

Determinanten der Leistungsveränderung im Mathematikunterricht – Ein Modellvergleich in unterschiedlichen Schulklassentypen

W. Schneider und B. Treiber

Einleitung

In einer Reihe neuerer Lehr-Lern-Modelle werden Eingangsfähigkeiten von Schülern und Unterrichtsmerkmale als Determinanten von Schülerleistungen aufgefaßt (s. Cronbach und Snow, 1977). Empirische Überprüfungen von Schulleistungsmodellen sind üblicherweise so vorgenommen worden, daß der relative Einfluß von Schülerfähigkeit und Unterrichtsstil auf ein spezifisches Leistungsmaß erfaßt und verglichen wurde. Auch die hier vorzustellende Studie befolgte dieses Prinzip, überprüfte jedoch zusätzlich die Frage, ob ein einziges Schulleistungsmodell für Schulklassen mit sehr unterschiedlichen Unterrichts-"Geschichten" valide ist. Zur Konstruktion des Schulleistungsmodells und Auswahl der Schulklassen-Typen: Es wurde von einem längsschnittlich angelegten Schulleistungsmodell mit einer Simplex-Struktur ausgegangen. Simplex-Modelle sind durch ein typisches Interkorrelationsmuster gekennzeichnet, das einen konsistenten Abfall der Koeffizientenwerte mit ansteigender Distanz zur Matrix-Diagonalen aufweist. Wie schon beispielsweise Jöreskog (1979) und

1) Eine englischsprachige Langversion dieses Artikels wurde unter dem Titel "Classroom Differences in the Determination of Achievement Changes" im American Educational Research Journal, Spring 1984, Vol. 21, No. 1, pp. 195-211 publiziert. Wir danken der American Educational Research Association für die Copyright-Genehmigung.

Werts, Linn und Jöreskog (1977) gezeigt haben, scheinen Simplex-Modelle insbesondere zur Analyse von schulischem Leistungsfortschritt geeignet zu sein. Ihre Struktur läßt sich als Pfadmodell beschreiben, bei dem signifikante Pfade nur zwischen unmittelbar benachbarten Meßzeitpunkten existieren. Für den Fall, daß das Modell sich an die beobachteten Daten anpassen läßt, können mit Ausnahme des ersten und letzten Meßzeitpunkts Reliabilitäten für die beobachteten Variablen und unkorrigierte Korrelationen zwischen den latenten Variablen berechnet werden (vgl. Werts et al., 1977).

Die drei relevanten Komponenten des Schulleistungsmodells (Fähigkeit, Unterricht und Schulleistung) wurden als latente Konstrukte konzipiert, die über ein Meßmodell mit ihren beobachteten Indikatoren verknüpft waren. Zur Erfassung der Schülerfähigkeit wurden dabei insgesamt sechs Subtests aus standardisierten Intelligenztests herangezogen. Weiterhin wurden über ein komplexes Kodierungsschema Beobachtungsdaten für die Instruktionsquantität und -qualität in den ausgewählten Schulklassen gewonnen, von denen für die vorliegende Untersuchung lediglich zwei Variablen berücksichtigt wurden ("Einführung von neuem Unterrichtsstoff" und "Wiederholung und Einübung von schon vorher eingeführten Unterrichtseinheiten"). Der abhängige bzw. endogene Variablenblock der Untersuchung wurde durch individuelle Unterschiede in der Mathematikleistung repräsentiert, die zu vier Meßzeitpunkten innerhalb eines Schuljahres erhoben wurde. Als Leistungsmaß wurde der Gesamtwert eines Mathematik-Tests herangezogen, der aus 36 multiple-choice-Items bestand und die jeweiligen Unterrichtsinhalte umfassend zu berücksichtigenden versuchte (zur genaueren Variablenbeschreibung vgl. Schneider und Treiber, 1984). Es wurde jeweils der gleiche Test vorgegeben, und auch die Verfahren zur Durchführung und Auswertung blieben für alle Meßzeitpunkte identisch.

Zur Rekrutierung unterschiedlicher Schulklassen-Typen

wurden Unterschiede in den für jede Klasse separat erfaßten Regressionen von Schulleistungs-Posttest-Werten auf Prätest-Werte näher analysiert. In Anlehnung an Burstein (1980); Burstein, Linn und Capell (1978) wurde angenommen, daß diese Unterschiede den Effekt unterschiedlicher Unterrichtsprozesse reflektieren können. So ist es vorstellbar, daß in Klassen mit flacher Regressionsgerade (LoSlope-Klassen) die ungenügende Vorhersage der Posttest-Werte durch die Eingangsleistung etwa dadurch zustandekommt, daß hier remedialer Unterricht vorherrscht, d.h. daß ungewöhnlich viel Unterrichtszeit auf die Unterstützung der weniger fähigen Schüler verwendet wird. Das Gegenteil sollte für Schulklassen mit steilen Regressionsgeraden (HiSlope-Klassen) zutreffen, in denen ein größeres Angebot an Lerngelegenheiten existiert, in denen also bessere Schüler ihre Fähigkeiten voller ausschöpfen können und weniger fähige Schüler gegenüber ihren Klassenkameraden stärker abfallen.

Aus insgesamt 113 sechsten Schulklassen (Hauptschulen des Rhein-Neckar-Kreises) wurden zwei Extremgruppen mit ungewöhnlich hohen bzw. niedrigen Prätest-Posttest-Regressionskoeffizienten (Mittelwerte: .90 vs. .41) ausgewählt. Die Prä-Posttest-Differenz zwischen den LoSlope- und HiSlope-Klassen erwies sich sowohl statistisch als auch praktisch als signifikant ($p = .001$; $w^2 = .80$). Insgesamt nahmen 195 Schüler und acht Mathematiklehrer an der Untersuchung teil.

Schritte der Datenanalyse: Zur Analyse des Schulleistungsmodells wurde auf das Computerprogramm LISREL IV (Jöreskog und Sörbom, 1978) zurückgegriffen. Es wurde eine sequentielle Strategie benutzt, um die spezifischen Eigenschaften der exogenen und endogenen Modellteile zu analysieren. Zunächst wurde geprüft, ob sich das angenommene Meßmodell für die beiden Gruppen aufrecht erhalten ließ. Für den Fall, daß das Simultanmodell zurückgewiesen wurde, sollten für beide Gruppen separate Schätzungen

des Meß- und Strukturmodells vorgenommen werden. In einem zweiten Schritt sollte dann untersucht werden, ob das für den endogenen Modellteil formulierte Längsschnittmodell die statistischen Eigenschaften eines Simplex-Modells aufweist. Der letzte Schritt betraf dann die Schätzung und Testung des vollständigen Modells (d.h. Meß- und Strukturmodells).

Ergebnisse

Testung des Meßmodells: Im ersten Analyseschritt wurde untersucht, ob die Faktorenstruktur für beide Schulklassen-Typen stabil ausfiel. Es wurde davon ausgegangen, daß das Meßmodell dann als zuverlässig gelten kann, wenn sowohl die Faktorenladungen als auch die Faktor-Interkorrelationen für beide Gruppen das gleiche Grundmuster aufweisen. Das Ergebnis einer simultanen konfirmatorischen Faktorenanalyse (COFAMM) zeigte jedoch, daß das zugrundegelegte Modell mit den Daten nicht kompatibel war ($\chi^2 = 143.61$ mit 59 Freiheitsgraden: $p \leq .01$). In der Folge wurden für jede Gruppe separate Berechnungen durchgeführt, da keine Anhaltspunkte für ein identisches Meßmodell vorlagen. Um zu der ökonomischsten Lösung zu kommen, wurde eine von Jöreskog (1978) vorgeschlagene exploratorische Vorgehensweise benutzt, mit der sich eine sog. "bestmöglich passende Einfachstruktur" für eine vorgegebene Faktorenzahl auffinden läßt. Während sich für die HiSlope-Klassen eine recht gute Datenanpassung erreichen ließ ($\chi^2 (17) = 25.04$; $p = .09$), mußte das angenommene Meßmodell für die LoSlope-Klassen zurückgewiesen werden ($\chi^2 (17) = 40.63$; $p \leq .01$). Wie die Inspektion der ersten Ableitungen der Anpassungsfunktion ergab, war insbesondere die Annahme unkorrelierter Meßfehler für beide Subgruppen sehr problematisch. Nach Sörbom (1975) sind Meßfehler-Korrelationen dann plausibel, wenn ähnliche Testverfahren zur Messung des gleichen Merkmals verwendet werden. Dieses Argument ließ sich auch auf die vorliegen-

de Untersuchung anwenden, bei der sechs Fähigkeitstests innerhalb einer kurzen Zeitspanne vorgenommen wurden. Von daher ließ sich eine Umformulierung des ursprünglichen Meßmodells für beide Gruppen rechtfertigen, bei der Meßfehler-Korrelationen für die Parameter mit den numerisch größten ersten Ableitungen im Ausgangsmodell unterstellt wurden. Für das modifizierte Modell konnte für jede Gruppe eine deutlich bessere Datenanpassung erreicht werden. Die Unterschiede in den χ^2 -Werten zwischen dem Ausgangsmodell und der modifizierten Version waren für beide Klassen-Typen auf dem 1%-Niveau signifikant (19.5 mit 7 Freiheitsgraden für die HiSlope-Klassen, 27.87 mit 5 Freiheitsgraden für die LoSlope-Klassen). Wiederum ließ sich für die HiSlope-Klassen die weitaus bessere Datenanpassung erzielen, wenn auch die Lösung für die LoSlope-Klassen noch als akzeptabel gelten muß.

Testung des Strukturmodells: Im zweiten Analyseschritt wurde die Annahme überprüft, ob der endogene Modell-Anteil, (d.h., der längsschnittliche Verlauf der Mathematikleistung) für beide Gruppen als Simplex formuliert werden kann. Da sich die Formulierungen des Quasi-Markovschen Simplex-Modells bei Jöreskog (1970) und Jöreskog (1979) geringfügig unterscheiden, wurden beide Modellspezifikationen geschätzt. In beiden Fällen ergaben sich jedoch identische Lösungen. Als erstes Ergebnis ließ sich festhalten, daß die angenommene Simplex-Struktur für die LoSlope-Klassen zurückgewiesen werden mußte ($\chi^2 = 4.81$ mit einem Freiheitsgrad; $p \leq .05$). Die Inspektion der ersten Ableitungen der Anpassungsfunktion ergab hohe Werte für die Parameter der jeweils nicht benachbarten Variablen, die auf Null gesetzt ('gefixt') worden waren. Es zeigte sich demnach, daß die Simplex-Annahme, der zufolge die partielle Korrelation zwischen zwei nicht benachbarten Variablen x_1 und x_3 bei Kontrolle von x_2 gleich Null ist, für diese Stichprobe nicht zutraf. Demgegenüber war das Simplex-Modell für die HiSlope-Klassen durchaus ak-

zeptabel ($\chi^2 = .07$ bei einem Freiheitsgrad; $p = .78$). Die in Modellen des akademischen Leistungszuwachses wiederholt gefundene Simplex-Struktur ließ sich demnach in der vorliegenden Untersuchung lediglich für die HiSlope-Klassen bestätigen. Da für den Fall von vier Meßzeitpunkten lediglich die beiden 'inneren' Variablen identifiziert sind (vgl. Jöreskog, 1979; Werts et al., 1977), waren Reliabilitätsschätzungen nur für diese beobachteten Variablen möglich (.54 und .43).

Vollständige Modelltests: Der letzte Analyseschritt betraf die Schätzung und Testung des vollständigen Modells. Das aus den COFAMM-Analysen resultierende modifizierte Meßmodell wurde dabei für beide Gruppen mit den Strukturkomponenten verknüpft, um die ökonomischste Lösung zu bestimmen. Erste Analysen zeigten, daß das theoretisch postulierte Modell in zwei Punkten modifiziert werden mußte: es erwies sich einmal als notwendig, die Simplex-Annahme für das endogene Submodell aufzugeben, nachdem das vollständige Modell für beide Gruppen geschätzt worden war. Die Hinzunahme der (exogenen) Fähigkeits- und Instruktionskomponenten bewirkte eine beträchtliche Veränderung des endogenen Submodells. Die direkten kausalen Pfadkoeffizienten zwischen den vier Meßzeitpunkten wiesen nun beträchtlich geringere numerische Werte auf (verglichen mit der oben beschriebenen Simplex-Lösung), und es mußten zusätzliche Pfade eingeführt werden, die die nicht benachbarten Leistungsmessungen miteinander in Beziehung setzten, d.h., es wurden alle Parameter mit signifikanten ersten Ableitungen freigesetzt (vgl. Abb. 1 und 2).

Weiterhin ließen sich nun für die ersten zwei bzw. drei Meßzeitpunkte signifikante kausale Effekte der beiden exogenen Konstrukte nachweisen: die entsprechenden Pfadkoeffizienten für den letzten Meßzeitpunkt konnten ohne jeden Informationsverlust eliminiert werden. Es sollte allerdings in diesem Zusammenhang betont werden, daß

nur die offensichtlichsten Fehlspekulationen bei der Testung der Modellannahmen korrigiert wurden; wie schon Maruyama und McGarvey (1980) dargelegt haben, wirft die Modifikation eines gegebenen theoretischen Modells insofern Probleme auf, als eine übermäßige Anpassung des Modells an (zufällige) Gegebenheiten der verfügbaren Stichprobe nicht ausgeschlossen werden kann.

Die Lösungen für beide Klassen-Typen sind in den Abb. 1 und 2 wiedergegeben. Es sollte herausgestellt werden, daß eine akzeptable Datenanpassung lediglich für die HiSlope-Klassen erreicht werden konnte ($\chi^2 = 32.31$ mit 39 Freiheitsgraden; $p = .77$), während sich für die LoSlope-Klassen ein signifikanter χ^2 -Wert (72.80 mit 36 Freiheitsgraden, $p \leq .01$) ergab, aus dem zu folgern ist, daß das Modell für diese Gruppe keine plausible Abbildung des zugrundeliegenden kausalen Prozesses darstellt.

Um die graphische Wiedergabe der Modelle übersichtlicher zu gestalten, wurden die Meßfehler und ihre Korrelationen in den Abb. 1 und 2 nicht wiedergegeben und stattdessen separat in Tab. 1 aufgeführt. Es läßt sich aus Tab. 1 ablesen, daß die offensichtlichste Fehlspezifikation des für die LoSlope-Klassen spezifizierten Modells die Beziehung zwischen den beiden Indikatoren des Unterrichtsprozesses betrifft. So deutet das negative Vorzeichen für den Meßfehler bei der zweiten Prozeßvariable darauf hin, daß die beiden ausgewählten Prozeßmerkmale für diese Subgruppe keinen gemeinsamen Faktor konstituieren können.

Der Vergleich der in den Abb. 1 und 2 wiedergegebenen Modelle legt die Annahme nahe, daß die hauptsächlichen Unterschiede in den Schätzparametern des Meßmodells liegen, während das Strukturgleichungsmodell in beiden Gruppen eine gewisse Ähnlichkeit aufweist. Es wird aufgrund der schlechten Anpassung des Modells für die LoSlope-Klassen allerdings kein Versuch gemacht, diese Befunde im Detail zu diskutieren. Die weiteren Überlegungen be-

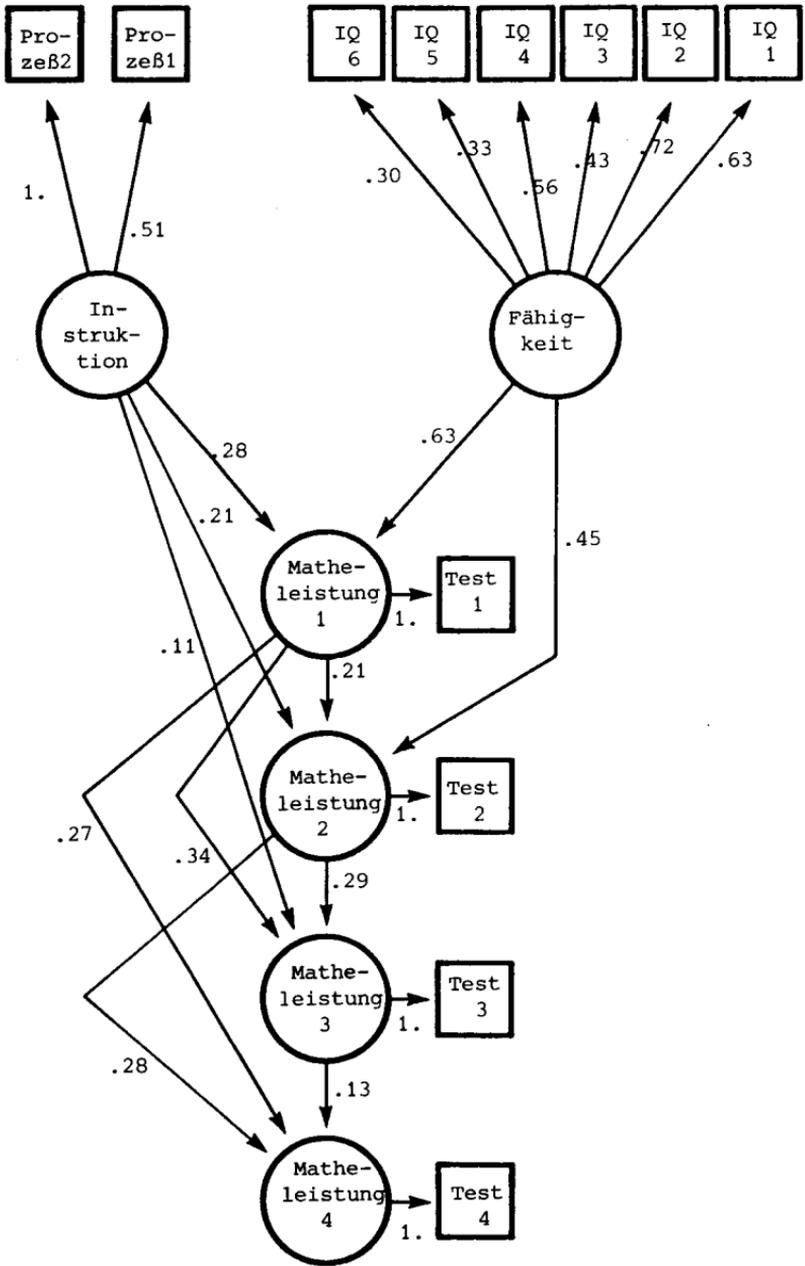


Abbildung 1: Bestangepasstes LISREL-Modell für die Entwicklung der Mathematikleistung in HiSlope-Klassen.

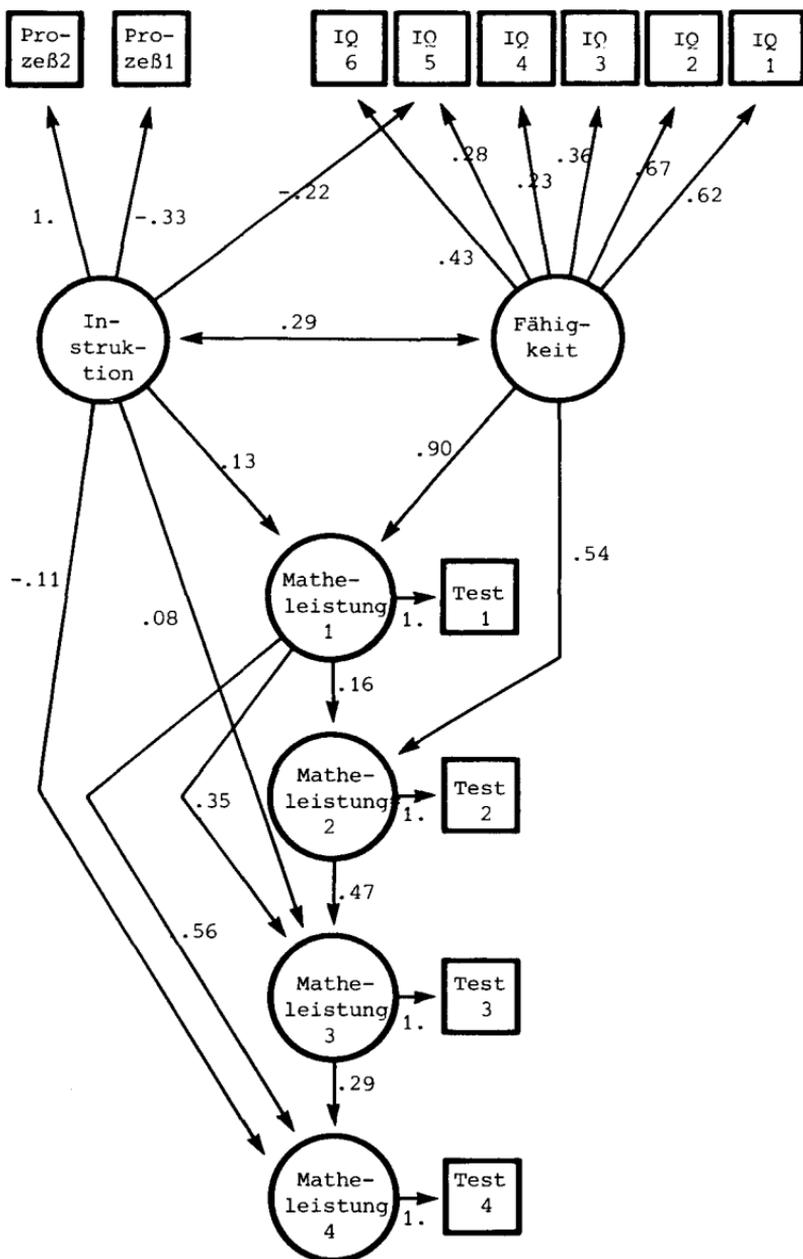


Abbildung 2: LISREL-Modell zur Entwicklung der Mathematikleistung in LoSlope-Klassen (nicht angepaßt).

Tabelle 1: Geschätzte Meßfehler¹⁾ und ihre Korrelationen aus den Meßfehlern der LISREL-Analysen für HiSlope und LoSlope-Klassen

<u>HiSlope-Klassen</u>	IQ- test 1	IQ- test 2	IQ- test 3	IQ- test 4	IQ- test 5	IQ- test 6	Pro- zeß 1	Pro- zeß 2
IQ-Test 1	.55							
IQ-Test 2	.00	.49						
IQ-Test 3	.16	.15	.82					
IQ-Test 4	.00	.00	.00	.64				
IQ-Test 5	-.11	.00	.00	.18	.89			
IQ-Test 6	.00	-.11	.00	.00	.00	.88		
Prozeß 1	.00	.00	.00	.00	.00	.00	.71	
Prozeß 2	.00	.00	.00	.00	.00	.00	.00	.00

<u>LoSlope-Klassen</u>	IQ- test 1	IQ- test 2	IQ- test 3	IQ- test 4	IQ- test 5	IQ- test 6	Pro- zeß 1	Pro- zeß 2
IQ-Test 1	.61							
IQ-Test 2	.00	.55						
IQ-Test 3	.13	.00	.88					
IQ-Test 4	.00	.00	.00	.94				
IQ-Test 5	.00	.00	-.14	.00	.93			
IQ-Test 6	.00	.00	.00	.00	.16	.80		
Prozeß 1	.00	.00	.12	-.22	-.35	-.18	.86	
Prozeß 2	-.17	.00	.00	.00	.00	.11	.00	-.03

¹⁾ Die Meßfehler sind in der Hauptdiagonalen wiedergegeben.

ziehen sich stattdessen auf den Vergleich des theoretisch postulierten mit dem empirisch bestätigten Modell für die HiSlope-Klassen.

Es gibt fundierte empirische Anhaltspunkte dafür, daß das theoretisch postulierte Modell die Veränderung von Schulleistungen im Fach Mathematik beschreiben und erklären kann. Die für das Meßmodell angenommene Einfachstruktur ließ sich für diese Stichprobe bestätigen. Weiterhin

konnte hier die Annahme empirisch belegt werden, daß die beiden exogenen Konstrukte unkorreliert sind. Demnach kann gefolgert werden, daß Parameter der Schülerfähigkeit und des Unterrichts die Mathematikleistung in den Hi-Slope-Klassen unabhängig voneinander beeinflussen. Was die numerische Ausprägung der Strukturkoeffizienten angeht, scheinen die Fähigkeits-Indikatoren im Vergleich mit den Unterrichtsmerkmalen für die beiden ersten Meßzeitpunkte mehr Vorhersagegewicht aufzuweisen. Demgegenüber tragen die beiden exogenen Faktoren jedoch nichts zur Erklärung der Leistungsvariation bei, die während der beiden letzten Testzeitpunkte beobachtet wurde. Dies ist höchstwahrscheinlich darauf zurückzuführen, daß die Fähigkeits-Indikatoren und Variablen des Instruktionsprozesses lediglich zum ersten Meßzeitpunkt erhoben wurden. Aufgrund der unterschiedlichen Zielsetzungen der Primärstudie (vgl. Treiber, 1980) wurden für diese Variablen keine weiteren Daten gesammelt.

Abschließende Bemerkung: Für die zukünftige Forschung lassen sich aus der vorliegenden Studie vor allem zwei Fragestellungen ableiten. Zum einen sollten für die Beschreibung und Erklärung von Leistungsänderungen über die Zeit hinweg elaboriertere Struktur- und Meßmodelle als die hier herangezogenen allzu einfachen Typen entwickelt werden. Es scheint notwendig, sowohl eine umfangreichere Datensammlung für exogene und endogene Modellvariablen wie auch eine kompliziertere (etwa nicht-lineare, nicht-additive und nicht-rekursive) Spezifikation ihre Interrelationen vorzunehmen.

Zum zweiten hat sich gezeigt, daß der konventionelle Erklärungsansatz insbesondere für die LoSlope-Klassen (also Klassen mit schlechter Vorhersagbarkeit von Posttest- aus Prätest-Werten) ungenügend blieb. Für Schulklassen dieses Leistungsstruktur-Typus sind alternative Erklärungsmodelle noch nicht verfügbar. Nachdem sich

die hier ausgewählten Instruktionsvariablen für die Lo-Slope-Klassen als inadäquat herausgestellt haben, erscheint es notwendig, unterschiedliche und/oder zusätzliche Maße zu konstruieren, die den Unterrichtsprozeß in diesen Klassen besser beschreiben.

Die Befunde der vorliegenden Studie unterstreichen die Annahme einer lediglich "lokalen" Anwendbarkeit von Schulleistungsmodellen (vgl. Snow, 1977b) und lassen die vor langer Zeit initiierten Versuche, generelle oder universale Erklärungsmodelle des schulischen Lernens zu entwickeln, als wenig erfolgversprechend erscheinen. Es scheint statt dessen ratsamer zu sein, den zeit- und kontextgebundenen Verlauf von Schulleistungsprozessen in vorgegebenen Lehr-Lernbedingungen zum Schwerpunkt zukünftiger Unterrichtsforschung zu machen.