

Kausalmodelle zur Beschreibung und Erklärung von Rechtschreibleistungen in der Grundschule – Eine Sekundäranalyse¹

Wolfgang Schneider

Zusammenfassung: Vorgestellt wird ein Versuch, die in der einschlägigen Literatur postulierte multifaktorielle Bedingtheit von Rechtschreibleistungen in der Grundschule über ein angemessenes methodisches Design zu prüfen. Zentral ist dabei die Frage, ob sich identische Kausalstrukturen für beginnende und geübte Rechtschreiber (Schüler der zweiten vs. vierten Klasse) nachweisen lassen.

Konventionelle Verfahren der Kausalanalyse hatten sich in der Primärstudie (Schneider 1980) als unökonomisch und wenig aussagekräftig erwiesen, so daß in der Sekundäranalyse auf eine flexiblere Prozedur zurückgegriffen wird. Mit diesem Verfahren zur Analyse von Strukturgleichungssystemen (LISREL) ist es möglich, für die Gruppe der Zweit- und Viertkläßler Modelle zu entwickeln und zu überprüfen, die mit den Ausgangsdaten kompatibel sind. Als wesentliches Ergebnis zeigt sich, daß die theoretisch postulierte Bedingungsstruktur nur für die Viertkläßler (eingeschränkt) bestätigt werden kann, während für die Schüler der zweiten Klassenstufe ein grundlegend verschiedenes Muster resultiert.

Testing causal models of spelling ability – A secondary analysis

Summary: In this study an attempt is made to find a causal pattern describing and explaining spelling in elementary school children. The main question concerns the structural equivalence of this pattern for beginning and advanced spellers (second vs. fourth graders). Whereas in a primary analysis (Schneider 1980) the limitations of traditional path analytic procedures made it impossible to handle the problem adequately, the advantages of a more flexible program system (LISREL) can be demonstrated in this secondary analysis. Parsimonious and rather elegant causal models are constructed and tested for both age groups, showing different causal patterns for beginning and advanced spellers. Only in case of the fourth graders the resulting causal structure has some similarity to the model derived from theoretical assumptions.

Einleitung

Obwohl in den letzten Jahren insbesondere im Rahmen der sog. Legasthenieforschung eine Vielzahl von Studien zur Erklärung von Rechtschreibleistungen und -defiziten durchgeführt worden sind, läßt sich aus den Befunden kein auch nur einigermaßen klares Bild rekonstruieren.

Sieht man einmal davon ab, daß eine Reihe von Untersuchungen zur Legasthenieproblematik schon durch die Art der Stichproben-Ziehung lediglich eingeschränkte Gültigkeit beanspruchen können (s. die Kritik von SCHLEE, 1976 am Legasthenie-Konzept), so liegt die fundamentale Schwäche annähernd aller Studien in ihrem restriktiven experimentellen Design begründet: es wurden von vornherein lediglich einige wenige Indikatoren ausgewählt,

deren Relevanz für die Rechtschreibleistung als gesichert galt (vgl. die Literaturübersicht bei ANGERMAIER, 1974; VALTIN, 1974). Es kann von daher kaum verwundern, daß zwar einerseits wahre Kataloge angeblich rechtschreibrelevanter Indikatoren erstellt wurden, es andererseits aber aufgrund der vielen widersprüchlichen Einzelbefunde kaum möglich war, hier eine befriedigende Integration zu leisten.

Bei der Durchsicht dieser Arbeiten fällt nun auf, daß viele Autoren relativ unabhängig von ihrem eigenen experimentellen Vorgehen der Auffassung sind, daß die Rechtschreibleistung als mehrdimensional bzw. ‚multikonditional‘ oder ‚polyfaktoriell‘ (s. ANGERMAIER, 1974) verursacht anzusehen ist. Die wenigen multivariaten Analysen in diesem Bereich blieben aber darauf beschränkt, multiple Indikatoren der Intelligenz bzw. der verbalen Fertigkeit faktorenanalytisch auf wenige Dimensionen zu reduzieren (vgl. ANGERMAIER, 1973).

Um nun die Hypothese der multifaktoriellen Determination der Rechtschreibleistung adäquater prüfen zu können, wurde von SCHNEIDER (1980) ein Kausalmodell konstruiert, dessen wesentliche Bestandteile kurz skizziert werden sollen; auf ihnen baut auch die hier näher darzustellende Sekundäranalyse auf.

1. Aufbau der Primäranalyse

In der genannten Vorläuferstudie sollte die Kausalanalyse lediglich illustrativen Zwecken dienen, wobei die Nützlichkeit eines theoretisch plausiblen Modells zur Erklärung und Vorhersage von Rechtschreibleistungen zu demonstrieren war. Da das Modell einigermaßen überschaubar bleiben sollte, wurden lediglich diejenigen Indikatoren aus einem größeren Variablenpool berücksichtigt, die sich schon in vorgeschalteten multiplen Regressionsanalysen zur Vorhersage der Rechtschreibleistung in unterschiedlichen Diktatsituationen (s.u.) bewährt hatten (vgl. SCHNEIDER, 1980, S. 160ff.). In das exemplarisch spezifizierte Kausalmodell (s. Abb. 1) gingen demnach Variablen ein, deren Relevanz für die Rechtschreibperformanz in zahlreichen Studien dokumentiert wurde.

Es waren dies im einzelnen die demographischen Merkmale Geschlecht, Alter und Schichtzugehörigkeit, die nichtkognitiven Variablen Konzentrationsfähigkeit (erfaßt über den Test d2 von BRICKENKAMP), Gedächtnis (repräsentiert über zwei eigens zusammengestellte Aufgaben zur Erfassung der Gedächtnisspanne und der Reproduktion von kategorisierbarem Wortmaterial) und Rechtschreibmotivation. Ergänzt wurde der Variablensatz durch die kognitiven Merkmale Intelligenz (CFT nach CATTELL & WEISS), Leseverständnis und Wortschatz (zwei Subtests aus dem Allgemeinen Schulleistungstest von FIPPINGER, 1967) sowie durch vier selbstentwickelte Diktatproben, in denen die orthographischen Merkmale Wortschwierigkeit und -frequenz systematisch kombiniert waren: diese Kriteriumsvariablen setzten sich also

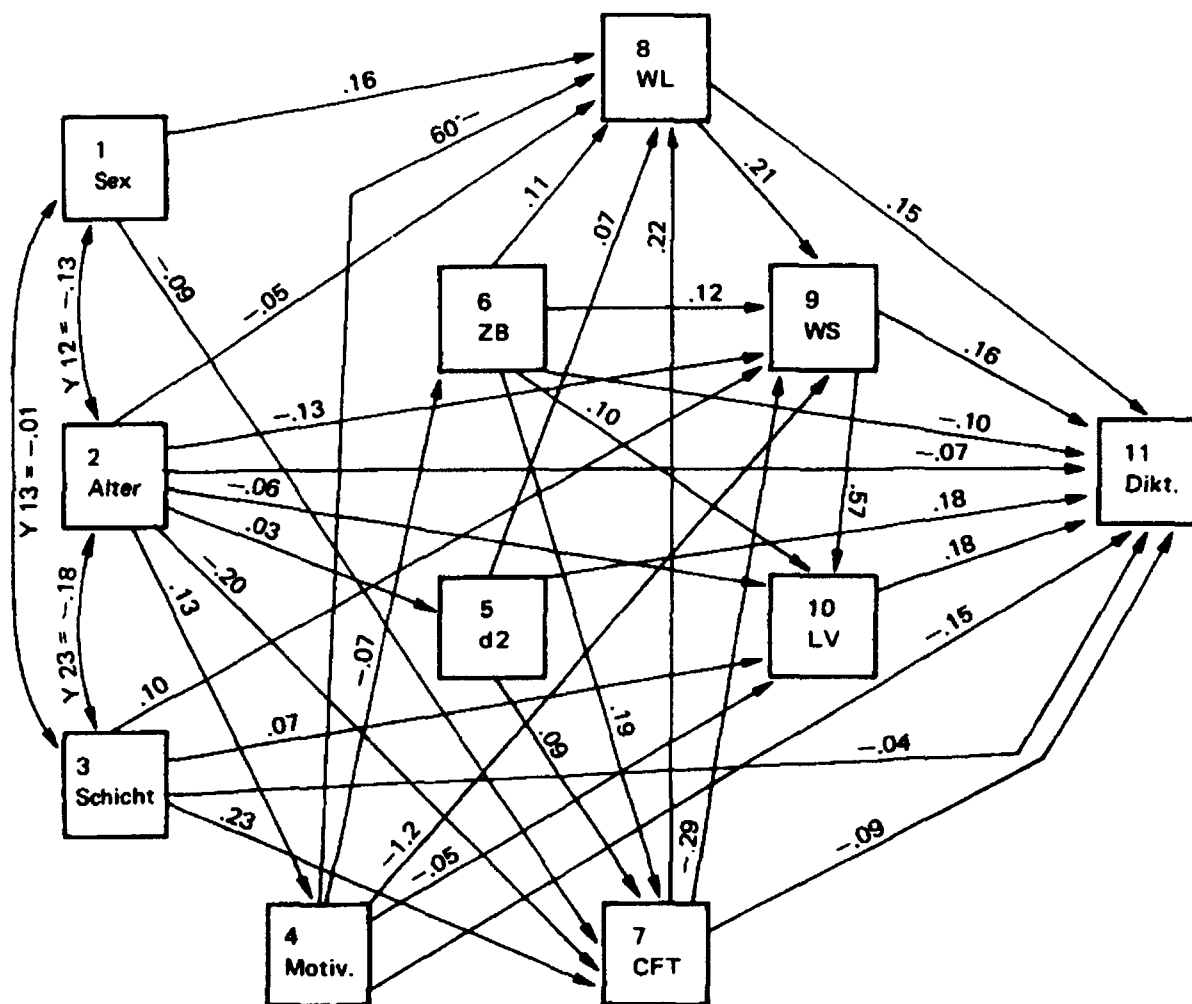


Abb. 1: Ergebnisse eines relativ komplexen rekursiven Pfadmodells, das lediglich auf manifesten Indikatoren aufbaut (entnommen aus SCHNEIDER 1980)

aus häufig/leichtem (HL), häufig/schwerem (HS), selten/leichtem (SL) und selten/schwerem (SS) Wortmaterial zusammen (zur detaillierten Beschreibung aller Indikatoren vgl. SCHNEIDER, a.a.O., S. 92ff.).

Die Auswahl der Personenstichprobe war eng mit der Frage verknüpft, ob die multiple Determination der Rechtschreibleistung über alle Klassenstufen der Grundschule hinweg als gleich anzusehen ist oder ob sich die Bestimmungsmuster für beginnende und fortgeschrittene Rechtschreiber deutlich unterscheiden. In den meisten Untersuchungen zur Rechtschreibproblematik waren es die Schüler der letzten Grundschulklassen, die als Versuchspersonen fungierten; es stand also nicht von vornherein fest, daß die dort berichteten Befunde auch für jüngere Probanden Geltung beanspruchen konnten. Es wurden deshalb in der Primärstudie sowohl Zweitkläßler ($N = 226$) als auch Viertkläßler ($N = 280$) berücksichtigt, um die Verallgemeinerungsmöglichkeiten der Ergebnisse zu prüfen.

Obwohl dem in Abb. 1 dargestellten Pfadmodell ein relativ sparsamer Variablensatz zugrunde gelegt worden war, lassen sich bei der Darstellung der

Befunde die Grenzen der üblicherweise verwendeten pfadanalytischen Prozeduren (im vorliegenden Fall das Programm YPFAD von ZIEGLER, 1972) erkennen: um die vorhandene Information voll ausschöpfen zu können, sind in beiden Klassenstufen getrennte Analysen für alle vier Diktatsituationen erforderlich. Da jedes resultierende Pfaddiagramm eine beträchtliche Komplexität aufweist, fallen Vergleiche der acht resultierenden Modelle per visuelle Inspektion äußerst unbefriedigend aus. Um diesen gravierenden Nachteil der Primäruntersuchung überwinden zu können, wurde nach einer eleganteren und gleichzeitig sparsameren Lösung des Problems gesucht, die schließlich auch in der Allgemeinen Methode für die strukturelle Analyse von Kovarianz- und Korrelationsmatrizen von JÖRESKOG (s. JÖRESKOG, 1977 u. 1978; JÖRESKOG & SÖRBOM, 1979) gefunden wurde.

2. Schritte der Sekundäranalyse²

Anwendungen des von JÖRESKOG & SÖRBOM (1978) entwickelten Computerprogramms LISREL IV (zur Analyse von *Linear Structural Relationships*) auf sozialwissenschaftliche Probleme sind in publizierter Form erst in den letzten Jahren im angloamerikanischen Raum bekanntgeworden; die vielfältigen Vorzüge und Verdienste der Methoden wurden hier in sozialpsychologischen Arbeiten (vgl. BENTLER & HUBA, 1979; BENTLER & SPECKART, 1979; KERLINGER, 1980), aber auch in entwicklungspsychologischen (CUNNINGHAM & BIRREN, 1980; McARDLE & HORN, 1980; OLSSON & BERGMAN, 1977; WEEKS, 1980) bzw. pädagogisch-psychologischen Studien (MARUYAMA & McGARVEY, 1980; WERTS et al., 1977 u. 1978) eindrucksvoll demonstriert.

Da im deutschsprachigen Bereich Beschreibungen des Ansatzes lediglich in unpublizierten Arbeiten (GRAFF, 1978; SCHMIDT, 1977) bzw. noch nicht erschienenen Manuskripten (MÖBUS, 1981) vorliegen, scheint es sinnvoll, die wesentlichen Eigenschaften der Methode kurz zu skizzieren.

2.1 Beschreibung des LISREL-Ansatzes

Der hauptsächliche Vorteil dieser Strukturgleichungs-Methode besteht zweifellos in ihrer Flexibilität. Es lassen sich mit LISREL unterschiedliche Formen von Kausalmodellen testen, wobei das in Abb. 1 wiedergegebene traditionelle Strukturgleichungs-System mit direkt beobachteten Indikatoren ebenso subsumiert werden kann wie auch komplexere Modelle, in denen latente Variablen als ‚Ursachen‘ für gemessene Merkmale betrachtet und zusätzlich verankert werden.

LISREL setzt sich aus zwei Teilen zusammen: während das *Meßmodell* die Beziehungen zwischen den beobachteten Variablen und den latenten Größen (hypothetischen Konstrukten) definiert, dient das *Strukturgleichungsmodell*

dazu, die Kausalbeziehungen zwischen den latenten Variablen festzulegen. Die algebraische Formulierung in der LISREL-Terminologie sieht dabei folgendermaßen aus:

$$(1) \text{ Meßmodell: } \begin{aligned} y &= \lambda_y \eta + \epsilon \\ x &= \lambda_x \xi + \delta \end{aligned}$$

$$(2) \text{ Strukturmodell: } B\eta = \Gamma\xi + \zeta$$

Bei der Formulierung des Meßmodells wird also davon ausgegangen, daß die manifesten exogenen (d.h. im Modell nicht weiter erklärbaren) x-Indikatoren und die manifesten endogenen (d.h. näher zu erklärenden) y-Indikatoren als lineare Funktion der mit λ_x bzw. λ_y gewichteten latenten exogenen (ξ) bzw. endogenen (η) Variablen und des zugehörigen Meßfehlers (ϵ bzw. δ) aufzufassen sind. Das auf der Ebene der latenten Variablen ansetzende Strukturmodell erklärt die endogenen latenten Größen über die vorausgehenden exogenen latenten Variablen, wobei der Spezifikationsfehler (ζ) die Güte der Prädiktion bestimmt.

Dieser Strukturgleichungs-Ansatz geht in mehrerer Hinsicht über die Möglichkeiten der multiplen Regressionsanalyse hinaus: neben der Unterscheidung von beobachteten Indikatoren und nicht gemessenen hypothetischen Konstrukten wirkt sich dies besonders in der Behandlung des Fehler-Konzepts aus. Man ist nun nicht mehr (wie bei der multiplen Regression) dazu gezwungen, von fehlerfrei erhobenen Prädiktor-Maßen auszugehen: das Modell erlaubt es vielmehr, zwischen Meßfehlern und Gleichungsfehlern zu unterscheiden. Die Flexibilität des Programms läßt sich daran verdeutlichen, daß sowohl rekursive als auch nicht-rekursive (Feedback-) Modelle geschätzt und getestet werden können, daß sowohl zufällige als auch (im Fall von Wiederholungsmessungen) systematische Meßfehler angenommen werden können und es schließlich möglich ist, identische Kausalmodelle simultan für zwei oder mehrere Gruppen zu schätzen, wobei einige oder alle Parameter für die jeweiligen Gruppen als gleich angenommen werden.

Die hauptsächlichsten Problempunkte des Berechnungsverfahrens betreffen die Identifikation, die Schätzung und die Testung des jeweils konstruierten Kausalmodells, zu denen eine kurze Erläuterung notwendig scheint.

a) Das Identifikationsproblem

Wie JÖRESKOG (1977, 1978) herausgestellt hat, hängt die Identifizierbarkeit eines Modells davon ab, ob die Schätzparameter eindeutig durch Σ (die geschätzte Populations-Kovarianzmatrix) determiniert sind. Es lassen sich nun mehrere äquivalente Strukturen denken, die die gleiche Σ -Matrix generieren können. Ein Parameter wird dann als identifiziert bezeichnet, wenn er in allen äquivalenten Strukturen den gleichen Wert aufweist. Von der Identifikation eines Modells wird also dann ausgegangen, wenn alle Parameter identische Werte in verschiedenen Kausalmodellen erhalten, die identische beobachtbare Daten (Korrelationen) generieren (globale Identifikation). Lokale Identifikation wird dann angenommen, wenn lediglich einige Parameter identifiziert werden können (zur detaillierten Diskussion vgl. FISHER, 1966; WILEY, 1973). Im LISREL wird die Identifikation der Parameter implizit getestet; für den Fall, daß sich die Informationsmatrix invertieren läßt, wird Identifizierbarkeit des Modells unterstellt; ist sie dagegen singular, muß davon ausgegangen werden, daß das Modell nicht identifiziert ist (s. JÖRESKOG, 1978; JÖRESKOG & SÖRBOM, 1978).

b) Modellschätzung und -testung

Das Problem bei der Generierung von Schätzungen für alle unbekannt Parameter des Modells besteht darin, die Populations-Kovarianzmatrix Σ an die beobachtete Kovarianzmatrix S anzupassen. Während prinzipiell mehrere Schätzmethoden (Methode der kleinsten Quadrate in der ungewichteten und verallgemeinerten Variante, Maximum-Likelihood-Schätzung) verfügbar sind, beschränkt sich LISREL auf die ‚full information maximum likelihood‘-Schätzung (FIML), bei der für die Bestimmung eines Parameters alle Informationen über alle übrigen Parameter benutzt werden. Die Schätzungen sind unter der Annahme der multivariaten Normalverteilung asymptotisch effizient.

Die Güte eines Modells läßt sich nun daran ablesen, wie exakt es die vorgegebenen Daten (Kovarianz- oder Korrelationsmatrix) reproduzieren kann. Die Validität eines Kausalmodells kann dabei (bei großen Stichproben) über die sog. ‚likelihood ratio-technique‘ getestet werden, wobei die Prüfgröße (goodness-of-fit statistic) χ^2 berechnet wird, deren Freiheitsgrade der Differenz zwischen bekannten und unbekannt Parametern entsprechen. Es läßt sich damit die Modell-Anpassung an die Daten bewerten; getestet wird hierbei gegen die allgemeine Hypothese, daß Σ irgendeine beliebige positiv definierte Matrix ist. Ein im Verhältnis zu den Freiheitsgraden hoher χ^2 -Wert deutet darauf hin, daß das hypostasierte Kausalmodell keine plausible Repräsentation des Kausalprozesses bieten kann. In diesem Fall stehen mehrere Möglichkeiten offen, um das inadäquate Modell zu modifizieren: es können beispielsweise diejenigen Parameter weggelassen werden, deren Schätzungen im Vergleich zu ihrem Standardfehler äußerst gering ausfallen (s. BENTLER, 1980). Weiterhin kann eine Inspektion der ersten Ableitung der Anpassungs-Funktion für die ‚gefixten‘ (als bekannt gesetzten) Parameter Wege aufzeigen, wie das Modell verbessert werden kann. Hohe Werte deuten hier darauf hin, daß die besagten Parameter besser frei zu schätzen sind. Schließlich läßt sich über die Analyse der Residuen, d.h. die Diskrepanzen in den Korrelationsmatrizen für die beobachteten und geschätzten Werte die Anpassungsgüte näher bestimmen; Diskrepanzen $> .10$ legen in der Regel eine Modell-Modifikation nahe.

JÖRESKÖG (1978) macht darauf aufmerksam, daß die χ^2 -Werte vorsichtig zu interpretieren sind, da sie beispielsweise in Abhängigkeit von der Stichprobengröße variieren. Es mag von daher in einigen Fällen interessanter sein, nicht auf die Absolutwerte zu rekurrieren, sondern die Differenzen in den χ^2 -Werten von mehreren konkurrierenden Modellen zu vergleichen.

2.2 Zur Verwendung des LISREL-Ansatzes in der vorliegenden Studie

Die LISREL-Prozedur sollte in der Sekundäranalyse gezielt dazu eingesetzt werden, das in Abb. 2 dargestellte Kausalmodell zu schätzen bzw. zu testen; letzteres war aus den wichtigsten Befunden der Primäranalyse abgeleitet worden. Die entscheidenden Vorteile des Ansatzes für die vorliegende Fragestellung sind darin zu sehen, daß hier ein einziges Modell all das enthält, was in der Primärstudie nur über acht unabhängige Schätzungen eruierbar war. Durch die Unterscheidung von beobachteten Indikatoren und latenten Konstrukten können die Variablen ökonomischer präsentiert werden, insbesondere aber auch die vier separaten Diktatproben in einem Rechtschreib-Faktor zusammengefaßt werden. Die Möglichkeit eines simultanen Gruppenvergleichs machte getrennte Berechnungen für die beiden Klassenstufen überflüssig.

Für das Verständnis von Abb. 2 ist wesentlich, daß die beobachteten Indikatoren jeweils als Rechtecke, die theoretischen Konstrukte als Kreise gezeichnet sind. Während der erste endogene Faktor nichtkognitive Merkmale wie

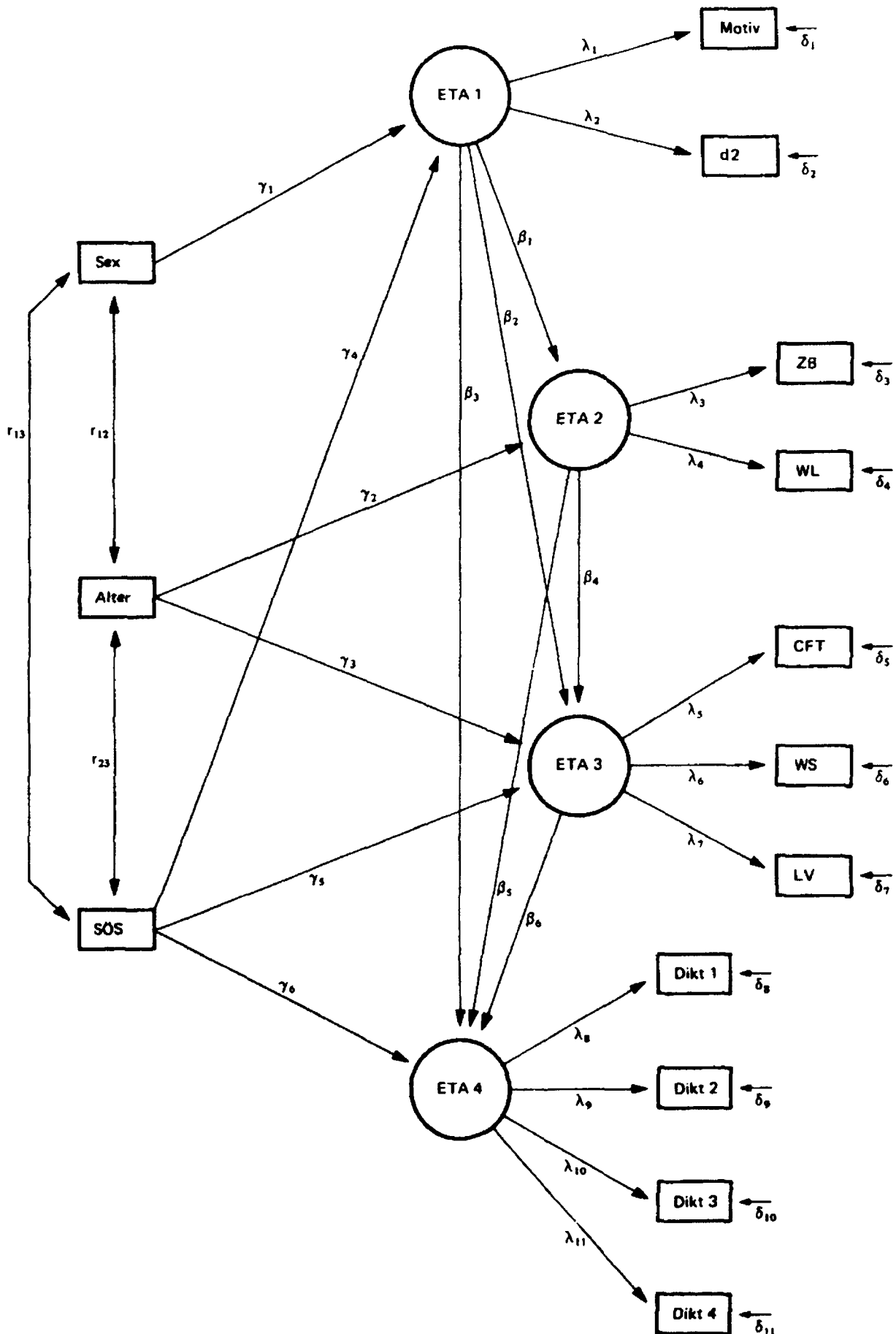


Abb. 2: Theoretisch plausibles LISREL-Modell zur Erklärung der Rechtschreibleistungen von Zweit- und Viertkläßlern

Konzentration (d2) und Rechtschreibmotivation (Motiv) einschließt, umfaßt der zweite die Gedächtnismerkmale wie Kapazität (ZB) und Gedächtnis für kategorisierbare Wortlisten (WL), wobei hier der eher strategiegeleitete Umgang mit Wortmaterial wesentlich scheint. Eta 3 repräsentiert Indikatoren der verbalen bzw. nichtverbalen intellektuellen Fähigkeit wie Wortschatz (WS), Leseverständnis (LV) und IQ (CFT).

Es wird nun angenommen, daß diese drei latente Variablen die resultierende Rechtschreibleistung (Eta 4) determinieren, die über die oben beschriebenen Diktat-Texte erfaßt wurde (die Durchnummerierung entspricht der oben berichteten Anordnung).

Es läßt sich nun leicht feststellen, daß die in Abb. 2 vorgestellte Konstruktion als Spezialfall des Allgemeinen Modells zu gelten hat, da für jede exogene latente Variable lediglich ein einziger Indikator verfügbar ist. Es wird also notwendig, x zu ‚fixen‘, d.h. die beobachteten mit den latenten Größen gleichzusetzen. Wenn dies auch JÖRESKOG & SÖRBOM (1978) zufolge in einigen Anwendungsbereichen durchaus sinnvoll sein mag, muß es im vorliegenden Fall als Aporie des Modells angesehen werden. Gerade für die Schichtvariable wäre es günstig gewesen, wenn mehrere Indikatoren verfügbar gewesen wären.

Wie schon erwähnt, wird bei LISREL das (rekursive) Kausalmodell auf der Ebene der latenten Variablen spezifiziert. Nichtkognitive Einflußgrößen wie Konzentration und Rechtschreibmotivation sollten sowohl Gedächtnis (Eta 2) wie auch Intellektuelle Fähigkeit (Eta 3) bzw. Rechtschreibkompetenz (Eta 4) direkt beeinflussen; es wurde weiter angenommen, daß Eta 2 die Konstrukte Eta 3 und Eta 4 determinieren sollte, während die Rechtschreibleistung wiederum direkt von der intellektuellen Ausstattung (Eta 3) abhängen sollte.

Für die exogenen latenten Größen wurden unterschiedlich viele Kausalbezüge vermutet. Das Geschlecht sollte lediglich den nichtkognitiven endogenen Faktor Eta 1 beeinflussen, während das Alter die Gedächtnisleistung und die intellektuelle Fähigkeit direkt determinieren sollte. Aufgrund des vielfach bestätigten Einflusses der Schichtvariable auf Merkmale wie Motivation, Gedächtnis und Intelligenz sowie die Rechtschreibleistung wurde davon ausgegangen, daß sie alle vier endogenen Faktoren direkt kausal beeinflussen würde.

Es wurde postuliert, daß das skizzierte Modell den kausalen Prozeß in beiden Altersgruppen widerspiegeln konnte. Wie schon erwähnt, bietet LISREL die Möglichkeit, das Modell für mehrere Gruppen simultan zu schätzen. Die Hypothese der strukturellen Äquivalenz des Modells kann nun unterschiedlich streng getestet werden, wobei entsprechend der Terminologie von BENTLER (1980) schwache, moderate und strenge Tests zu differenzieren sind. Während in einem schwachen Test lediglich das Modell und die Schätzprozedur für beide Gruppen gleich sein müssen, werden in moderaten Tests einige wichtige theoretische Parameter (z. B. die Faktorladungen) als identisch angenommen. In der strengen Testvariante geht man davon aus, daß alle Parameter der zweiten Stichprobe denen der ersten Stichprobe äquivalent sind.

2.2.1 Explorative Bestimmung des Meßmodells

In der vorliegenden Studie wurde zur Beurteilung der faktoriellen Invarianz eine Methode angewandt, die in etwa der von CUNNINGHAM & BIRREN (1980) eingesetzten Prozedur entsprach. In einem ersten Schritt wurden lediglich die Faktorladungs-Matrizen als gleich angenommen ($\lambda_{y1} = \lambda_{y2}$); für den Fall eines erfolgreichen Modelltests waren weitere Restriktionen vorgesehen, die insbesondere die faktorielle Struktur der Eta-Variablen und die Beziehungen zwischen latenten exogenen und endogenen Größen betrafen und den Tatbestand eines strengen Tests erfüllten. Der moderate Test bestand also darin, daß das Modell mit der Restriktion geschätzt wurde, daß alle λ -Elemente über beide Gruppen hinweg gleich sein sollten. Die Ergebnisse dieser ersten Analyse fielen negativ aus: der Algorithmus kam zu keiner konvergierenden Lösung (es wurden mehr als 250 Iterationen verzeichnet), und der resultierende χ^2 -Wert von 671.94 (bei 135 Freiheitsgraden) wies auf eine hochsignifikante Abweichung des Modells von den Ausgangsdaten hin. Die Beziehungen zwischen manifesten Indikatoren und latenten Variablen wiesen demzufolge keine äquivalente Struktur für beide Gruppen auf, so daß dieses leicht restringierte Modell zurückgewiesen werden mußte.

In einem zweiten Schritt wurde das Modell für beide Gruppen separat geschätzt. Wiederum deutete die χ^2 -Prüfgröße auf eine schlechte Modellanpassung an die Daten hin ($\chi^2_{(62)} = 369.98$ für die Zweitkläßler, $\chi^2_{(62)} = 137.99$ für die Viertkläßler). Das Modell wurde wiederum nicht völlig identifiziert und konnte deshalb nicht aufrechterhalten werden. Es schien, als ob das postulierte Muster der Faktorenladungen im Hinblick auf die verfügbaren Daten allzu restriktiv formuliert worden war.

Da die hier intuitiv ausgewählte ‚mentale Rotation zur Einfachstruktur‘ kein daten-kompatibles Meßmodell ergeben hatte, mußte nun nach einer Möglichkeit gesucht werden, das mit der Datenstruktur am besten zu vereinbarenden Meßmodell herauszufinden. Eine angemessene Lösung schien darin zu bestehen, das von JÖRESKOG (1978) beschriebene Verfahren zur Entdeckung der ‚best-fitting‘ (absolut besten) Einfachstruktur für eine jeweils vorgegebene Anzahl von Faktoren zu benutzen.

Diese explorative Vorgehensweise ist durch folgende Durchführungsschritte charakterisiert:

- (1) Berechnung einer explorativen Faktorenanalyse (EFAP) mit Varimax-Rotation;
- (2) Durchführung einer obliquen Procrustes-Rotation (da die Faktoren als korreliert angenommen werden);
- (3) Analyse der Promax-Faktormatrix; nachdem die größte Faktorladung für jede Spalte identifiziert worden ist, wird der Varimax-Faktor so rotiert, daß die übrigen Ladungen der betreffenden Zeile Null werden. Es resultiert eine schiefwinklige Lösung, bei der jeder Faktor-Vektor durch einen Test abgebildet wird;
- (4) Durchführung einer Konfirmatorischen Faktorenanalyse (COFAMM) mit den gleichen ‚gefixten‘ Null-Werten, wobei die gleichen von Null verschiedenen Faktorladungen resultieren, für diese aber nun Standardfehler verfügbar sind. Die Ladungen lassen sich nun zu ihren Standardfehlern in Beziehung setzen, insignifikante Schätzungen werden auf Null gesetzt;

Tabelle 1: Ergebnisse der EFAP-Analysen für die Zweit- und Viertkläßler (Promax-Lösungen)

a) Promax-Lösung (N = 4 Faktoren) für die Zweitkläßler (die höchsten Ladungen sind unterstrichen)

	F1	F2	F3	F4		
Motiv	-.036	.169	.326	.042		
d2	.137	.207	.064	.120		
ZB	.026	-.145	-.218	.175		
WL	.159	.130	-.124	.329		
CFT	-.136	.007	.032	<u>.653</u>	Faktor-Interkorrelationen	
WS	-.050	.032	-.605	.252		F1 F2 F3
LV	.032	.042	<u>-.864</u>	-.090	F2	-.21
Dikt.1	.008	.867	.006	-.040	F3	.24 -.29
Dikt.2	.001	<u>.905</u>	-.051	.053	F4	.10 .30 -.64
Dikt.3	-.780	.235	-.158	-.093	Anpassungsgüte:	
Dikt.4	<u>-1.052</u>	-.132	.089	.146	$\chi^2(17) = 9.91 (p = .91)$	

b) Promax-Lösung (N = 4 Faktoren) für die Viertkläßler

	F1	F2	F3	F4		
Motiv	.048	-.024	.019	<u>.519</u>		
d2	.117	.135	.351	.185		
ZB	-.114	.013	.181	.127		
WL	.049	-.113	<u>.772</u>	.001	Faktor-Interkorrelationen	
CFT	.046	.594	-.033	.156		F1 F2 F3
WS	-.107	<u>.806</u>	-.102	-.062	F2	-.30
LV	.034	.698	.025	-.081	F3	-.31 .54
Dikt.1	-.948	.005	-.130	.121	F4	.19 -.09 .10
Dikt.2	<u>-.957</u>	.053	.030	-.069	Anpassungsgüte:	
Dikt.3	.215	.136	-.018	-.361	$\chi^2(17) = 18.79 (p = .34)$	
Dikt.4	-.353	.076	.313	-.242		

(5) Durchführung einer zweiten COFAMM-Analyse mit diesem reduzierten Modell. Im allgemeinen wird hier eine Lösung resultieren, die nicht signifikant schlechter als die ursprüngliche ausfällt, dafür aber eine klare Einfachstruktur mit vielen Null-Ladungen aufweist.

Die mit dieser Prozedur erzielten interessantesten Ergebnisse sind in den Tab. 1-3 wiedergegeben, in denen die Lösungen für die Zweit- und Viertkläß-

ler gegenübergestellt sind. EFAP-Analysen wurden in beiden Altersgruppen für $N = 2$ bis $N = 4$ Faktoren durchgeführt; die theoretisch am meisten plausible vierfaktorielle Lösung erreichte dabei in beiden Stichproben die beste Datenanpassung. Während ein dreifaktorielles Modell für die Daten der Viertkläßler nicht aufrechterhalten werden konnte ($\chi^2_{(25)} = 51.68$, $p < .001$), war es mit den Daten der Zweitkläßler durchaus kompatibel ($\chi^2_{(25)} = 21.39$, $p < .67$). Eine bessere Anpassung ließ sich aber auch hier für die vierfaktorielle Lösung erzielen (vgl. Tab. 1).

Bemerkenswert scheint dabei, daß sich das Faktorenmodell für die Zweitkläßler besser an die Daten anpassen läßt als das für die Viertkläßler entwickelte Modell. Beträchtliche Unterschiede bestehen sowohl im Muster der Faktorenladungen als auch in der Interkorrelation der einzelnen Faktoren.

Tab. 2 gibt die Resultate des vierten Schritts der JÖRESKOG-Prozedur wieder: über die COFAMM-Lösung sind Standardfehler für die von Null verschiedenen Faktorladungen verfügbar, die deren Signifikanz bzw. Insignifikanz beurteilen erlauben (signifikante Ladungen sind dabei mindestens doppelt so groß wie ihre Standardfehler).

Wenn auch die Prüfgröße χ^2 zu erkennen gibt, daß die Modelle in beiden Altersgruppen aufrechterhalten werden können, zeigt die genauere Inspektion der Standardfehler, daß weitere Faktorladungen auf Null gesetzt und die Modelle damit weiter restringiert werden sollten. Die Lösungen für diese restringierten Modelle sind in Tab. 3 dargestellt; sie dürften so etwas wie die ‚bestpassende‘ Einfachstruktur für die Beziehungen zwischen den beobachteten Indikatoren und den latenten Konstrukten repräsentieren.

Es muß in diesem Zusammenhang erwähnt werden, daß gemäß einem Vorschlag von JÖRESKOG (1978) die ersten Ableitungen der Maximum-Likelihood-Funktion daraufhin überprüft wurden, ob die auf Null gesetzten Ladungen auch tatsächlich bei der Modellschätzung ohne Bedeutung geblieben sind. Da in der Regel alle ersten Ableitungen auch nahe Null lagen, ließ sich die Richtigkeit dieser Maßnahme in fast allen Fällen bestätigen. Als einzige Ausnahme kann die Ladung des Wortschatz-Tests auf Faktor 1 in der Gruppe der Viertkläßler gelten: die Freisetzung des Schätzparameters führte zu einer signifikant besseren Anpassung des Modells an die Daten.

Der nächste Schritt bestand darin, die in den Tabellen 2 und 3 aufgeführten COFAMM-Lösungen auf Differenzen in der Anpassungsgüte hin zu untersuchen. Da die Unterschiede zwischen zwei χ^2 -Werten wiederum χ^2 -verteilt sind, können zwei Lösungen daraufhin geprüft werden, ob sie sich in der Anpassung an die Daten signifikant unterscheiden.

Die Ergebnisse dieser Vergleiche zeigten, daß die Lösungen des restringierten COFAMM-Modells (Tab. 3) nicht signifikant schlechter als die des ursprünglichen Modells ausfielen; ganz im Gegenteil erwies sich bei den Zweitkläßlern die restringierte Variante als signifikant überlegen. Von daher schien es gerechtfertigt, diese Lösungen als neue Startwerte für das LISREL-Meßmodell heranzuziehen, das hier durch die in Abb. 2 dargestellte Strukturkomponente ergänzt wurde.

Tabelle 2: Ergebnisse der COFAMM-Analyse (vollständiges Modell) für beide Klassenstufen

(Standardfehler sind in Klammern wiedergegeben; auf Null gesetzte Ladungen sind durch ein ‚+‘ gekennzeichnet)

a) Lösung für die Zweitkläßler

	F1	F2	F3	F4
Motiv	.022 (.069)	.198 (.086)	-.309 (.135)	-.015 (.152)
d2	-.154 (.072)	.194 (.085)	-.091 (.145)	.134 (.160)
ZB	-.063 (.070)	-.175 (.087)	.212 (.140)	.218 (.152)
WL	-.214 (.075)	.088 (.091)	.083 (.179)	.389 (.188)
CFT	0.0 ⁺	0.0 ⁺	0.0 ⁺	.620 (.113)
WS	.006 (.059)	-.004 (.078)	.581 (.135)	.311 (.135)
LV	0.0 ⁺	0.0 ⁺	.812 (.097)	0.0 ⁺
Dikt.1	.009 (.044)	.906 (.074)	-.048 (.095)	-.103 (.095)
Dikt.2	0.0 ⁺	.940 (.060)	0.0 ⁺	0.0 ⁺
Dikt.3	.773 (.099)	.347 (.052)	.204 (.133)	-.212 (.128)
Dikt.4	1.019 (.111)	0.0 ⁺	0.0 ⁺	0.0 ⁺

Anpassungsgüte: $\chi^2_{(17)} = 10.24$ (p = .89)

<u>Faktor-Interkorrelationen:</u>	F1	F2	F3
F2	.10		
F3	.20	.39	
F4	.15	.43	.62

b) Lösung für die Viertkläßler

	F1	F2	F3	F4
Motiv	0.0 ⁺	0.0 ⁺	0.0 ⁺	.570 (.120)
d2	-.080 (.076)	.228 (.097)	.285 (.118)	.205 (.102)
ZB	.127 (.074)	.049 (.090)	.158 (.092)	.134 (.101)
WL	0.0 ⁺	0.0 ⁺	.843 (.232)	0.0 ⁺
CFT	-.112 (.076)	.608 (.089)	.039 (.082)	.192 (.102)
WS	0.0 ⁺	.791 (.065)	0.0 ⁺	0.0 ⁺
LV	-.123 (.075)	.743 (.089)	.076 (.079)	-.014 (.092)
Dikt.1	.911 (.074)	-.056 (.049)	-.090 (.053)	.126 (.056)
Dikt.2	1.050 (.051)	0.0 ⁺	0.0 ⁺	0.0 ⁺
Dikt.3	-.266 (.093)	.127 (.094)	.022 (.090)	-.406 (.123)
Dikt.4	.369 (.070)	.156 (.084)	.216 (.096)	-.138 (.092)

Anpassungsgüte: $\chi^2_{(17)} = 16.59$ (p = .48)

<u>Faktor-Interkorrelationen:</u>	F1	F2	F3
F2	.46		
F3	-.20	-.25	
F4	-.31	-.25	.09

Tabelle 3: Ergebnisse für das restringierte COFAMM-Modell („best-fitting simple structure“) in beiden Klassenstufen
(Die höchsten Ladungen in jeder Spalte sind unterstrichen; auf Null gesetzte Ladungen sind durch ein ‚+‘ gekennzeichnet)

a) Lösung für die Zweitkläblier

	F1	F2	F3	F4
Motiv	0.0 ⁺	.192	-.312	0.0 ⁺
d2	-.154	.216	0.0 ⁺	0.0 ⁺
ZB	0.0 ⁺	-.186	.173	.237
WL	-.209	0.0 ⁺	0.0 ⁺	.514
CFT	0.0 ⁺	0.0 ⁺	0.0 ⁺	<u>.581</u>
WS	0.0 ⁺	0.0 ⁺	.559	.318
LV	0.0 ⁺	0.0 ⁺	<u>.807</u>	0.0 ⁺
Dikt.1	0.0 ⁺	.909	0.0 ⁺	-.142
Dikt.2	0.0 ⁺	<u>.945</u>	0.0 ⁺	0.0 ⁺
Dikt.3	.770	.365	.243	-.253
Dikt.4	<u>1.017</u>	0.0 ⁺	0.0 ⁺	0.0 ⁺

Faktor-Interkorrelationen:

	F1	F2	F3
F2	.11		
F3	.20	.38	
F4	.15	.49	.70

Anpassungsgüte:

$$\chi^2_{(28)} = 13.42 \quad (p = .99)$$

b) Lösung für die Viertkläblier

	F1	F2	F3	F4
Motiv	0.0 ⁺	0.0 ⁺	0.0 ⁺	<u>.540</u>
d2	0.0 ⁺	.218	.286	.232
ZB	0.0 ⁺	0.0 ⁺	.241	0.0 ⁺
WL	0.0 ⁺	0.0 ⁺	<u>.786</u>	0.0 ⁺
CFT	0.0 ⁺	.593	0.0 ⁺	.198
WS	.136	.705	0.0 ⁺	0.0 ⁺
LV	0.0 ⁺	<u>.743</u>	0.0	0.0
Dikt.1	.870	0.0 ⁺	-.095	.123
Dikt.2	<u>1.075</u>	0.0 ⁺	0.0 ⁺	0.0 ⁺
Dikt.3	-.234	0.0 ⁺	0.0 ⁺	-.456
Dikt.4	.417	.176	.176	0.0 ⁺

Faktor-Interkorrelationen:

	F1	F2	F3
F2	.31		
F3	.22	.32	
F4	-.33	-.33	.05

Anpassungsgüte:

$$\chi^2_{(29)} = 27.63 \quad (p = .54)$$

2.2.2 Ergebnisse der LISREL-Analysen

Wie schon aus der visuellen Inspektion der Faktorladungs-Matrizen in Tab. 3 abzuleiten ist, ließ sich ein für beide Klassenstufen simultan geschätztes Modell nicht identifizieren: wiederum mußte die Annahme eines in beiden Gruppen gleichartigen Meßmodells zurückgewiesen werden. Es blieb von daher nichts anderes übrig, als verschiedene Modelle jeweils separat für beide Altersstufen zu formulieren und zu testen. Während es aufgrund theoretischer Erwägungen gerechtfertigt war, die Rechtschreibkompetenz als ‚letzte‘ endogene Variable (Kriteriumsvariable) im Strukturmodell anzunehmen, war kein a-priori-Wissen über die beste Rangordnung der drei übrigen endogenen Konstrukte verfügbar. Die explorative Prüfstrategie sah demnach den Test und Vergleich konkurrierender Kausalmodelle vor, in denen die Position der latenten Variablen systematisch variiert wurde. Mehr als zwanzig Modelle mit plausiblen und unplausiblen rekursiven bzw. nicht-rekursiven Inter-Relationen³ wurde spezifiziert, geschätzt und getestet und schließlich auf ihre Kompatibilität mit den Daten hin verglichen.

In Tab. 4 sind für jede Stichprobe die beiden jeweils am besten passenden Modelle wiedergegeben, die sich im Hinblick auf ihre Daten-Kompatibilität

Tabelle 4: Anpassungsgüte für konkurrierende Kausalmodelle (für Zweit- und Viertkläßler getrennt wiedergegeben)

a) Modell-Vergleiche für die Zweitkläßler

Modell 1 (F4 → F3 → F1 → F2)	:	$\chi^2_{(60)} = 43.56$	(p = .94)
Modell 2 (F3 → F4 → F1 → F2)	:	$\chi^2_{(59)} = 45.60$	(p = .90)
Reduziertes Modell 1	:	$\chi^2_{(51)} = 35.44$	(p = .95)

b) Modell-Vergleiche für die Viertkläßler

Modell 1 (F4 → F3 → F2 → F1)	:	$\chi^2_{(56)} = 67.28$	(p = .14)
Modell 2 (F4 → F2 → F3 → F1)	:	$\chi^2_{(59)} = 71.72$	(p = .12)
Reduziertes Modell 1	:	$\chi^2_{(49)} = 56.35$	(p = .22)

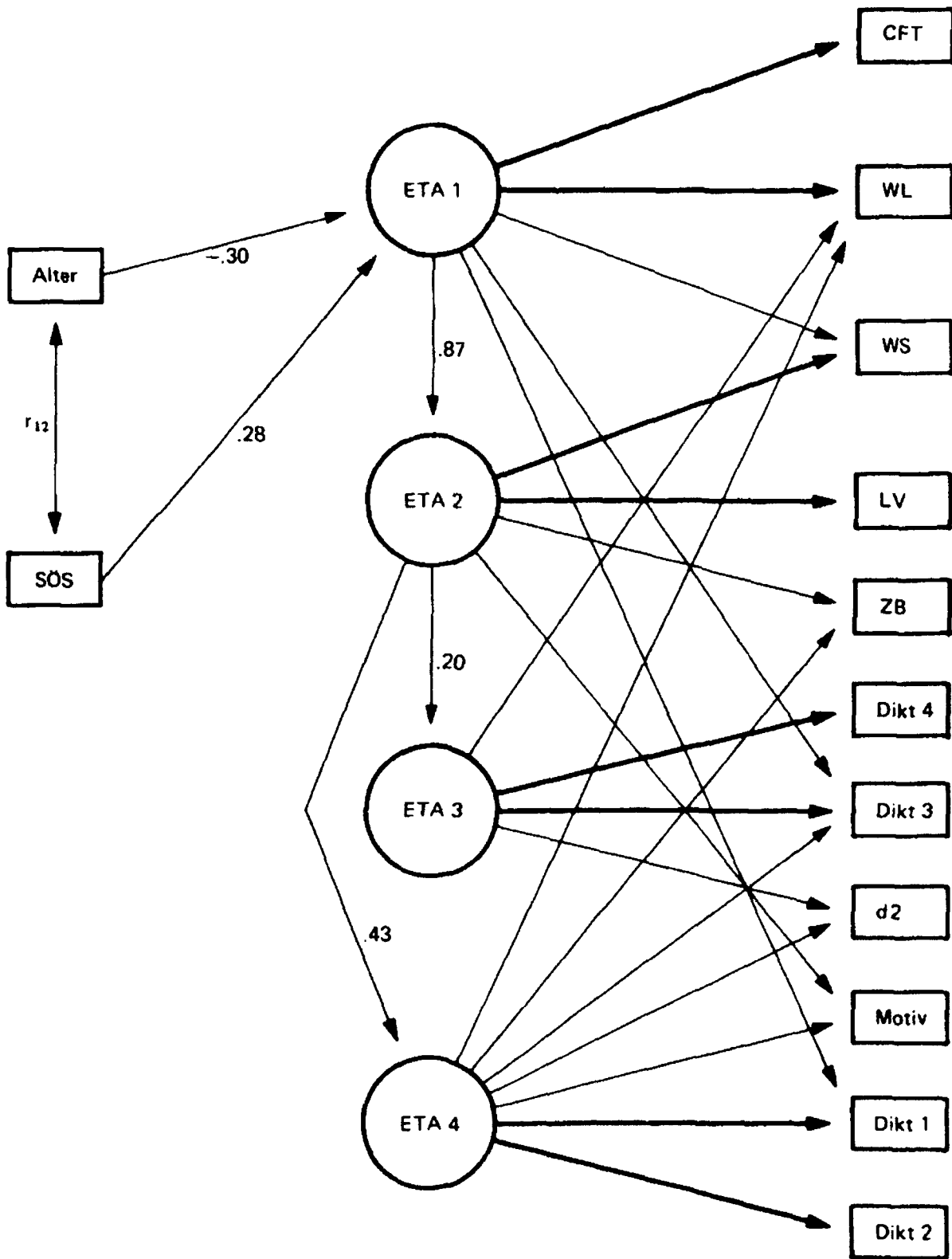


Abb. 3: Das empirisch ‚best-passende‘ LISREL-Modell für die Zweitkläßler

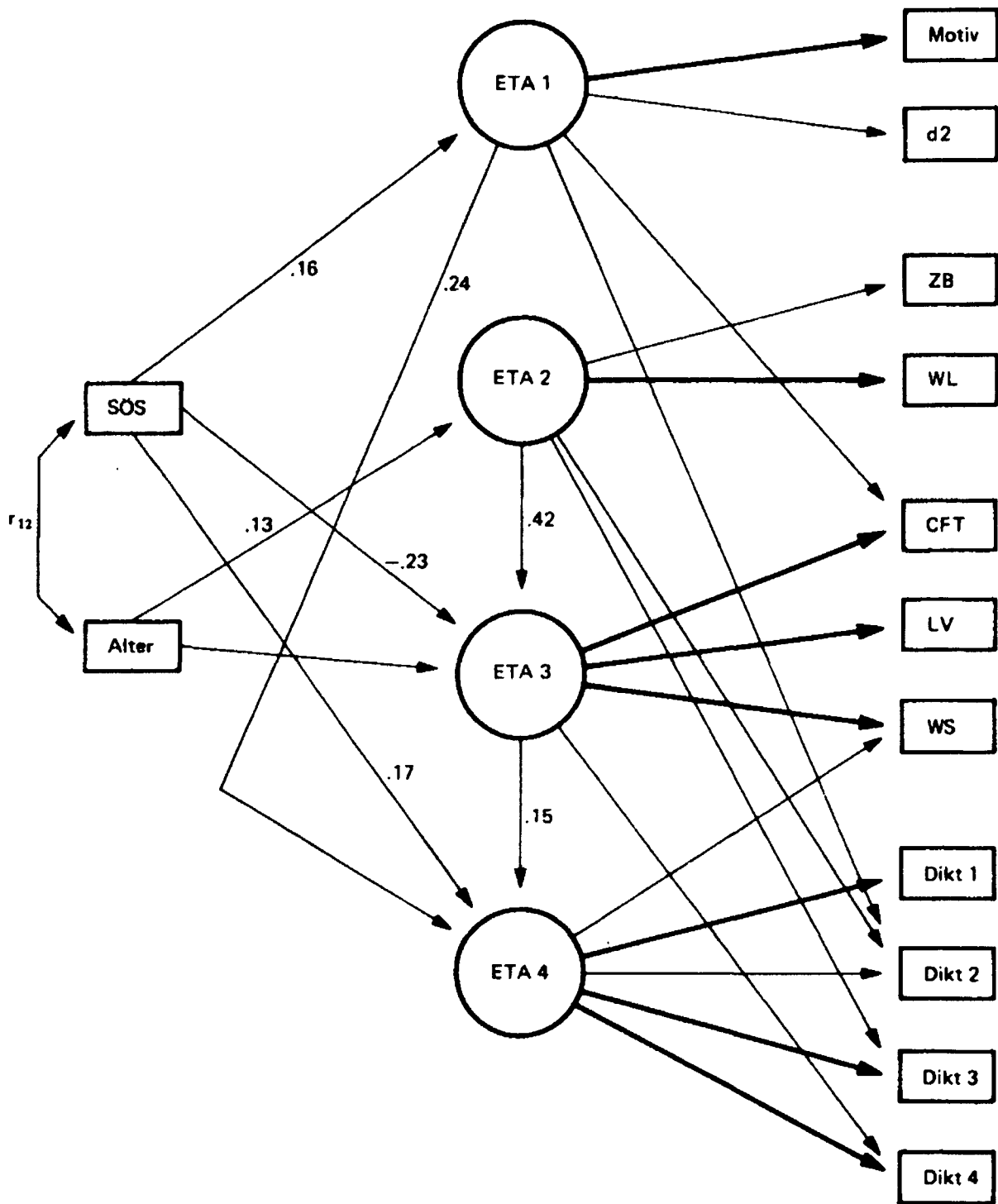


Abb. 4: Das empirisch ‚best-passende‘ LISREL-Modell für die Viertkläßler

nur geringfügig unterscheiden. Zusätzlich werden die Resultate der Schätzungen für ein reduziertes Modell dargestellt, bei dem die Geschlechtsvariable ausgespart wurde. Für dieses sparsamere Modell ließ sich in beiden Altersgruppen ein (wenn auch nicht signifikant) besserer Anpassungswert erzielen (vgl. Tab. 4), so daß im folgenden diese Lösungen näher diskutiert werden sollen. Vergleicht man die in Abb. 3 und Abb. 4 wiedergegebenen Pfadmodelle mit dem theoretisch plausiblen Modell in Abb. 2, so kann nur für die Lösung der Viertkläßler eine gewisse Ähnlichkeit konstatiert werden. Wenn man für die Interpretation lediglich die (besonders markierten) größeren Faktorladungen heranzieht, lassen sich die Eta-Variablen ähnlich wie in Abb. 2 benennen. Es scheint also durchaus sinnvoll, hier einen nichtkognitiven (motivationalen) Faktor Eta 1 anzunehmen, der den Rechtschreibfaktor Eta 4 direkt beeinflusst; weiterhin einen Gedächtnisfaktor Eta 2, der direkt den IQ-Faktor Eta 3 determiniert, während letzterer wiederum direkt auf den Rechtschreibfaktor einwirkt.

Obwohl sich die in Abb. 2 postulierten Beziehungen zwischen den zwei exogenen Variablen und den Eta-Variablen grundsätzlich auch in dem empirisch haltbaren Modell nachweisen lassen, weist letzteres gerade im Hinblick auf die strukturellen Beziehungen zwischen den endogenen latenten Variablen doch ein beträchtlich sparsameres Muster auf. Von den sechs Beta-Koeffizienten in Abb. 2 läßt sich gerade noch die Hälfte in Abb. 4 wiederfinden; mit Ausnahme des Pfads von Gedächtnis zu IQ sind keine engen Beziehungen zwischen den endogenen Variablen nachweisbar.

Im Hinblick auf das Meßmodell (s. Tab. 5) gilt geradezu das Gegenteil: das empirisch bestätigte Kausalmodell ist im Vergleich zum theoretisch plausiblen (in Abb. 2) deutlich komplexer strukturiert. Es muß von daher nochmals betont werden, daß die Etikettierung der Eta-Variablen als ein übersimplifizierendes Verfahren zu werten ist, da zahlreiche Verbindungsglieder ignoriert werden, die im vorliegenden Kontext nur schwer zu erklären bzw. zu interpretieren sind.

Was nun den Vergleich des theoretisch plausiblen Modells mit der empirischen Lösung für die Zweitkläßler angeht, so sind hier nur wenige Gemeinsamkeiten auszumachen. Offensichtliche Diskrepanzen betreffen das strukturelle Muster zwischen den exogenen Variablen und den endogenen Konstrukten, die Beziehungen zwischen den latenten Faktoren sowie das Meßmodell für die endogenen Variablen.

Es scheint hier zunächst einmal evident, daß die exogenen Variablen relativ bedeutungslos für den endogenen Teil des Modells sind; so führen etwa nur zwei Pfade zur ersten endogenen latenten Variable hin. Unterschiede in Alter und Schichtzugehörigkeit dürften demnach die Rechtschreibfertigkeit von Zweitkläßlern nur marginal beeinflussen. Ähnlich unbedeutend scheinen auch nichtkognitive Einflußgrößen wie Konzentration und Motivation zu sein: Eta 1 setzt sich hier aus einer Kombination von Indikatoren des Gedächtnisses und der Intelligenz zusammen, deren gemeinsamer Nenner möglicherweise in der Erfordernis von Inferenzprozessen zu sehen ist. Eta 2 um-

Tabelle 5: Lambda-Schätzungen für die best-angepaßten LISREL-Modelle (für Zweit- und Viertkläßler getrennt)

a) Lösung für die Zweitkläßler

	ETA 1	ETA 2	ETA 3	ETA 4
Motiv	0.0	-.310	0.0	.202
d2	0.0	0.0	-.153	.218
ZB	0.0	.337	0.0	-.151
WL	.421	0.0	-.215	.131
CFT	.609	0.0	0.0	0.0
WS	-.450	1.161	0.0	0.0
LV	0.0	.819	0.0	0.0
DIKT.1	-.107	0.0	0.0	.871
DIKT.2	0.0	0.0	0.0	.953
DIKT.3	-.398	.379	.788	.319
DIKT.4	0.0	0.0	1.000	0.0
Unerklärte Varianz	80%	24%	96%	82%

b) Lösung für die Viertkläßler

	ETA 1	ETA 2	ETA 3	ETA 4
Motiv	.616	0.0	0.0	0.0
d2	.232	.478	0.0	0.0
ZB	0.0	.228	0.0	.083
WL	0.0	.595	0.0	0.0
CFT	.141	0.0	.579	0.0
WS	0.0	0.0	.684	.174
LV	0.0	0.0	.745	0.0
DIKT.1	.113	0.0	0.0	.870
DIKT.2	0.0	.166	0.0	1.020
DIKT.3	-.385	0.0	0.0	-.200
DIKT.4	0.0	.244	.143	.432
Unerklärte Varianz	97%	98%	63%	85%

faßt Merkmale der verbalen Intelligenz, die direkt auf die Faktoren Eta 3 und Eta 4 einwirken. Ist es schon erstaunlich genug, daß das Meßmodell für diese Altersgruppe zwei separate Rechtschreibfaktoren Eta 3 und Eta 4 enthält, so muß sicherlich noch mehr überraschen, daß keinerlei kausale Beziehung zwischen den beiden Größen entdeckt werden kann: die Rechtschreibfaktoren sind als unabhängig einzustufen. Dieser Befund legt die Schlußfolgerung nahe, daß in dieser Klassenstufe Rechtschreibfertigkeit nicht als eindimensionale Größe aufgefaßt werden kann, sondern stark von orthographischen Merkmalen des Textes (z. B. der Wortschwierigkeit) abhängt.

3. Schlußfolgerungen

Da die vorgelegte Untersuchung vorwiegend explorativen Charakter hatte, soll hier nur auf die hauptsächlichsten Befunde eingegangen werden. Es ist festzuhalten, daß sich die wesentliche Fragestellung der Studie klar beantworten ließ: die Hypothese einer für beide Altersgruppen äquivalenten Kausalstruktur und damit eines für beginnende und geübte Rechtschreiber gleichermaßen gültigen Bedingungsmodells der Rechtschreibleistung ließ sich durch die Daten nicht stützen und mußte zurückgewiesen werden. Es ließ sich zeigen, daß relativ unterschiedliche Kausalmodelle mit den Daten aus beiden Stichproben vereinbar waren, wobei lediglich das Strukturmodell für die vierten Klassen Ähnlichkeiten zu dem theoretisch postulierten Rechtschreibmodell aufwies.

Aus diesen Befunden kann der Schluß gezogen werden, daß wohl nur wenige Befunde der Rechtschreibforschung dazu benutzt werden können, um die Kompetenz von beginnenden Rechtschreibern zu erklären. Die Ergebnisse lassen sich demnach nicht über die Gruppe der geübteren Rechtschreiber hinaus verallgemeinern, an der sie zumeist erhoben worden sind.

Was nun die spezifischen Muster der Strukturgleichungsmodelle für beide untersuchten Stichproben angeht, so sind hier vorsichtige Interpretationen angezeigt. Die hier eingesetzte Strategie zur Auffindung der ‚best-passenden‘ Einfachstruktur hat explorativen Charakter und begünstigt eine Überanpassung des Modells an die verfügbaren Daten (sample overfit). Von daher scheint eine Kreuzvalidierung der beiden empirisch bestätigten Kausalmodelle unerlässlich, um die Validität der Befunde besser beurteilen zu können. Aufgrund der geringen Stichprobengrößen ließ sich ein solcher Test hier nicht durchführen.

Eine letzte Bemerkung zielt auf den Befund, daß die Modell-Anpassung für die Zweitkläßler wesentlich besser gelang. Es mag in diesem Zusammenhang eine Rolle gespielt haben, daß in der vorliegenden Untersuchung lediglich Indikatoren für die kognitive bzw. nicht-kognitive Entwicklung der Schüler sowie Merkmale der Text-Struktur systematisch berücksichtigt wurden, während beispielsweise Variablen wie Qualität oder Quantität des Unterrichts fehlten. Es kann nun durchaus sein, daß diese Fehl-Spezifikation des Modells

die Schätzung für die Viertkläßler ungleich stärker negativ beeinflußt hat, da die Effekte von Lehrerverhalten und Differenzen im Unterrichtsangebot in dieser Altersstufe deutlicher ausgeprägt sein sollten.

Die hier aufgezeigten wie auch andere Mängel dieser Untersuchung sind darauf zurückzuführen, daß die Primärstudie zu einem anderen Zweck geplant war und deshalb einige Voraussetzungen der Sekundäranalyse nicht erfüllen konnte. Um die oben skizzierten Befunde in ihrem Gültigkeitsbereich näher abschätzen zu können, scheint deshalb eine umfassende Replikationsstudie unerlässlich zu sein.

Anmerkungen

- 1 Für kritische Kommentare und Anregungen danke ich Peter M. Bentler, Claus Möbus, Klaus Opwis, Ralf Schwarzer und Martin Sieber; Klaus Opwis weiterhin für seine Unterstützung bei der Durchführung der umfangreichen Rechenarbeiten.
- 2 Der Begriff ‚Sekundäranalyse‘ wird hier gemäß der Definition von GLASS (1976) benutzt; er meint die Re-Analyse des gleichen Datensatzes mit einer angemesseneren statistischen Methode.
- 3 Diese zeitaufwendige Prozedur schien schon deshalb notwendig, weil nachgewiesen werden konnte, daß manchmal auch unsinnige Modelle mit den Daten kompatibel sein können (vgl. MÖBUS, 1981).

Literatur

- Angermaier, M.: Drei Faktorenanalysen zum Thema ‚Legasthenie‘. *Z. f. exp. angew. Psychol.*, *20*, 1–19 (1973)
- Angermaier, M.: *Legasthenie – Verursachungsmomente einer Lernstörung*. Weinheim: Beltz 1974
- Bentler, P.M.: Multivariate analysis with latent variables: Causal modeling. *Ann. Rev. Psychol.*, *31*, 419–456 (1980)
- Bentler, P.M. & Huba, G.J.: Simple minitheories of love. *J. Person. & Soc. Psychol.*, *37*, 124–130 (1979)
- Bentler, P.M. & Speckart, G.: Models of attitude-behavior relations. *Psychol. Rev.*, *86*, 452–464 (1979)
- Brickenkamp, R.: *Test d2 Aufmerksamkeits-Belastungstest*. Göttingen: Hogrefe 1972
- Cattell, R.B. & Weiss, R.: *Grundintelligenztest CFT2 Skala 2*. Braunschweig: Westermann 1972
- Cunningham, W.R. & Birren, J.E.: Age changes in the factor structure of intellectual abilities in adulthood and old age. *Educ. & Psychol. Measmt.*, *40*, 271–290 (1980)
- Fippinger, F.: *Allgemeiner Schulleistungstest für 4. Klassen (AST4)*. Weinheim: Beltz 1967
- Fisher, F.M.: *The identification problem in econometrics*. New York: McGraw-Hill 1966
- Glass, G.V.: Primary, secondary, and meta-analysis of research. *Educ. Researcher*, *5*, 3–8 (1976)
- Graff, J.: *Ein Modell der Theorie der kognitiven Dissonanz*. Unveröff. Dissertation, Hamburg 1978
- Horn, J.L. & McArdle, J.J.: Perspectives on mathematical/statistical model building (MASMOB) in research on aging. In: L.W. Poon (ed.), *Aging in the 1980's: Psychological issues*. Washington, D.C.: APA 1980

- Jöreskog, K.G.: Structural equation models in the social sciences: Specification, estimation, and testing. In: P.R. Krishnaiah (ed.), *Applications of statistics*. Amsterdam: North Holland 1977
- Jöreskog, K.G.: Structural analysis of covariance and correlation matrices. *Psychometrika*, **43**, 443–477 (1978)
- Jöreskog, K.G. & Sörbom, D.: LISREL IV Users Guide. Chicago: Nat. Educ. Res. 1978
- Jöreskog, K.G. & Sörbom, D.: *Advances in factor analysis and structural equation models*. Cambridge, Mass.: Abt Books 1979
- Kerlinger, F.N.: Analysis of covariance structure tests of a criterial referents theory of attitudes. *Multivar. Behav. Res.*, **15**, 403–422 (1980)
- Maruyama, G. & McGarvey, B.: Evaluating causal models: An application of maximum-likelihood analysis of structural equations. *Psychol. Bull.*, **87**, 502–512 (1980)
- Möbus, C.: Zur Abbildung von Entwicklungssequenzen mittels Strukturgleichungssystemen (Test-Retest-Situationen). In: G. Rudinger (Hrsg.), *Methoden der Entwicklungspsychologie*. Stuttgart: Kohlhammer 1981
- Olsson, U. & Bergman, L.R.: A longitudinal factor model for studying change in ability structure. *Multivar. Behav. Res.*, **12**, 221–242 (1977)
- Schlee, J.: *Legasthenieforschung am Ende?* München: Urban & Schwarzenberg 1976
- Schmidt, P.: *Zur praktischen Anwendung von Theorien: Grundlagenprobleme und Anwendung auf die Hochschuldidaktik*. Unveröff. Dissertation, Mannheim 1977
- Schneider, W.: *Bedingungsanalysen des Recht-Schreibens*. Bern: Huber 1980
- Valtin, R.: *Legasthenie – Theorien und Untersuchungen*. Weinheim: Beltz 1974
- Weeks, D.G.: A second-order longitudinal model of ability structure. *Multivar. Behav. Res.*, **15**, 353–365 (1980)
- Werts, C.E., Linn, R.L. & Jöreskog, K.G.: A simplex model for analyzing academic growth. *Educ. & Psychol. Measmt.*, **37**, 745–756 (1977)
- Werts, C.E., Linn, R.L. & Jöreskog, K.G.: Reliability of college grades from longitudinal data. *Educ. & Psychol. Measmt.*, **38**, 89–95 (1978)
- Wiley, D.E.: The identification problem for structural equation models with unmeasured variables. In: A.S. Goldberger & O.D. Duncan (eds.), *Structural equation models in the social sciences*. New York: Seminar Press 1973
- Ziegler, R.: *Theorie und Modell*. München: Oldenbourg 1972

Dr. Wolfgang Schneider
 z. Zt. Stanford University
 Department of Psychology
 Jordan Hall, Bldg. 420
 Stanford, California 94305
 USA

ab 1.10.82:
 Max-Planck-Institut für
 Psychologische Forschung
 Leopoldstr. 24
 D-8000 München 40