

American Educational Research Journal

<http://aerj.aera.net>

Classroom Differences in the Determination of Achievement Changes

Wolfgang Schneider and Bernhard Treiber

Am Educ Res J 1984; 21; 195

DOI: 10.3102/00028312021001195

The online version of this article can be found at:
<http://aer.sagepub.com/cgi/content/abstract/21/1/195>

Published on behalf of



American
Educational
Research
Association

<http://www.aera.net>

By



<http://www.sagepublications.com>

Additional services and information for *American Educational Research Journal* can be found at:

Email Alerts: <http://aerj.aera.net/cgi/alerts>

Subscriptions: <http://aerj.aera.net/subscriptions>

Reprints: <http://www.aera.net/reprints>

Permissions: <http://www.aera.net/permissions>

Classroom Differences in the Determination of Achievement Changes

WOLFGANG SCHNEIDER and BERNHARD TREIBER
University of Heidelberg

This study addresses three themes that recur in the research on student achievement: (a) developmental modeling of intraindividual changes in achievement over time; (b) examination of the differences among subgroups within a classroom in the determinants of achievement; (c) description of the interactions among instructional variables in determining achievement differences. Eight classrooms were preselected on the basis of their widely differing slopes obtained in a regression analysis of pre- and posttest achievement scores. Mathematics achievement differences among sixth graders were analyzed in a four-wave design and explained by aptitude and instructional variables in a structural equation framework provided by LISREL. The results demonstrate the local nature of achievement models in that neither their measurement nor structural components proved generalizable across both groups of classrooms. Mention is also made, however, of technical problems and analytical ambiguities in the interpretation of these results.

Recent models of achievement are faced with a number of descriptive and explanatory problems that are rarely solved successively, let alone simultaneously (i.e., within one investigation). First, the changes in achievement over time are inadequately dealt with in conventional pre-post designs. They are better handled by more extended longitudinal studies with multiple measurements. Appropriate longitudinal studies are still rare, however. Where they do exist, they make insufficient use of available

We wish to thank Lyn Corno and David Rogosa for their critical comments and Klaus Opwis for his assistance in data analysis.

descriptive models in analyzing inter- and intraindividual characteristics of achievement changes.

Instructional variables, when introduced into achievement models, tend to be treated as a uniform block of concurrent determinants. Mention is rarely made, however, of the fact that at least some of these variables may be negatively intercorrelated and so conflict with regard to a given criterion (e.g., classroom mean achievement). Effective instruction may then more resemble a careful balance of partially reconcilable inputs than the concerted effort of maximizing their impact for a given educational goal. Again, studies that describe these more complicated interdependencies in explicit measurement models of instructional practices are rare.

Achievement processes usually are explained by characteristics of students and their learning environments. Studies that compare the relative importance of both classes of determinants have a long tradition, for example, in the debate of the ascriptive versus achieved nature of educational (and later occupational) opportunities. It is also in this tradition that explicit teaching-learning models attempt to maximize the degree and range of instructional control over critical aspects of achievement processes and likewise to reduce the impact of dispositional (e.g., ability) and structural (e.g., social or ethnic) achievement conditions. Empirical tests of these models, however, have frequently been of an illustrative or exploratory nature only, overfitting to the idiosyncrasies of a given sample. Comparative studies of competing models of achievement across several groups of students or confirmatory tests of an a priori preselected and prespecified model for a given student group are still rare. This is particularly unfortunate for arriving at an accurate understanding of the kind and range of diversity in the causal determination of student achievement under traditional teaching-learning conditions. The present study concentrates on these research problems by introducing and testing a specific structural model of achievement processes under traditional classroom conditions in a longitudinal, structural-equation framework.

The dynamic features of achievement processes are best captured by longitudinal data obtained at two or more occasions and analyzed in explicit multiwave models that describe interindividual differences in achievement change. Despite their apparent analytical advantages (cf. Baltes, Cornelius, & Nesselroade, 1979; Bielby & Hauser, 1977; Jöreskog & Sörbom, 1979), these models should still be viewed with some caution (see Rogosa, 1979, p. 287). Critical assumptions (e.g., measurement model and error specification) are usually oversimplified and neglect the unique difficulties of specifying adequate structural-descriptive models for multiwave panel data even for only one variable over time. In analyzing achievement processes, the present study therefore assumed a simplex structure for the longitudinal achievement data. Such models have been

termed simplex models by Guttman (1954) because of the typical pattern of intercorrelations they give rise to, showing a consistent decline of coefficient values with increasing distance from the matrix diagonal. Simplex models seem particularly appropriate for the study of academic growth (cf. Jöreskog, 1979; Werts, Linn, & Jöreskog, 1977). Its structure can be expressed as a path model with paths occurring only between adjacent measurement periods. When this model fits the observed data, then reliabilities of the observed variables and unattenuated correlations between the true variables can be estimated except for the first and last period (cf. Werts et al., 1977). These advantages lead us to give a simplex structure to the endogenous part of the model introduced here, as depicted in Figure 1. In this typical LISREL path diagram observed indicators are enclosed in squares, whereas the latent factors are represented by circles. The arrows between the observed indicators and latent factors signify factor loadings, whereas the arrows between the latent factors designate structural or path coefficients. As the model permits the distinction between errors in measurement (δ) and errors in equation (ζ), these are included in the path diagram, denoted by smaller single-headed arrows. According to the LISREL

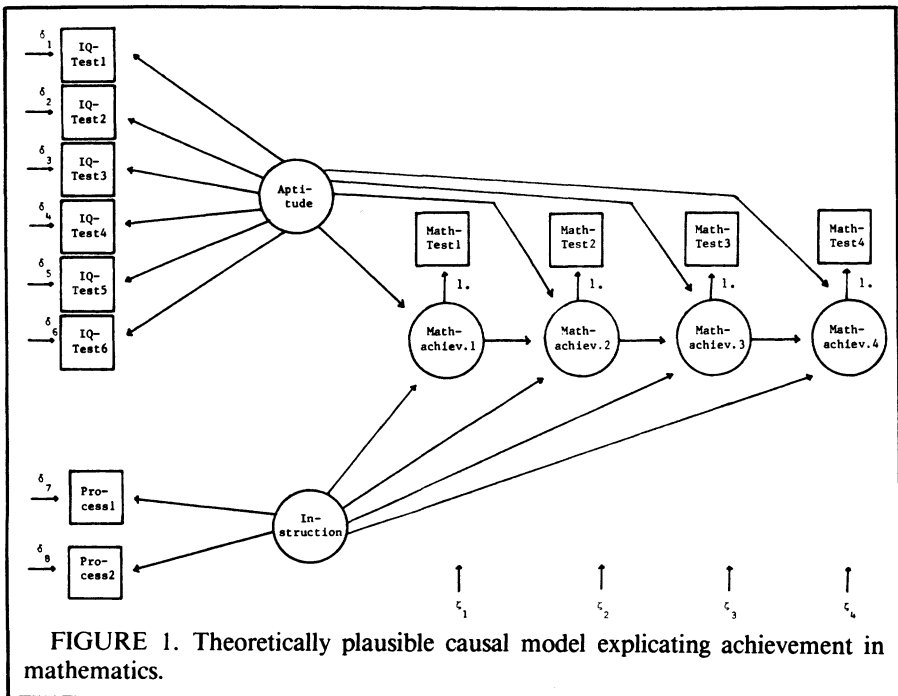


FIGURE 1. Theoretically plausible causal model explicating achievement in mathematics.

notation, two-headed arrows indicate correlations or covariances between indicators or constructs.

Student entry abilities and classroom instructional characteristics show up in a number of recent teaching-learning models as (specific or joint) determinants of student achievement (cf. Cronbach & Snow, 1977). They also appear as exogenous constructs of the achievement model in Figure 1. Researchers have tried to empirically test their models by comparing the relative impact of student ability and instructional style on some measure of achievement. This was also done in the present study. An implicit assumption of these tests has been that all features of a given model (i.e., its measurement and structural-relational parts) would be valid for all elements within a sample (of classrooms or students). This assumption was not adopted in the present study. Rather, an explicit attempt has been made to find out whether the same achievement model can be applied to classrooms with widely differing instructional "histories." This is particularly interesting for understanding the naturally occurring variation of structural effects of various student- versus classroom-level conditions on achievement. Previous studies have occasionally compared model characteristics (e.g., mean vectors, covariance matrices) between groups under the assumption that the distribution of their latent variables is different for different groups. The grouping criterion in the present study will be within-classroom slope differences in their pre- and posttest achievement regressions.

Following up an idea suggested by Wiley (1970), Burstein (1979), Burstein, Linn, and Capell (1978), and Burstein, Miller, and Linn (1979) argued that differences in within-group slopes can reflect the effects of substantive differences in instructional processes. Classrooms with greater opportunities to learn may have allowed certain students to take fuller advantages of their abilities, and may also have caused less able students to fall behind their classmates. Hence, when the posttest scores of these students are regressed onto their entry achievement scores, a steep slope will result. The opposite would be true, of course, for classrooms in which an explicit attempt is made to give disproportionate amounts of instructional support to less able students: Their entry achievement would then predict their posttest achievement only minimally. Thus, slope heterogeneity across classrooms may be regarded as a function of instructional treatment differences between them. It was for this reason that slope differences have been used in the present study to identify and select exceptional classrooms. The next step was to find out more about the instructional characteristics of these classrooms that might serve as explanatory variables. As to the achievement model depicted in Figure 1, the assumption is that classrooms with widely differing slopes also require models that differ in their structural and/or measurement components to

describe the determination of achievement processes within them. This will be explicated more fully later on.

The exogenous construct “Instruction” of the achievement model introduced here should, therefore, be sensitive to instructional differences that have a two-sided effect. They reflect differences in classroom teaching, resulting in within-classroom slope differences, and they also determine to what degree student aptitude variables become effective for student achievement processes.

A rather direct implication of current aptitude-treatment interaction (ATI) research is an obvious and an often empirically confirmed interaction (see Cronbach & Snow, 1977; Snow, 1977b) of ability and instructional treatments that differ in the information-processing burdens they place on the learner. Accordingly, regression slopes of cognitive outcomes on ability greatly depend on teaching conditions that relieve learners from difficult reading, analyzing complex concepts and coding new material, that is, that compensate particularly for, or circumvent less able learners’ weaknesses.

It is against this theoretical background that the following two process variables were chosen to represent the Instruction construct: frequency with which new subject matter was introduced; and frequency with which already covered subject matter was reviewed, repeated, and practiced.

The measurement model of the latent variable Instruction shown in Figure 1 is assumed to differ between the two groups of classrooms for substantive reasons, because the task of organizing the conflicting requirements of introducing and reviewing new subject matter has led to different instructional treatments in each of the two classroom types.

The theoretical framework of the present study consists then of the following components:

- an *explanatory* component that introduces three different hypothetical constructs—“Achievement,” “Aptitude,” and “Instruction”—as analytic units for describing and explaining achievement processes;
- a *structural* component that specifies the causal links between these three constructs as indicated in Figure 1;
- a *longitudinal* component that describes the statistical properties of a multiwave achievement data set in a simplex growth model (see above); and
- a *measurement* component that defines the relationship among these three unmeasured constructs and their observed indicators.

In this study, Aptitude is represented by six intelligence subtests (to be described later), and Instruction is represented by two observation process variables, which have already been introduced above. The longitudinal construct Achievement, however, has only a single, observed variable at each measurement to define it.

The resulting achievement model (see Figure 1) is used to compare the dependency of achievement processes over time on student aptitude and instruction in two preselected types of classrooms. Their major difference is their previous achievement determination as indexed by within-classroom pre- and posttest achievement regression slopes. It is assumed that slope differences between classrooms may also be seen as a result of the degree to which student aptitude and instruction can differentially determine student achievement outcomes. The expectation is, of course, that the achievement model depicted in Figure 1 can locate the exact source of the slope differences between both groups of classrooms, that is, in one or more of the model components listed above.

METHOD

Sample

This study was conducted during the math instruction of eight sixth-grade classes in four elementary schools in a school district in southwest Germany. These classrooms had been preselected out of the total population of 113 sixth-grade classes in this school district because of their unusually high versus low within-classroom, pre-posttest, standardized regression coefficients (slopes): .412 versus .900 were the respective slope means for each of the two resulting extreme groups (of 2×4 classrooms), a difference that proved to be statistically and practically significant ($p = .001$, $w^2 = .802$). Altogether, 195 students and their 8 math teachers participated.

Variables and Instruments

Achievement. Individual differences in mathematics achievement over 4 measurement points within a school year represented the dependent or endogenous variable block of this study. The achievement measures were the composite score on 36 multiple-choice items of a special mathematics test. This test was designed to reflect the subject matter taught that year in most of the math classes sampled, according to the school-district's adopted textbooks and curriculum aids. To establish content validity, textbooks and teacher ratings of opportunity-to-learn were examined to see how closely the material covered matched the content on the achievement test. Results of the examination revealed that all the teachers had covered most of the mathematics material that was tested.

Aptitude. This category includes scores on six standardized German intelligence tests:

- "Language comprehension" as subtest 6 of the *Kognitiver Fähigkeitstest KFT 4-13* by Heller, Gädike, and Weinläder (1976) and as subtest 2 of the *Aufgaben zum Nachdenken AzN 4+* by Hylla (1966);

- “Computational skills” as subtest 9 of the *Prüfsystem für Schul- und Bildungsberatung* by Horn (1969);
- “Induction” as subtest Q 3 of the *Kognitiver Fähigkeitstest KFT 4-13* by Heller et al., 1976);
- “Rotation” as subtest 7 of the *Prüfsystem für Schul- und Bildungsberatung* by Horn (1969); and
- “Visual identification” as subtest 10 of the *Prüfsystem für Schul- und Bildungsberatung*

These tests were administered in classrooms by trained administrators in December 1976.

Instructional process. Observational data of the instructional quantity and quality in these classrooms were collected with a complex coding system, details of which are found in Treiber (1982). A uniform observation length of 200 instructional minutes (5 lessons of 40 min each) was chosen. Classroom events were coded once every 2 minutes with this low-inference coding instrument (of objective counts of discrete behaviors). The maximum frequency of any event, if coded at all, would therefore be 100. However, only two variables were created from the data collected with this system: introduction of new subject matter, and review and practice of subject matter already introduced and taught. The frequencies coded in each of the eight classrooms were summed for each of the two categories considered here for the total of 100 event points. The resulting summed scores constituted the basic observational variables used in further analyses.

Data Collection

The longitudinal design covered four mathematical achievement tests, each administered within a single lesson: at the beginning of December 1976 (Math Test 1), end of March 1977 (Math Test 2), end of August 1977 (Math Test 3), and mid November 1977 (Math Test 4). Procedures for administering and scoring these tests remained the same over all four time points. The six intelligence subtests were administered in December 1976 within a single lesson. The math instruction of these eight classrooms was observed in 5 to 7 lessons over a period of 2 months, before the first achievement testing (in December 1976) but shortly after the beginning of the school year.

Data Analysis

Achievement models with the properties described above are adequately handled by methods for the analysis of covariance and correlation matrices, which were developed by Jöreskog (1977, 1978; Jöreskog & Sörbom, 1979). Thus, the appropriate computer program, LISREL IV (Jöreskog & Sörbom, 1978), was used in the present study. The flexibility of this structural equation approach has been demonstrated in several studies (cf. Bentler &

Woodward, 1978; Cunningham & Birren, 1980; Munck, 1979; Weeks, 1980; Whitely, 1980).

Because rather detailed introductions to the LISREL approach are available elsewhere (e.g., Bentler, 1980; Maruyama & McGarvey, 1980), only a short description will be given here. Although the models testable by LISREL still include the traditional structural equation systems (path models) with directly observed variables, the computer program is particularly designed to handle more complex models, including latent variables treated as underlying causes of the measured variables. LISREL consists of two parts: Whereas the measurement model defines the relationship between observed variables and unmeasured hypothetical constructs, the structural equation model ("causal" model) is used to specify the causal links among the latent variables. Maximum likelihood estimates of measurement and causal parameters are obtained simultaneously; LISREL makes use of all information in the data about each parameter in generating its estimates.

The efficiency of a given model is evaluated by a chi-square goodness-of-fit statistic with degrees of freedom equal to the difference between the number of known relationships and unknown parameters; the chi-square is a direct function of the discrepancy between the sample correlation matrix and that reproduced by the parameter estimates of the model. A large chi-square (relative to the degrees of freedom) indicates that the model does not provide a plausible representation of the causal process. In this case several possibilities exist to modify the model (cf. Bentler, 1980): Parameters whose estimates are small compared to their standard errors can be omitted and the resulting model reestimated; an inspection of the fit function with respect to the fixed (known) parameters can suggest ways of relaxing the model by introducing more parameters. If the obtained chi-square corresponds to a probability level greater than .05, it is concluded that the proposed model cannot be rejected.

Steps of Analysis

One of the advantages of the LISREL approach is that parts of the model can be tested separately. Thus, in case of poor data fit it is possible to detect the source of misspecification and reformulate the interrelationships in the submodel that has proven to be inadequate.

In the present investigation a sequential strategy was used to analyze specific properties of the exogenous and endogenous parts of the proposed model. The plan of the data analyses was as follows:

In a first step we tested whether the assumed measurement model held across the two groups. If the simultaneous model has to be rejected, independent (one sample) solutions should be carried out for the measurement and structural model. In a second step we examined whether the

multiwave-one-variable longitudinal submodel describing the endogenous part possessed the statistical properties of a simplex model. In the last step we estimated and tested the full model depicted in Figure 1.

RESULTS

Estimation of Math Achievement Parameters

For each of the four mathematical achievement tests, classical test analyses were done first to demonstrate the adequacy of the test items employed concerning their mean difficulty, discriminability, skewness, and consistency. To ensure that these achievement tests were also comparable over the four different measurement points, their Rasch-scalability was analyzed by a specially developed computer program (Rasch Item Analysis; cf. Wakenhut, 1974, pp. 139–143). Here, the conformity of item parameters as established in independent subject groups (student samples at different measurement points) was tested. Ability parameters could thus be estimated in these Rasch-analyses for students' summed scores in significantly reduced versions of the four math tests, and a person-specific parameter could be attributed to each student for each of the four measurement points. Alpha reliabilities of individual scores on the six cognitive aptitude measures were obtained also, ranging from .78 to .91.

Multivariate Analyses

Subsequent analyses did not use the overall correlation matrix based on all students of the total sample, but instead used the two matrices that were calculated separately for each of the two classroom (HiSlope vs. LoSlope) types. The same structural model, explained above, was used for both subgroups, however, to merge those measures that could readily be interpreted as multiple indicators of a single theoretical construct.

Measurement model tests. According to the sequence of analysis proposed above, the first problem to be analyzed concerned the stability of factor structure in both groups. It was assumed that a measurement model would be reliable if it constrained both factor loadings and factor intercorrelations to have the same pattern across the two groups. But the result of a simultaneous confirmatory factor analysis (COFAMM) indicated that the proposed model did not fit the data ($\chi^2 = 143.61$ with 59 degrees of freedom; $p < .01$). As a consequence, independent solutions were developed for each group. There seemed to be no evidence that the same measurement model could hold for both samples.

To arrive at the most parsimonious solution, an exploratory procedure suggested by Jöreskog (1978) was used to find a so-called “best-fitting simple structure” for a given number of factors, that is, a neat simple structure with many zero loadings. While the results for the HiSlope classes

showed a reasonably good data fit ($\chi^2 [17] = 25.04; p = .09$), the proposed measurement model had to be rejected for the LoSlope classes ($\chi^2 [17] = 40.63; p < .01$). The inspection of the first derivatives of the fit function indicated that, in particular, the assumption of uncorrelated measurement errors was questionable for both samples. As Sörbom (1975) pointed out, a correlation between errors of measurement within occasions seems plausible when similar tests are used to measure the same trait. This could certainly be true for the six aptitude tests administered within only a short time. Hence, a slightly modified measurement model was formulated for both samples, taking into account correlations between measurement errors for the parameters with the largest first order derivations in the model presented above. An estimation of this modified model resulted in a considerably better data fit for each of the two groups. The differences in chi-square were 19.5 with seven degrees of freedom for the HiSlope classes and 27.87 with five degrees of freedom for the LoSlope classes, which are both significant at the 1 percent level. Though the model provided an acceptable fit for each of the two samples, a far better solution was obtained for the HiSlope classes. As can be seen from Table I, the simple structure of factor loadings postulated in Figure 1 did not receive total support, but the resulting pattern did not differ too much from the expected structure.

Structural model tests. In the second step of our analysis, the simplex assumption was tested for both groups. Because the formulations of the quasi-Markov simplex model given by Jöreskog (1970) and Jöreskog (1979) are slightly different, both model specifications were estimated. However,

TABLE I
Factor Loadings of the Best-fitting COFAMM Solutions for HiSlope and LoSlope Classes

	HiSlope classes		LoSlope classes	
	Aptitude	Instruction	Aptitude	Instruction
IQ-Test 1	.69 (.63)	-.34 (-.19)	.68 (.62)	.00 (.00)
IQ-Test 2	1.00 (.72)	.00 (.00)	1.00 (.67)	.00 (.00)
IQ-Test 3	.64 (.43)	.00 (.00)	.44 (.36)	.00 (.00)
IQ-Test 4	.63 (.56)	-.36 (-.18)	.31 (.23)	.00 (.00)
IQ-Test 5	.36 (.33)	.00 (.00)	.87 (.28)	-.89 (-.22)
IQ-Test 6	.44 (.30)	-.23 (-.15)	.65 (.43)	.00 (.00)
Process 1	.00 (-.19)	.69 (.51)	.00 (.00)	-.78 (-.33)
Process 2	.00 (.00)	1.00 (1.00)	.00 (-.05)	1.00 (1.00)
	Goodness-of-fit: $\chi^2 (10) = 5.54; p = .85$		Goodness-of-fit: $\chi^2 (12) = 12.76; p = .39$	

Note. Numbers in parentheses refer to the estimates given by the full model solution (depicted in Figures 2 and 3). To guarantee greater clearness, coefficients less than .20 are omitted from these figures.

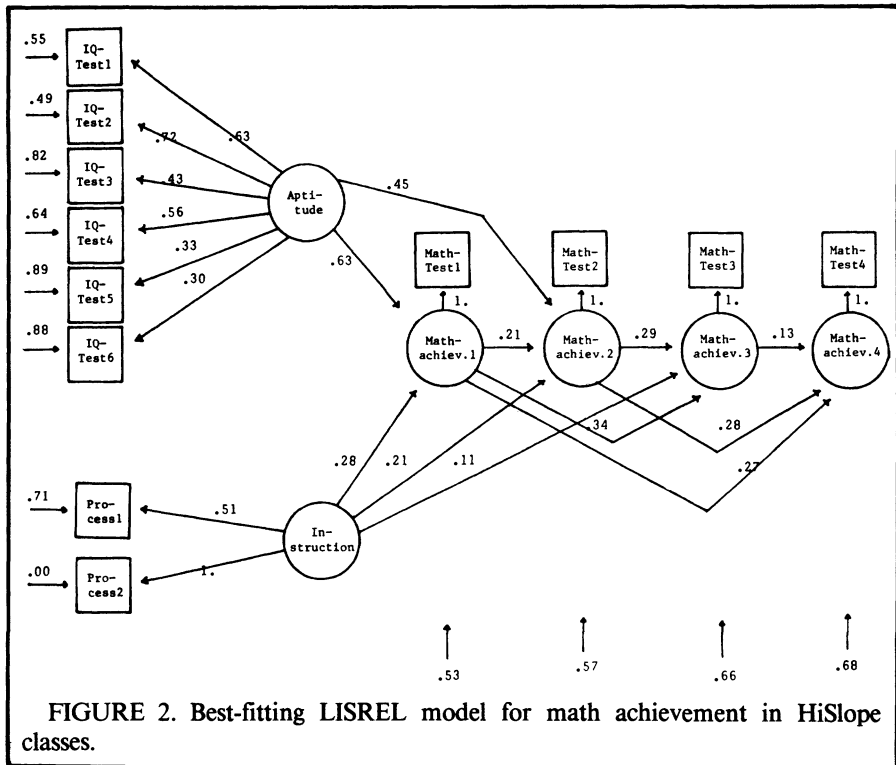
identical solutions were obtained in both cases. The results indicated that the assumed simplex structure had to be rejected for the LoSlope classes ($\chi^2 = 4.81$ with one degree of freedom; $p < .05$). An inspection of the first derivatives of the fit function yielded large values for the parameters of nonadjacent variables that were fixed to zero. Therefore, the simplex assumption that implied that the partial correlation between two nonadjacent variables x_1 and x_3 was zero, with x_2 controlled, could not hold for this sample. In contrast, the formulation of the quasi-Markov simplex model was acceptable for the HiSlope classes. For these classes, a reasonably good fit could be obtained ($\chi^2 [1] = .07$; $p = .78$). Thus it is only for the HiSlope classes that the simplex structure frequently found in models of academic growth could be confirmed.

As with four repeated measurements, only the two "inner" variables are identified (cf. Jöreskog, 1979; Werts et al., 1977), and estimated reliabilities could be obtained only for these observed variables (.54 and .43, respectively).

Full model test. The last step of analysis concentrated on the estimation and testing of the full model depicted in Figure 1. For each group the modified measurement model resulting from the COFAMM analyses was combined with the structural component to identify the most parsimonious solution. Preliminary analyses indicated that the theoretically proposed model had to be modified in two aspects. First, it proved necessary to give up the simplex assumption of the endogenous submodel when the full model was estimated for both groups. When the aptitude and instruction components were included, the endogenous submodel changed considerably. The direct causal links between the four measurement periods decreased in magnitude (compared with the solution for the simplex model described above), and it was necessary to introduce additional paths connecting the nonadjacent achievement measurements (cf. Figures 2 and 3).

Second, only on the first two or three occasions was a significant causal influence of the two exogenous constructs detected; so the trivial coefficients for the last measurement point could be fixed to zero without any loss of information. Because modifying a theoretical model poses problems insofar as a sample overfit cannot be ruled out (cf. Maruyama & McGarvey, 1980), only the most obvious misspecifications were corrected when testing the model assumptions.

The solutions for both samples are depicted in Figures 2 and 3. It must be pointed out that the model fit the data only for the HiSlope classes ($\chi^2 = 32.31$ with 39 degrees of freedom; $p = .77$), whereas the solution for the LoSlope classes yielded a significant $\chi^2 = 72.80$ with 36 degrees of freedom ($p < .01$), indicating that the model does not provide a plausible representation of the underlying causal process for this group.



To simplify the graphical representation of these models, measurement errors and their correlations are not included in Figures 2 and 3, but are depicted separately in Table II. As can be seen from this table, the most obvious misspecification of the model for the LoSlope classes concerns the relationship between the two process indicators: the negative sign (!) for the measurement error in the process variable 2 indicates that the two process variables cannot constitute a common factor for this sample. Some further evidence for this phenomenon can be found in Table I; here it is shown that the patterns of factor loadings for the instruction component differ considerably for both groups.

A comparison of the models depicted in Figures 2 and 3 suggests that the most important differences do concern the estimation parameters of the measurement model, whereas the structural equation model shows some similarities in both groups. Because of the bad data fit of the model for the LoSlope classes, there is no reason to discuss the results in detail. Instead, some further comments should be made on the comparison of the

theoretically assumed model and its empirically confirmed counterpart for HiSlope classes.

There is strong evidence that the proposed model described and explained educational achievement in mathematics for this sample. The simple structure postulated for the measurement model was confirmed for this group. In addition, the assumption that the two exogenous constructs are uncorrelated was empirically supported. Thus one can argue that aptitude and instructional parameters do independently influence math achievement for the HiSlope classes. With regard to the magnitude of the structural coefficients, aptitude indicators seem to have more predictive power for math achievement during the first two measurement periods, compared with the causal influence of the instruction component. But in contrast to the model assumptions made in Figure 1, the two exogenous factors do not contribute to an explanation of the achievement variation measured on the last two occasions. This is probably due to the fact that aptitude indicators and instruction process variables were obtained only during the first measurement period. Because of the different purposes of

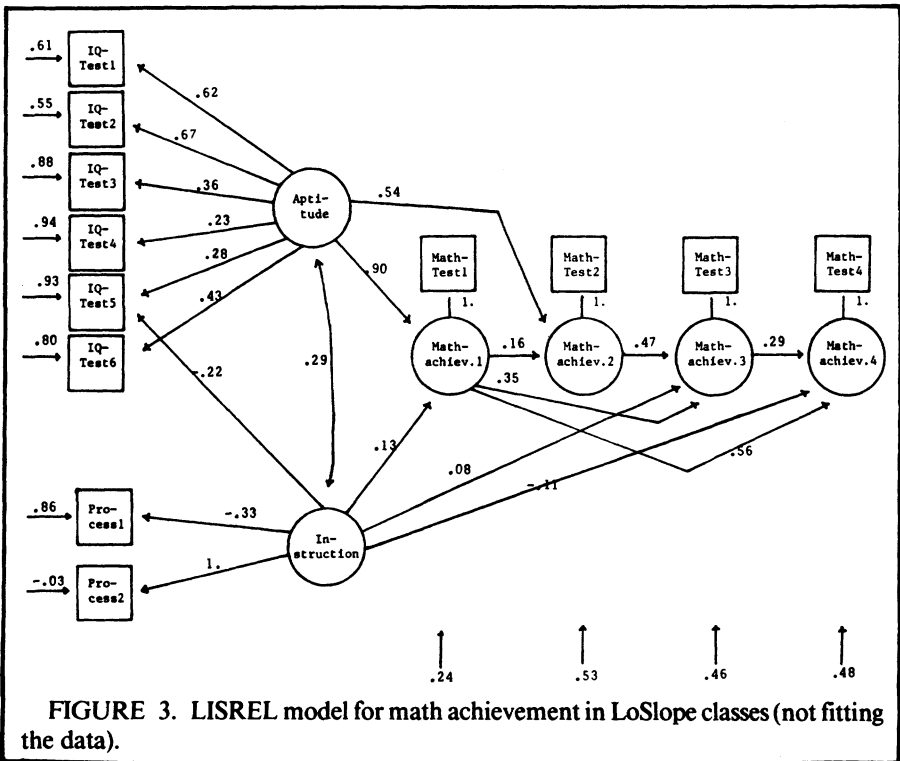


FIGURE 3. LISREL model for math achievement in LoSlope classes (not fitting the data).

TABLE II
Estimates of Measurement Errors and Correlations Between Measurement Errors from LISREL Analyses for HiSlope and LoSlope Classes

HiSlope classes								
	IQ-Test 1	IQ-Test 2	IQ-Test 3	IQ-Test 4	IQ-Test 5	IQ-Test 6	Process 1	Process 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		
Process 1	.00	.00	.00	.00	.00	.00	.71	
Process 2	.00	.00	.00	.00	.00	.00	.00	.00
LoSlope classes								
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		
Process 1	.00	.00	.12	-.22	-.35	-.18	.86	
Process 2	-.17	.00	.00	.00	.00	.11	.00	-.03

Note. Measurement errors are depicted in the main diagonal.

the original investigation (cf. Treiber, 1980), no further data were collected for these variables.

DISCUSSION

The limited empirical research on student achievement, mostly conducted within a cross-sectional, one-sample design, suggested the three themes that were addressed in the present study: (a) specification of achievement growth in developmental models of interindividual differences over time; (b) examination of differential patterns of achievement determination for subgroups of classrooms; and (c) description of the interplay of instructional variables in the determination of achievement.

An attempt was made to provide answers to these research problems by selecting two groups of classrooms with widely differing slopes of their pretest achievement regression, and by observing the resulting differences in the determination of achievement by instructional and ability factors.

The basic assumption was that achievement in classrooms with a steep regression slope may indicate an instructional organization that combines two characteristics: first, a focus on introducing new subject matter and on maintaining high achievement demands for high-ability students; and second, remedial, practice, and review work for low-ability students.

The results of the present study showed that the two classroom subgroups differed in all three aspects: First, the assumed simplex structure of the longitudinal achievement measures was valid only for classes with a high regression slope coefficient (HiSlope). Second, the measurement components of the explanatory model used here did not generalize across both groups of classrooms. Only in HiSlope classes could the observed variables be tied to their respective latent constructs, as expected in the original measurement model. The two instructional process variables seem to be structured differently in the two types of classrooms observed. In HiSlope classes, the introduction of new subject matter seemed to be integrated with extensive review and practice of the material. However, in LoSlope classes, a common Instruction factor could not be constructed. Finally, though the structural components of the explanatory model developed here received empirical support, various modifications had to be made before an adequate model fit was achieved. These modifications stem from the simplex submodel for the four-wave achievement measures and their time-related dependency on ability and instructional determinants. Whereas the latter seem to influence student achievement differences most at occasions 1 and 2, achievement at measurement points 3 and 4 is insufficiently explained, by both the exogenous determinants of aptitude and instruction and the endogenous simplex structure of previous to subsequent achievement.

The present study leaves two problems to future research. First, describing and explaining achievement change over time may require more elaborate structural and measurement models than the apparently all-too-simple type adopted here. Both a more complex multiwave data collection for exogenous and endogenous model variables and a more complicated (i.e., nonlinear, nonadditive, and nonrecursive) specification of their interrelationship may be necessary. Second, conventional explanations are particularly weak for classrooms with flat slopes of their pre-posttest achievement regression. Alternative models are not yet available for classrooms with this type of achievement structure.

The results of this study underline the "local" applicability of achievement models (see Snow, 1977a) and shed doubts on historical attempts to develop general or universal explanations of school learning. Rather, the time- and context-bound nature of achievement processes under given teaching-learning conditions should be the focus of future classroom research.

REFERENCES

- BALTES, P. B., CORNELIUS, S. W., & NESSELROADE, J. R. (1979). Cohort effects in developmental psychology. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development*. New York: Academic Press.
- BENTLER, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology*, 31, 419–456.
- BENTLER, P. M., & WOODWARD, J. A. (1978). A Head Start re-evaluation: Positive effects are not yet demonstrable. *Evaluation Quarterly*, 2, 493–510.
- BIELBY, W. T., & HAUSER, R. M. (1977). Structural equation models. *Annual Review of Sociology*, 3, 137–161.
- BURSTEIN, L. (1979). Issues in the aggregation of data. In D. C. Berliner (Ed.), *Review of research in education* (Vol. 8). Washington, DC: American Educational Research Association.
- BURSTEIN, L., LINN, R. L., & CAPELL, F. J. (1978). Analyzing multilevel data in the presence of heterogeneous within-class regressions. *Journal of Educational Statistics*, 3, 347–383.
- BURSTEIN, L., MILLER, M. D., & LINN, R. L. (1979). *The use of within-group slopes as indices of group outcomes* (CSE Report Series). Los Angeles: Center for the Study of Evaluation, University of California.
- CRONBACH, L. J., & SNOW, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.
- CUNNINGHAM, W. R., & BIRREN, J. E. (1980). Age changes in the factor structure of intellectual abilities in adulthood and old age. *Educational and Psychological Measurement*, 40, 271–290.
- GUTTMAN, L. A. (1954). A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences*. New York: Columbia University Press.
- HELLER, M. S., GÄDIKE, A.-K., & WEINLÄDER, H. (1976). *Kognitiver Fähigkeitstest für 4. bis 13. Klassen* (KFT 4-13). Weinheim: Beltz.
- HORN, W. (1969). *Prüfsystem für Schul- und Bildungsberatung* (PSB). Göttingen: Hogrefe.
- HYLLA, E. (1966). *Aufgaben zum Nachdenken AzN4+*. Weinheim: Beltz.
- JÖRESKOG, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology*, 23, 121–145.
- JÖRESKOG, K. G. (1977). Structural equation models in the social sciences: Specification, estimation, and testing. In P. R. Krishnaiah (Ed.), *Applications of statistics*. Amsterdam: North Holland.
- JÖRESKOG, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, 43, 443–477.
- JÖRESKOG, K. G. (1979). Statistical estimation of structural models in longitudinal-developmental investigations. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development*. New York: Academic Press.
- JÖRESKOG, K. G., & SÖRBOM, D. (1978). *LISREL IV: Analysis of linear structural relationships by the method of maximum likelihood*. Chicago: International Educational Services.

- JÖRESKOG, K. G., & SÖRBOM, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt Books.
- MARUYAMA, G., & MCGARVEY, B. (1980). Evaluating causal models: An application of maximum-likelihood analysis of structural equations. *Psychological Bulletin*, 87, 502–512.
- MUNCK, I. M. (1979). *Model building in comparative education*. Stockholm: Almqvist & Wiksell.
- ROGOSA, D. (1979). Causal models in longitudinal research: Rationale, formulation, and interpretation. In J. R. Nesselrode & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development*. New York: Academic Press.
- SNOW, R. E. (1977a). Individual differences and instructional theory. *Educational Researcher*, 6(10), 11–15.
- SNOW, R. E. (1977b). Research on aptitudes: A progress report. In L. S. Shulman (Ed.), *Review of research in education* (Vol. 4). Itasca, IL: Peacock.
- SÖRBOM, D. (1975). Detection of correlated errors in longitudinal data. *British Journal of Mathematical and Statistical Psychology*, 28, 138–151.
- TREIBER, B. (1980). *Qualifizierung und Chancenausgleich in Schulklassen* (Bde. 1 u. 2). Frankfurt: Lang.
- TREIBER, B. (1982). Modeling achievement changes in classrooms. *Evaluation in education*. Manuscript in preparation.
- WAKENHUT, R. (1974). *Messung gesellschaftlich-politischer Einstellungen mit Hilfe der Rasch-Skalierung*. Bern: Huber.
- WEEKS, D. G. (1980). A second-order longitudinal model of ability structure. *Multivariate Behavioral Research*, 15, 353–365.
- WERTS, C. E., LINN, R. C., & JÖRESKOG, K. G. (1977). A simplex model for analyzing academic growth. *Educational and Psychological Measurement*, 37, 745–756.
- WHITELY, S. E. (1980). Modeling aptitude test validity from cognitive components. *Journal of Educational Psychology*, 72, 750–769.
- WILEY, D. E. (1970). Design and analysis of evaluation studies. In M. C. Wittrock & D. E. Wiley (Eds.), *The evaluation of instruction: Issues and problems*. New York: Holt, Rinehart and Winston.

AUTHORS

- WOLFGANG SCHNEIDER, Research Associate, Max Planck Institute for Psychological Research, Leopoldstr. 24, D-8000 München 40, West Germany. *Specializations*: Educational measurement and data analysis; cognitive development.
- BERNHARD R. TREIBER, Senior Researcher, International Survey Research Corp., Degenfeldstrasse 5, D-8000 München 40, West Germany. *Specializations*: Schooling and cognitive development.