkeit der Meßfehler) aufgehoben.

Wie schon erwähnt, lassen sich alle beschriebenen Modellvarianten auf ihre Kompatibilität mit den Daten hin überprüfen, wobei LISREL auch die Möglichkeit bietet, alternative Modellversionen direkt miteinander im Hinblick auf ihre Anpassungsgüte zu vergleichen. Es muß hier allerdings beachtet werden, daß eine fortwährende Testung von ineinander geschachtelten Modellen an den gleichen Daten das Signifikanzniveau in unkontrollierbarer Weise verändert (vgl. Möbus 1986).

Neuere Modellentwicklungen. In neuerer Zeit sind Prozeduren zur Analyse von Strukturgleichungsmodellen entwickelt worden, über die sich bestimmte Problem- punkte von LISREL (Version 6) überwinden lassen. So ist der besondere Vorzug von EQS (Equations; Bentler 1985) für die Analyse von Längsschnittdaten darin zu sehen, daß multivariate Normalverteilung der Daten nicht länger wie bei LISREL VI unabdingbare Voraussetzung ist. Bentler (1986, 1988) weist darauf hin, daß multivariate Normalität gerade bei Längsschnittstudien oft dann nicht gegeben ist, wenn über mehrere Jahre hinweg die gleichen Meßinstrumente verwendet werden. Die Gefahr von „Deckeneffekten“ scheint hier also sehr groß. EQS stellt eine asymptotisch verteilungsfreie Schätzprozedur zur Verfügung, die allerdings größere Stichprobenwerte voraussetzt, als dies etwa für die bei LISREL gebräuchlichen Maximum-Likelihood-Schätzungen erforderlich ist (vgl. Tanaka 1987). Da für die Programm-Eingabe keine Matrix-Notation erforderlich ist, kann EQS als vergleichsweise benutzerfreundlich eingestuft werden.

Tabelle 2 Vergleich der Programmeigenschaften von EQS, LISREL VI* und LISCOMP

| PROGRAMM-SPEZIFIKATIONEN | EQS | LISREL | LISCOMP |
|----------------------------------|------|--------|---------|
| Test auf multivariate Normalität | JA | NEIN | JA |
| Eliminierung von „Ausreißern“ | JA | NEIN | NEIN |
| Matrix-Notation erforderlich | NEIN | JA | JA |
| „Gemischtes“ Meßmodell | NEIN | NEIN | JA |
| PARAMETER-SCHÄTZUNG | | | |
| Normale Theorie | JA | JA | JA |
| Elliptische Theorie | JA | NEIN | NEIN |
| Verteilungsfreie Theorie | JA | NEIN | JA |
| Automat. Startwerte | JA | JA | JA |
| Mehrgruppenvergleich | NEIN | JA | JA |
| Latente Mittelwertsstrukturen | NEIN | JA | JA |
| OUTPUT-INFORMATION | | | |
| Automatische Modellmodifikation | NEIN | JA | NEIN |
| Standardisierte Lösung | JA | JA | JA |
| Pfaddiagramm | JA | NEIN | NEIN |

* Die neueste Version (LISREL VII) stand dem Verfasser nicht zur Verfügung

Das von Muthén (1984, 1987) entwickelte Programmsystem LISCOMP (Analysis of Linear Structural Equations Using a Comprehensive Measurement Model) ermöglicht die simultane Analyse dichotomer, kategorialer, normaler und nicht-normaler kontinuierlicher Variablen und damit die Analyse „gemischter“ Meßmodelle. Tabelle 2 stellt die drei gängigsten Programme zur Analyse von Strukturgleichungsmodellen in ihren Leistungsmöglichkeiten gegenüber, wobei allerdings selektiv vorgegangen wurde. Aufgrund der rasanten Entwicklung der letzten Jahre ist davon auszugehen, daß die noch

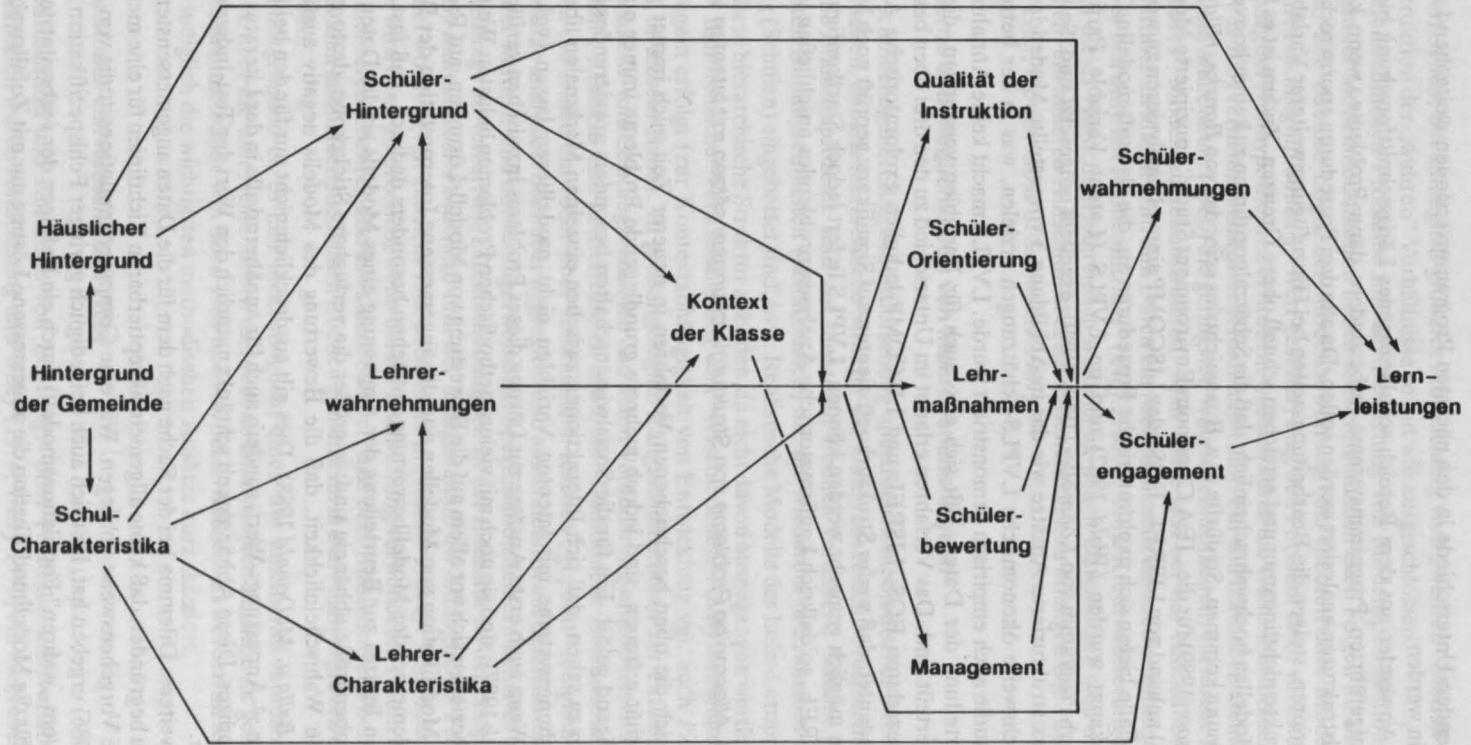


Abb. 3 Basismodell der IEA-„Classroom Environment“-Studie (modifiziert nach Bourke u. Mitarb., 1989)

bestehenden Unterschiede in den nächsten Programmversionen weitgehend aufgehoben sein werden.

Für Anwender aus dem Bereich schulbezogener Längsschnittforschung bieten alle drei aufgeführten Programmsysteme insbesondere dann Probleme, wenn komplexe Modellstrukturen analysiert werden sollen. Da alle drei Prozeduren enorm rechenintensiv operieren, steigen die Bearbeitungszeiten bei Hinzufügung weiterer Variablen bzw. Konstrukte nichtlinear an und erreichen schnell obere Grenzen. Zudem ist es bei „großen“ Modellen hochwahrscheinlich, daß die Schätzalgorithmen zu keiner konvergierenden Lösung kommen. So dürfte es z. B. aussichtslos sein, das von *Bourke, Hildyard* und *Anderson* (1989) für die „IEA Classroom Environment Study“ konzipierte Modell (vgl. Abb. 3) anhand von LISREL, EQS oder LISCOMP sinnvoll analysieren zu wollen. Für solche Fälle bieten sich exploratorische Prozeduren an, die als „soft modeling“-verfahren bekannt wurden (*Wold* 1982) und im LVPLS (Latent Variable Partial Least Squares)-Programm von *Lohmöller* (1984, 1989), exzellent umgesetzt sind. Selbst für so komplex strukturierte Ansätze wie das in Abbildung 3 vorgestellte Modell lassen sich vergleichsweise ökonomische LVPLS-Schätzungen erzielen, was in der betreffenden IEA-Studie auch empirisch demonstriert wurde. LVPLS macht keine Annahmen über die Verteilung der Daten, läßt sich also auch für Variablen verwenden, die extrem schiefverteilt sind. Das Verfahren erlaubt im Unterschied zu den drei oben beschriebenen Prozeduren EQS, LISREL und LISCOMP lediglich exploratorische Analysen. Dies bedeutet, daß weder Strukturkoeffizienten auf Signifikanz geprüft noch Modelltestungen möglich gemacht werden können. LVPLS liefert jedoch Startkonfigurationen für LISREL, so daß sich konfirmatorische Analysen problemlos anschließen lassen.

5.2.2.3 Allgemeine Probleme von Strukturgleichungsansätzen mit latenten Variablen

Wenn sich die oben beschriebenen Verfahren in neuerer Zeit auch immer größerer Popularität erfreuen, sind jedoch mehrere grundlegende Probleme immer noch nicht befriedigend gelöst. Ein für die Sozialwissenschaften besonders gravierendes Problem ist darin zu sehen, daß sich Interaktionen zwischen einzelnen Merkmalen über Strukturgleichungssysteme mit latenten Variablen nicht modellieren lassen (vgl. *Pekrun* 1990). Wenn auch erste Ansätze zur Lösung dieses Problems inzwischen verfügbar sind (*Hayduk* 1987), sind sie noch mit vielen methodischen Problemen belastet. Weitere Probleme beziehen sich vor allem auf die Bewertung von Modell-Lösungen, auf Richtlinien bei der Modifikation von Modellen und die angemessene Interpretation der Befunde.

Probleme bei der Modellbewertung entstehen besonders dadurch, daß fast alle verfügbaren Indizes zur Beurteilung der Anpassung eines Modells an die Daten von der Stichprobengröße abhängig sind: je größer die verfügbare Stichprobe, desto größer ist auch die Wahrscheinlichkeit, daß die Bewertung des Modells negativ ausfällt (vgl. *Marsh, Balla u. McDonald* 1988). Dies gilt ausdrücklich nicht nur für den bei LISREL üblichen χ^2 -Anpassungs-Wert, sondern auch für annähernd alle in der Literatur verwendeten Indizes. Diese Abhängigkeit schränkt natürlich den Wert der Test-Indizes bedeutsam ein.

Ein weiteres Dilemma bei der Suche nach dem für die Daten angemessensten Modell ist darin begründet, daß keine allgemein akzeptierbaren Richtlinien für eine methodisch probate Vorgehensweise vorliegen. Wie eine Computersimulationsstudie von *MacCallum* (1986) ergeben hat, läßt sich auch bei lediglich geringer Fehlspezifikation (Abweichung vom „wahren“ Populationsmodell) durch ein Befolgen der verbreitetsten Faustregeln für die Modellmodifikation die angemessene Lösung nur mit Zufallswahrschein-

lichkeit erreichen; bei anfänglich größerer Abweichung vom „wahren“ Modell ist eine Rekonstruktion der „wahren“ Verhältnisse so gut wie ausgeschlossen.

Die offensichtlichen Probleme mit der Bewertung von Modellspezifikationen und nachfolgender Modellmodifikation haben selbstverständlich auch Folgen für die Interpretation. Die dubiose Logik der Modellbewertung (ein theoretisches Modell wird durch einen *nichtsignifikanten* χ^2 -Wert gestützt) mag den Forscher dazu verleiten, eine Bestätigung der Nullhypothese zu erstreben, was sich durch die Wahl einer kleinen Stichprobe ($N \leq 100$) verhältnismäßig leicht realisieren läßt. Für solchermaßen kleine Stichproben fand *MacCallum* (1986) die geringsten Wahrscheinlichkeiten, ein für die Population „wahres“ Modell zu rekonstruieren. Man sollte sich bei der Modellbewertung immer darüber im klaren sein, daß Daten ein Modell niemals eindeutig bestätigen können: „they can only fail to disconfirm it“ (*Cliff* 1983, p. 118). Kreuzvalidierungen eines einmal in dieser Hinsicht erfolgreichen Modells an neuen Datensätzen sind demnach unerlässlich, um näheren Aufschluß über seine Bewährtheit zu erlangen (vgl. *Cudeck* u. *Browne* 1983; *Opwis, Gold* u. *Schneider* 1987).

6 Probleme und Möglichkeiten von Mehrebenenanalysen

Ein für schulbezogene Längsschnittforschung spezifisches methodisches Problem ist darin zu sehen, daß die hierarchische Struktur der Daten zusätzlich berücksichtigt werden muß. Bildungsprozesse sind auf mehreren Ebenen verankert: Schüler gehören bestimmten Bildungseinheiten (Schulklassen) an, die wiederum in Einheiten höherer Ordnung (Schulen) eingebettet sind, usw. Realistische Modelle des Lehr-Lernprozesses sollten diese hierarchische Struktur abbilden, da nicht davon ausgegangen werden kann, daß die hierarchische Verschachtelung der verschiedenen Analyseeinheiten nach Zufallsregeln erfolgt (zur Kontextabhängigkeit von Entwicklung vgl. auch *Pekrun* u. *Helmke*, in diesem Band). In großangelegten Längsschnittstudien zur Erfassung von Persönlichkeits- und Leistungsvariablen und deren Veränderung sind demnach systematisch Schul-, Klassen- und Lehreffekte zu erwarten. Traditionelle *Einebenenanalysen*, bei denen die Daten entweder über verschiedene Klassen „gepooled“ lediglich auf der Schülerebene analysiert oder aber aggregiert und ausschließlich auf den Klassenebenen verarbeitet werden, sind notwendigerweise mit Spezifikationsfehlern behaftet und führen demzufolge zu verzerrten Effektschätzungen. In der Literatur ist der erstgenannte Fehler als *individueller Fehlschluß* und der letztgenannte als *ökologischer Fehlschluß* bezeichnet worden (vgl. *Burstein* 1980; *Schneider* u. *Helmke* 1986a). Von *Mehrebenenanalysen* wird dann gesprochen, wenn die zur Auswertung verwendeten Daten Messungen auf mehreren Ebenen, also etwa dem Schüler- und dem Klassenniveau beinhalten, und die Daten aus unterschiedlichen Ebenen gemeinsam in den Analyseprozeß einbezogen werden.

Aus Platzgründen ist es hier nicht möglich, einen nur annähernd repräsentativen Überblick über die unterschiedlichen mehrebenenanalytischen Ansätze zu geben (vgl. hierzu die Sammelbände von *Bock* 1989 und von *Saldern* 1986). Es werden im folgenden deshalb lediglich die wichtigsten methodischen Ansätze kurz skizziert.

6.1 Das Modell von Keesling und Wiley

Zur Identifikation und Kontrolle von Leistungs- oder Persönlichkeitsvariablen und Klasseneffekten wurde von *Keesling* und *Wiley* (1974; *Wiley* 1976) ein mehrebenenana-

lytischer Ansatz entwickelt, bei dem einmal der Einfluß der Klassenzugehörigkeit bei der Regression von Schulleistungen auf individuelle Hintergrundsmerkmale (z.B. Intelligenz, Sozialschicht) berücksichtigt und zum anderen gleichzeitig der Effekt von Gruppen- (Instruktions-) Merkmalen auf die auf Klassenebene aggregierten Schulleistungen um den Einfluß individueller Unterschiede in den Klassen korrigiert wird. Konkret sieht dies so aus, daß zunächst auf Schülerebene eine Kovarianzanalyse durchgeführt wird, wobei als abhängige Variable die individuelle Klassenzugehörigkeit und als Kovariaten die Schüler-Hintergrundsmerkmale fungieren. Über diese „gepoolte“ Innerklassen-Regression erhält man für jede Schulklasse einen geschätzten Schulleistungswert, der dann zusätzlich in die sich anschließende Regression der beobachteten Klassenmittelwerte auf Erklärungsvariablen der Schulklassenebene (z. B. Lehrer- und Unterrichtsmerkmale) aufgenommen wird. *Keesling* und *Wiley* gehen davon aus, daß durch die Berücksichtigung des geschätzten Schulleistungswerts, der als summierter Personeneffekt aufzufassen ist, eine verbesserte Schätzung der auf der Gruppenebene aufzuklärenden Klassenunterschiede in der abhängigen Variablen erfolgen kann. Empirische Anwendungen dieses Modells finden sich z. B. bei *Treiber* (1981) bzw. *Treiber* und *Schneider* (1981).

6.2 Das SVS- (systematically varying slopes-)Modell von *Burstein* und *Linn*

Burstein, *Linn* und Mitarbeiter (*Burstein*, *Linn* u. *Capell* 1978; *Burstein* u. *Linn* 1986) haben einen mehrebenenanalytischen Auswertungsansatz empfohlen, der unter der Bezeichnung „slopes as outcomes“ bzw. als SVS-Modell bekannt geworden ist. In diesem Ansatz wird unterstellt, daß aufgrund unterschiedlicher Kontextbedingungen in unterschiedlichen Schulklassen die Innerklassen-Regressionskoeffizienten (z. B. der Vortest-Nachtestbeziehung) *systematisch* heterogen ausfallen, demnach etwa auf unterschiedliche Unterrichtspraktiken in den einzelnen Klassen rückführbar sind. Flache Prätest-Posttest-Regressionssteigerungen könnten etwa als Effekte kompensatorischer Unterrichtsbemühungen interpretiert, steile Regressionsgeraden (d.h., hoch-positive Beta-Koeffizienten) dagegen auf eher traditionelle Unterrichtspraktiken zurückgeführt werden, bei denen sich Schüler mit höheren Eingangsfähigkeiten in Relation zu schwächeren Schülern stärker entfalten können.

Das mehrebenenanalytische Verfahren erfolgt in zwei Schritten. Zunächst werden die spezifischen Innerklassen-Regressionskoeffizienten für die einzelnen Schulklassen ermittelt. Im zweiten Schritt werden diese Regressionskoeffizienten als abhängige Variablen in einer Interklassen-Regressionsanalyse betrachtet, bei der hochaggregierte Schülervariablen (z. B. durchschnittliche Vortestleistungen und Hintergrundsmerkmale) zusammen mit verschiedenen auf Klassenebene erfaßten Instruktionsmerkmalen als Prädiktoren fungieren.

6.3 Das HLM-Modell von *Bryk* und *Raudenbush*

Das oben im Zusammenhang mit individuellen Veränderungsanalysen schon näher beschriebene Hierarchical Linear Modeling-Programm von *Bryk* und *Raudenbush* (1987) läßt sich auch sinnvoll für die Analyse hierarchisch strukturierter Daten aus dem schulischen Bereich verwenden. *Bryk* und *Raudenbush* (1987, 1989) setzen dabei in ihren Illustrationsbeispielen am SVS-Ansatz von *Burstein* und Mitarbeitern an, wählen jedoch nicht die Klassen-, sondern die Schulebene, um die Brauchbarkeit von HLM für

solche Fragestellungen zu demonstrieren. Für alle Schüler einer bestimmten Schule wird in einem ersten Schritt ein „within-subject“-Regressionsmodell spezifiziert, in dem die Schulleistungen als Funktion unterschiedlicher Hintergrundmerkmale der Schüler vorhergesagt werden. Wie bei SVS-Modellen üblich, werden im zweiten Schritt die für die unterschiedlichen Schulen bei der Vorhersage des Kriteriums ermittelten Regressionskoeffizienten im „between-subjects“-Modell nun selber als Kriteriumsvariablen definiert: es geht darum, die Varianz zwischen den Regressionskoeffizienten auf Unterschiede zwischen den Schulen (z. B. ihrer Infrastruktur, Organisation etc.) zurückzuführen.

Die Entscheidung der Autoren, anstatt von Klasseneffekten ausschließlich Schuleffekte zu analysieren, macht es möglich, die „within-subject“-Regressionskoeffizienten robust zu schätzen. Vorteilhaft ist es in diesem Zusammenhang sicherlich auch, daß empirische Bayes-Verfahren zur Schätzung verwendet werden (weitere Anwendungen empirischer Bayes-Methoden finden sich in dem von *Bock* (1989) herausgegebenen Sammelband). Das Problem des Ansatzes scheint mir jedoch darin zu liegen, daß Klasseffekte nicht weiter berücksichtigt werden.

6.4 Strukturgleichungsmodelle mit latenten Variablen

Da die oben beschriebenen regressionsanalytischen Ansätze keine befriedigende Lösungen zur Behandlung der Meßfehler-Problematik anbieten, liegt es nahe, Strukturgleichungsmodelle mit latenten Variablen für mehrebenenanalytische Fragestellungen heranzuziehen. Die Möglichkeit, Kovarianzstrukturen für unterschiedliche Gruppen (Schulklassen) via LISREL zu vergleichen, scheint jedoch methodisch problematisch: Zum einen sind Modellschätzungen für Stichproben von maximal 30 bis 40 Elementen (Schüler) mit einiger Sicherheit extrem verzerrt, zum anderen werden die Grenzen von LISREL auch für eine mittelgroße Studie mit etwa 20 Schulklassen dann leicht erreicht, wenn simultane Gruppenvergleiche für alle Klassen durchgeführt werden sollen.

Als Konsequenz aus diesen Überlegungen schlagen *Burstein, Kim* und *DeLandshere* (1989) eine Prozedur vor, bei der (1) auf irgendeine Weise Cluster von Schulklassen gebildet werden; (2) geprüft wird, ob die Kovarianzstrukturen innerhalb dieser Cluster invariant sind; (3) im Falle einer Bestätigung von (2) die Daten für jedes resultierende Cluster von Schulklassen „gepoolt“ werden, und (4) die Invarianz der Kovarianzstrukturen zwischen den Clustern getestet wird. Für den Fall, daß Makro-Merkmale (z. B. unterschiedliche Unterrichtsprozesse) als Grundlage der Clusterbildung dienen, kann eine statistisch signifikante Variabilität der Kovarianzstrukturen zwischen den Clustern von Schulklassen auf die Makro-Merkmale zurückgeführt werden.

Eigene Untersuchungen mit diesem Ansatz (*Schneider* u. *Helmke* 1986b; *Schneider* u. *Treiber* (1984) basierten auf den Annahmen des oben beschriebenen SVS-Modells, gingen also davon aus, daß Unterschiede in den Innerklassen-Regressionskoeffizienten substantielle Unterschiede in Unterrichtsprozessen widerspiegeln können. In der Studie von *Schneider* und *Treiber* (1984) wurden beispielsweise zunächst für alle beteiligten Schulklassen Posttest-Prätest-Regressionen durchgeführt, um Extremgruppen von Schulklassen mit äußerst hohen (High-Slope-) oder flachen (Low-Slope-) Innerklassen-Regressionskoeffizienten zu selektieren. Für die gebildeten Cluster ließ sich dann auf Individualebene über LISREL prüfen, ob das gleiche Schulleistungsmodell an die Daten beider Extremgruppen gleichermaßen gut angepaßt werden konnte. Der Befund, daß sich das spezifiziertere Modell nur für die „High-Slope“-Klassen bestätigen ließ, konnte in einer breiter angelegten Folgestudie (*Schneider* u. *Helmke* 1986b) weit-

gehend repliziert werden. Die Ergebnisse wurden insgesamt als Bestätigung der These gewertet, daß Schulleistungsmodelle nur „lokal“ anwendbar sind (für weitere LISREL-Anwendungen bei der Analyse hierarchisch strukturierter Daten vgl. *Wisembaker u. Schmidt 1986; Muthén u. Satorra 1989*).

6.5 Zusammenfassende Bewertung methodischer Ansätze zur Analyse hierarchisch strukturierter Daten

Ein grundsätzliches methodisches Problem vieler schulbezogener Längsschnittstudien ist darin zu sehen, daß in ihren Datenanalysen der Mehrebenencharakter des Gegenstandsbereichs nicht angemessen modelliert wird. Mehrebenenanalytische Ansätze sollten insbesondere dann verwendet werden, wenn die Datenbasis der Längsschnittuntersuchungen ausreichend groß ist, also gewährleistet werden kann, daß auf der obersten Hierarchie-Ebene (z. B. Schulklasse oder Schule) genügend Einheiten verfügbar sind, um multivariate Statistikprozeduren einzusetzen.

Nicht alle der hier skizzierten mehrebenenanalytischen Modellansätze stellen Optionen dar, die für die Anwendung zu empfehlen sind. Das Modell von *Keesling und Wiley* gilt strenggenommen nur unter der Voraussetzung, daß die schulklassenspezifischen Regressionssteigerungen parallel verlaufen, m.a.W. nicht signifikant voneinander abweichen. Für den Fall signifikanter Modellabweichungen treten in der Praxis erhebliche Probleme auf, wenn ein Modell neu spezifiziert werden muß. Das SVS-Modell basiert demgegenüber auf der Annahme, daß sich die schulklassenspezifischen Regressionssteigerungen signifikant unterscheiden. Seine grundlegende Schwachstelle ist darin zu sehen, daß die Schätzung der Regressionssteigerungen auf sehr kleinen Fallzahlen basiert. In diesem Punkt wirkt das HLM-Verfahren von *Bryk und Raudenbush* deutlich solider, zumindest wenn es für Forschungsarbeiten herangezogen wird, die sich mit den Kontexteffekten von Schulen (und nicht mit denen von Schulklassen) beschäftigen. Allen drei genannten Modellansätzen ist gemeinsam, daß sie auf manifesten Variablen basieren, die Meßfehlerproblematik demnach ausgeklammert bleibt.

Gerade in diesem Punkt bieten die vorgestellten LISREL-Modelle große Vorteile. Trotz ihrer großen Flexibilität ist es mit ihnen jedoch noch immer nicht möglich, Wechselwirkungsbeziehungen (deren Annahme in Mehrebenenmodellen unentbehrlich ist) angemessen zu modellieren. Weiterhin macht es Schwierigkeiten, die für unterschiedliche Gruppen als unterschiedlich angenommenen Fehlerstrukturen (die Heteroskedastizität von Fehlern) angemessen zu verarbeiten. Aus diesen Ausführungen geht hervor, daß auch weiterhin keine Methode verfügbar ist, in der alle aufgeführten Probleme minimiert sind. Dennoch steht es außer Zweifel, daß die in neuerer Zeit präsentierten Methoden zur Behandlung der Mehrebenenproblematik von Schulleistungsmodellen eindeutige Fortschritte darstellen. Angesichts der hier eröffneten Möglichkeiten ist schwer zu rechtfertigen, wenn man sich in schulbezogenen Längsschnittstudien auf ein-ebenenanalytische Auswertungen beschränkt.

7 Schlußbemerkungen

Der Versuch, methodische Probleme und Möglichkeiten schulbezogener Längsschnittforschung aufzuzeigen, konnte aus Platzgründen nur unvollständig ausfallen. Die Darstellung fokussierte auf Analyseverfahren, die bevorzugt mit Intervall- oder Ordinaldaten operieren, verzichtete also weitgehend auf die Beschreibung von Prozeduren, die

Vorhersagemodelle für qualitative Daten prüfen (vgl. hierzu z. B. *Rudinger, Chaselon, Zimmermann* u. *Henning* 1985). Mein Anliegen bestand im wesentlichen darin, methodische Probleme und Möglichkeiten bei der Auswertung komplexer, groß angelegter Längsschnittstudien im schulischen Bereich offenzulegen. Von den Ausführungen sollten demnach alle diejenigen Forscher profitieren können, denen Varianzanalysen mit Meßwiederholungen für einzelne Kriteriumsvariablen allein nicht genügen, die also primär daran interessiert sind, möglichst viele ihrer Beobachtungen simultan zu verarbeiten.

Die Quintessenz meiner Ausführungen läßt sich so zusammenfassen, daß neuere Versionen von Strukturgleichungsmodellen wohl am ehesten dazu geeignet sind, erhöhten Ansprüchen an die Auswertung von schulbezogenen Längsschnittdaten zu genügen. Unabhängig von den in jüngerer Zeit beobachtbaren Fortschritten in der Programmentwicklung (vgl. etwa *McArdle* u. *Epstein* 1987; *Rudinger, Andres, Rietz* u. *Schneider* 1989) sollte jedoch nicht übersehen werden, daß auch diese Prozeduren nach wie vor Schwachstellen aufweisen, die bei der Interpretation von Ergebnissen zu berücksichtigen sind. Ein besonders gravierendes Problem für die Praxis dürfte darin bestehen, daß die korrekte Handhabung von Strukturgleichungsmodellen große methodologische Expertise und technisches Wissen voraussetzt. Es besteht hier die Gefahr, daß sich mit Einführung immer sophistizierterer Modellkonstruktionen der Abstand zwischen denjenigen, die diese Modelle anwenden können, und dem großen Rest immer mehr vergrößert (vgl. *Hertzog* 1990). Unerwünschte Effekte dieser Entwicklung sind darin zu sehen, daß (a) zunehmend Mißbrauch mit den Methoden betrieben wird und (b) fragwürdige Forschungsergebnisse von denjenigen unkritisch rezipiert werden, die zwischen angemessenen und unangemessenen Anwendungen von Strukturgleichungsmodellen nicht unterscheiden können. Schließlich könnte sich auch in Zukunft ein negativer Trend in die Richtung ergeben, daß sich LISREL-Anwender von der großen Gemeinde pädagogisch-psychologischer Forscher zunehmend wissenschaftlich entfremden. Ein Teilziel der vorliegenden Arbeit bestand darin, dieser Entwicklung entgegenzuwirken.

Literatur

- Anderson, L.W., Ryan, D.W., Shapiro, B.J.* (Eds.): The IEA Classroom Environment Study. Pergamon Press, New York 1989
- Andres, J.*: Grundlagen linearer Strukturgleichungsmodelle. Lang, Frankfurt a.M. 1989
- Appelbaum, M.I., McCall, R.B.*: Design and analysis in developmental psychology. In: *P.H. Mussen* (Hrsg.), Handbook of child psychology: History, theory, and methods (3rd ed.) (Bd. 1, S. 415–476). Wiley, New York 1983
- Baltes, P.B., Nesselroade, J.R.*: History and rationale of longitudinal research. In: *J.R. Nesselroade, P.B. Baltes* (Hrsg.), Longitudinal research in the study of behavior and development (S. 1–39). Academic Press, New York 1979
- Bentler, P.M.*: Multivariate analysis with latent variables: Causal modeling. Annual Review of Psychology 31, 1980, 419–456
- Bentler, P.M.*: Theory and implementation of EQS: A structural equations program. BMDP Statistical Software Corp., Los Angeles 1985
- Bentler, P.M.*: EQS – Ein Ansatz zur Analyse von Strukturgleichungsmodellen für normal- bzw. nichtnormal verteilte quantitative Variablen. In: *C. Möbus, W. Schneider* (Hrsg.), Strukturmodelle für Längsschnittdaten und Zeitreihen (S. 27–56). Huber, Bern 1986
- Bentler, P.M.*: Causal modeling via structural equation systems. In: *J.R. Nesselroade, R.B. Cattell* (Hrsg.), Handbook of multivariate

- experimental psychology (2nd edition) (S. 317–336). Plenum Press, New York 1988
- Bock, R.D.* (Ed.): Multilevel analysis of educational data. Academic Press, New York 1989
- Bourke, S., Hildyard, A., Anderson, L.W.:* Variables, models, and instruments. in: *L.W. Anderson, D.W. Ryan, B.J. Shapiro* (Hrsg.), *The IEA Classroom Environment Study* (S. 15–43). Pergamon Press, New York 1989
- Bryk, A.S., Raudenbush, S.W.:* Application of hierarchical linear models to assessing change. *Psychological Bulletin* 101, No. 1, 1987, 147–158
- Bryk, A.S., Raudenbush, S.W.:* Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. In: *R.D. Bock* (Hrsg.), *Multilevel analysis of educational data* (S. 159–204). Academic Press, New York 1989
- Burstein, L.:* The analysis of multilevel data in educational research and evaluation. In: *D. Berliner* (Hrsg.), *Review of research in education* (Bd. 8, S. 158–233). American Educational Research Association, Washington, DC 1980
- Burstein, L., Kim, K.S., DeLandshere, G.:* Multilevel investigations of systematically varying slopes: Issues, alternatives, and consequences. In: *R.D. Bock* (Hrsg.), *Multilevel analysis of educational data* (S. 237–276). Academic Press, New York 1989
- Burstein, L., Linn, R.L.:* Analysen von Erziehungseffekten aus einer Mehrebenenperspektive: Zwischen- und Innerklassenbeziehungen im Mathematikleistungsbereich. In: *M.V. Saldern* (Hrsg.), *Mehrebenenanalyse: Beiträge zur Erfassung hierarchisch strukturierter Realität* (S. 18–43). Beltz, Weinheim 1986
- Burstein, L., Linn, R.L., Capell, F.J.:* Analyzing multi-level data in the presence of heterogeneous within-class regressions. *Journal of Educational Statistics* 3, 1978, 47–383
- Byrne, B.M.:* *A primer of LISREL*. Springer-Verlag, New York 1989
- Cliff, N.:* Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research* 18, 1983, 115–126
- Cohen, J., Cohen, P.:* *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd edition). Erlbaum, Hillsdale, NJ 1983
- Cooley, W.W., Lohnes, P.R.:* *Evaluation research in education*. Wiley, New York 1976
- Cronbach, L.J., Furby, L.:* How should we measure „change“ – or should we? *Psychological Bulletin* 74, 1970, 68–80
- Cudeck, R., Browne, M.W.:* Cross-validation of covariance structures. *Multivariate Behavioral Research* 18, 1983, 147–167
- Draper, N.R., Smith, H.:* *Applied regression analysis* (2nd edition). Wiley, New York 1981
- Fend, H.:* *Gesamtschule im Vergleich*. Beltz, Weinheim 1982
- Gaensslen, H., Schubö, W.:* *Einfache und komplexe statistische Analyse*. Reinhardt, München 1973
- Gollob, H.F., Reichardt, C.S.:* Taking account of time lags in causal models. *Child Development* 58, 1987, 80–92
- Goldstein, H.:* *The design and analysis of longitudinal studies*. Academic Press, London 1979
- Hayduk, L.A.:* *Structural equation modeling with LISREL*. The John Hopkins University Press, Baltimore 1987
- Helmke, A., Schneider, W., Weinert, F.E.:* Quality of instruction and classroom learning outcomes: The German contribution to the IEA Classroom environment study. *Teaching and Teacher Education* 2, 1986, 1–18
- Hertzog, C.:* On the utility of structural equation models for developmental research. In: *P.B. Baltes, D.L. Featherman, R.M. Lerner* (Hrsg.), *Life-span development and behavior* (Bd. 10, S. 257–290). Erlbaum, Hillsdale, NJ 1990
- Janson, C.G.:* Some problems of longitudinal research in the Social Sciences. In: *F. Schulzinger, S.A. Mednick, J. Knop* (Hrsg.), *Longitudinal research: Methods and uses in Behavioral Science* (S. 19–55). Martinus Nijhoff Publishing, Boston 1981
- Jöreskog, K.G., Sörbom, D.:* *LISREL VI – Analysis of linear structural relationships by maximum likelihood instrumental variables, and least squares methods* (Users Guide). Scientific Software, Mooresville 1984
- Keesling, J.W., Wiley, D.E.:* *Regression models of hierarchical data*. Paper presented at the meetings of the Psychometric Society, Stanford 1974
- Keeves, J.P.* (Ed.): *Educational research, methodology, and measurement: An international handbook*. Pergamon Press, New York 1988
- Loehlin, J.C.:* *Latent variable models*. Erlbaum, Hillsdale, NJ 1987
- Lohmöller, J.B.:* *LVPLS program manual*. Zentralarchiv für Empirische Sozialforschung, Köln 1984
- Lohmöller, J.B.:* *Latent variable path modeling with Partial Least Squares*. Physica-Verlag, Heidelberg 1989

- MacCallum, R.*: Specification searches in covariance structure modeling. *Psychological Bulletin* 100, 1986, 107–120
- Magnusson, D.*: Some methodology and strategy problems in longitudinal research. In: *F. Schulsinger, S.A. Mednick, J. Knop* (Hrsg.), *Longitudinal research – Methods and uses in behavioral science*. Martinus Nijhoff Publishing, Boston 1981
- Marsh, H.W., Balla, J.R., McDonald, R.P.*: Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin* 103, 1988, 391–410
- Maxwell, A.E.*: Limitations on the use of the multiple linear regression model. *British Journal of Mathematical and Statistical Psychology* 28, 1975, 51–62
- Mayeske, G.W., Beaton, A.E.*: *Special studies of our nation's students*. U.S. Government Printing Office, Washington, DC 1975
- McArdle, J.J., Epstein, D.*: Latent growth curves within developmental structural equation models. *Child Development* 58, 1987, 110–133
- Möbus, C.*: Analyse von Rohdaten mit LISREL: Allgemeines Lineares Modell, stochastische Differenzen- und Differentialgleichungssysteme. In: *C. Möbus, W. Schneider* (Hrsg.), *Strukturmodelle zur Analyse von Längsschnittdaten: LISREL, Pfad- und Varianzanalyse* (S. 57–127). Huber, Bern 1986
- Möbus, C., Schneider, W.* (Hrsg.): *Strukturmodelle zur Analyse von Längsschnittdaten: LISREL, Pfad- und Varianzanalyse*. Huber, Bern 1986
- Muthén, B.O.*: A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* 49, 1984, 115–132
- Muthén, B.O.*: LISCOMP – Analysis of linear structural equations using a comprehensive measurement model. Scientific Software, Inc., Mooresville, IN 1987
- Muthén, B., Satorra, A.*: Multilevel aspects of varying parameters in structural models. In: *R.D. Bock* (Hrsg.), *Multilevel analysis of educational data* (S. 87–99). Academic Press, New York 1989
- Nesselroade, J.R., Stigler, S.M., Baltes, P.B.*: Regression toward the mean and the study of change. *Psychological Bulletin* 88, 1980, 622–637
- Opwis, K., Gold, A., Schneider, W.*: Möglichkeiten der Kreuzvalidierung von Strukturgleichungsmodellen. *Psychologische Beiträge* 29, 1987, 60–77
- Pedhazur, E.J.*: *Multiple regression in behavioral research* (2nd edition). Holt, Rinehart u. Winston, New York 1982
- Pekrun, R.*: Die Entwicklung leistungsbezogener Identität bei Schülern. In: *H.-P. Frey, K. Hausser* (Hrsg.), *Identität* (S. 43–57). Enke, Stuttgart 1987
- Pekrun, R.*: Social support, achievement evaluations, and self-concepts in adolescence. In: *L. Oppenheimer* (Hrsg.), *The self-concept* (S. 107–119). Springer-Verlag, Berlin 1990
- Pfeifer, A., Schmidt, P.*: LISREL: Die Analyse komplexer Strukturgleichungsmodelle. Fischer, Stuttgart 1987
- Purves, A.C.*: *Literature education in ten countries*. Wiley, New York 1973
- Quack, L.*: Untersuchung zur Bedingungsanalyse der Schulleistung durch kognitive und nicht-kognitive Merkmale der Schülerpersönlichkeit. In: *K.F. Klauer, H.H. Kornadt* (Hrsg.), *Jahrbuch für empirische Erziehungswissenschaft* (S. 93–116). Schwann, Düsseldorf 1979
- Robitaille, D.F., Garden, R.A.* (Eds.): *The IEA study of mathematics II: Contexts and outcomes of school mathematics*. Pergamon Press, New York 1989
- Rogosa, D.R.*: Myths about longitudinal research. In: *K.W. Schaie, R.T. Campbell, W.M. Meredith, C.E. Rawlings* (Hrsg.), *Methodological problems in aging research* (S. 171–209). Springer-Verlag, New York 1988
- Rogosa, D.R., Brandt, D., Zimowski, M.*: A growth curve approach to the measurement of change. *Psychological Bulletin* 90, 1982, 726–748
- Rudinger, G., Andres, J., Rietz, C., Schneider, W.*: Structural equation models for studying intellectual development. Paper presented at European Science Foundation Workshop on Methodological Issues in Longitudinal research „Stability and change: Methods and models for data treatment“. Soria Moria, Norway 1989
- Rudinger, G., Chaselon, F., Zimmermann, J., Henning, H.J.*: *Qualitative Daten – Neue Wege sozialwissenschaftlicher Methodik*. Urban u. Schwarzenberg, München 1985
- Schneider, W.*: *Bedingungsanalysen des Rechtschreibens*. Huber, Bern 1980
- Schneider, W.*: Strukturgleichungsmodelle der zweiten Generation: Eine Einführung. In: *C. Möbus, W. Schneider* (Hrsg.), *Strukturmodelle für Längsschnittdaten und Zeitreihen* (S. 13–26). Huber, Bern 1986

- Schneider, W.*: Problems of longitudinal studies with children: Practical, conceptual, and methodological issues. In: *M. Brambring, F. Lösel, H. Skowronek* (Hrsg.), *Children at risk: Assessment, longitudinal research, and intervention* (S. 313–335). De Gruyter, New York 1989a
- Schneider, W.*: Möglichkeiten der frühen Vorhersage von Leseleistungen im Grundschulalter. *Zeitschrift für Pädagogische Psychologie* 3, 1989b, 157–168
- Schneider, W.*: The longitudinal study of motor development: Methodological issues. In: *A.F. Kalverboer, B. Hopkins, R.H. Geuze* (Hrsg.), *Motor development in early and later childhood: Longitudinal approaches*. Cambridge University Press, Cambridge (im Druck)
- Schneider, W., Edelstein, W.*: Inventory of European longitudinal studies on the Behavioral and Medical Sciences. Max-Planck-Institut für Bildungsforschung, Berlin 1990
- Schneider, W., Helmke, A.*: Mehrebenenanalytische Ansätze zur Erklärung von Schulleistungen. In: *M.V. Saldern* (Hrsg.), *Mehrebenenanalyse: Beiträge zur Erfassung hierarchisch strukturierter Realität* (S. 170–193). Beltz, Weinheim 1986a
- Schneider, W., Helmke, A.*: The role of classroom differences in achievement changes. *European Journal of the Psychology of Education* 1, 1986b, 81–91
- Schneider, W., Näslund, J.C.*: Cognitive prerequisites of reading and spelling: A longitudinal approach. In: *A. Demetriou, M. Shayer, A. Efklides* (Hrsg.), *The modern theories of cognitive development go to school*. Routledge, London (im Druck)
- Schneider, W., Treiber, B.*: Classroom differences in the determination of achievement changes. *American Educational Research Journal* 21, 1984, 195–211
- Siegler, R.S., Jenkins, E.*: How children discover new strategies. Erlbaum, Hillsdale, NJ 1989
- Tanaka, J.S.*: „How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child Development* 58, 1987, 134–136
- Travers, K.J., Westbury, J.* (Eds.): *The IEA study of mathematics I: Analysis of mathematics curricula*. Pergamon Press, New York 1989
- Treiber, B.*: Qualifizierung und Chancengleichheit in Schulklassen. Teil 2: Empirische Untersuchungen. Lang, Frankfurt a.M. 1980
- Treiber, B.*: Bildungseffekte in Mehrebenenanalysen individueller Schulleistungen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 13, 1981, 217–226
- Treiber, B., Schneider, W.*: Schulisches Lernen im sozialökologischen Kontext. In: *K.J. Klauer, H.J. Kornadt* (Hrsg.), *Jahrbuch für Empirische Erziehungswissenschaft* (S. 181–215). Schwann, Düsseldorf 1981
- Treiber, B., Weinert, F.E.*: *Gute Schulleistungen für alle?* Aschendorff, Münster 1985
- Verdonik, F., Sherrod, L.R.*: An inventory of longitudinal research on childhood and adolescence. Social Science Research Council, New York 1984
- Von Saldern, M.*: Mehrebenenanalyse: Beiträge zur Erfassung hierarchisch strukturierter Realität. Beltz, Weinheim 1986
- Wiley, D.E.*: Another hour, another day. In: *W.H. Sewell, D. Alwin* (Hrsg.), *Schooling and achievement in American society*. Academic Press, New York 1976
- Wisnabaker, J.M., Schmidt, W.H.*: Die Anwendung der Strukturgleichungsanalyse auf Mehrebenenanalysen. In: *M.V. Saldern* (Hrsg.), *Mehrebenenanalyse: Beiträge zur Erfassung hierarchisch strukturierter Realität* (S. 122–134). Beltz, Weinheim 1986
- Wohlwill, J.F.*: Cognitive development in childhood. In: *O.G. Brim, Jr., J. Kagan* (Hrsg.), *Constancy and change in human development* (S. 359–444). Harvard University Press, Cambridge, Mass. 1980
- Wold, H.*: Soft modeling: The basic design and some extensions. In: *K.G. Jöreskog, H. Wold* (Hrsg.), *Systems under indirect observation: Causality, structure, prediction* (Bd. 2, S. 1–54). North Holland, Amsterdam 1982