DOCUMENT RESUME

ED 227 148 TM 830 159

AUTHOR Hendrickson, Leslie; Jones, Barnie
TITLE A Study of Longitudinal Causal Models Comparing Gain Score Analysis with Structural Equation Approaches.
PUB DATE Dec 82
NOTE 50p.; Table 2 may be marginally legible due to small print.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Achievement Gains; Achievement Tests; *Data Analysis; Elementary Education; Error of Measurement; Estimation (Mathematics); Goodness of Fit; *Longitudinal Studies; *Models; Multiple Regression Analysis; Path Analysis; Predictor Variables; Reading Achievement; *Research Methodology; Scores
IDENTIFIERS *Causal Models; Gain Scores; LISREL Computer Program; Philadelphia School District PA; *Structural Equation Models

ABSTRACT

The logic of using a gain score approach versus longitudinal causal models is studied in this secondary analysis of a complex data base. The gain score model used by the Federal Reserve Bank and the School District of Philadelphia in their "What Works in Reading?" study is successively refined using the LISREL structural equation program. First the Philadelphia data base is described and then difficulties of using gain score models are discussed. Regression estimates of the different models are described. Procedures dealing with identification, specification, and collinearity are examined. A sensitivity analysis of measurement and specification error shows the degree to which estimated parameters are affected by researchers' assumptions. The reanalysis shows improvements in the understanding of achievement test data and the logic of how to analyze data bases with longitudinal dependent variables. (Author)

A STUDY OF LONGITUDINAL CAUSAL MODELS COMPARING GAIN SCORE ANALYSIS WITH STRUCTURAL EQUATION APPROACHES

December 1982

Leslie Hendrickson, Ph.D.
Research, Analysis & Planning
Eugene School District
200 North Monroe
Eugene, Oregon 97402

Barnie Jones
Department of Sociology
University of Oregon
Eugene, Oregon 97403

ABSTRACT

FOR

A STUDY OF LONGITUDINAL CAUSAL MODELS
COMPARING GAIN SCORE ANALYSIS WITH
STRUCTURAL EQUATION APPROACHES

The logic of using a gain score approach versus longitudinal causal models is studied in this secondary analysis of a complex data base. The gain score model used by the Federal Reserve Bank and the School District of Philadelphia in their "WHAT WORKS IN READING?" study is successively refined using the LISREL structural equation program. First, the Philadelphia data base is described and then difficulties of using gain score models are discussed.

Regression estimates of the different models are described. Procedures dealing with identification, specification and collinearity are exemplified. A sensitivity analysis of measurement and specification error shows the degree to which estimated parameters are affected by researchers' assumptions. The reanalysis shows improvements in the understanding of achievement test data and the logic of how to analyze data bases with longitudinal dependent variables.

TABLE OF CONTENTS

	Page
INTRODUCTION	1
DATA ANALYSIS	6
Gain Score Model Used by Philadelphia Researchers	6
A Structural Equation Model of the Data	9
Analysis and Discussion of Models 1 and 2	14
A STRUCTURAL EQUATION MODEL WITH NON-ZERO ERROR ASSUMPTIONS.	21
Discussion of Model 4	24
Goodness of Fit and Longitudinal Assumptions.	26
Comparison of Models 2 and 4	27
THE ANALYSIS OF MEASUREMENT AND SPECIFICATION ERROR	29
Results of the Sensitivity Analysis.	32
Results of Testing a Model with Realistic Error Assumption	39
CONCLUDING COMMENTS.	38
REFERENCES	40

LIST OF TABLES AND FIGURES

TABLE 1	Code Names, Definitions, Means and Standard Deviations for Eleven Independent Variables and Three Dependent Variables	4
TABLE 2	Correlation Matrix for 14 Variables in Philadelphia Achievement Study	5
TABLE 3	Variance-Covariance Matrix of Gain Score With the Two Achievement Scores	7
MODEL 1	Gain Score	8
MODEL 2	A Longitudinal Model	11
TABLE 4A	Lisrel Estimates of Model 1 and Model 2	15
TABLE 4B	Lisrel Estimates of Model 1 and Model 2 With Insignificant Paths Suppressed	17
MODEL 3	Independent Measurement Model	23
MODEL 4	The Full Model	25
TABLE 5	Comparison of Model 2 With Model 4	28
TABLE 6	Resulting Parameter Estimates Given Error Assumptions	33

INTRODUCTION

Current trends in applied research have witnessed widespread adaptation of multiple regression techniques to program evaluations. While regression analysis is a powerful technique, it owes much of its power to highly restrictive and often unrealistic assumptions. The interpretation of regression results, especially the assessment of the relative impact or importance of independent variables, can be treacherous. There is no assurance that consumers of applied research, in this case school administrators, educational researchers, teachers, or politicians will understand its limitations.

This paper compares methodological procedures used to analyze longitudinal data. It critically compares the use of gain scores to structural equation approaches. The analytic techniques discussed here are applicable to any longitudinal analysis. These general techniques are exemplified in the secondary analysis of data from the "WHAT WORKS IN READING " study conducted by the School District of Philadelphia.

The data base examined in this study is the result of a joint effort in 1977 and 1978 by the Federal Reserve Bank and the School District of Philadelphia to study factors affecting the reading achievement of 1,800 elementary school children. Approximately 8,000 copies of their report "WHAT WORKS IN READING" and its summary have been distributed throughout the world (Kean, Summers, Raivetz and Farber, 1979).

Philadelphia data are reanalyzed using LISREL. LISREL provides an extraordinarily flexible framework for parameter estimation of complex models, and is well adapted to a wide variety of models, including recursive or non-recursive, as well as models incorporating latent structures (Joreskog and Sorbom 1981).

Following an introduction to the data base, the analysis proceeds in three steps. First, specification of the dependent variable is examined. The original report (Kean et al., 1979) treated reading improvement as a net change or gain score. Results of using the gain score as a dependent variable are compared to results when reading at time one and time two are treated as separate dependent variables in a longitudinal model, (see Models 1 and 2).

Second, eleven independent variables are re-examined to incorporate a latent, 10-factor structure and the results of this analysis are compared to the previous results. This analysis exemplifies the use of a factor structure to control for collinearity. Third, the 10 factor model is subjected to a sensitivity analysis (see Land and Felson, 1978) with regard to random measurement error in the dependent variables, and to specification error due to the omission of theoretically important independent variables. This analysis demonstrates how small changes in model specification and residual assumptions can modify results.

Information on 25 reading teachers, 25 principals, 94 teachers, 68 reading aides, and the 1,800 students yielded 245 variables which were analyzed by Philadelphia researchers using multiple regression techniques. The sample selection process was done by school. Average Total Reading Achievement Development Scale Scores (ADSS) on the California Achievement Tests (CAT-70) for 1974 and 1975 were summed over grades 1-4. The 190 schools studied were ranked on the difference of these sums. The final sample contained ten "high-high", five "middle-middle", and ten "low-low" schools which gave representation from all eight administrative sub-districts of the city. The sample totaled 25 schools. The students in these schools were representative of students in schools having high, middle, and low success in reading achievement. Data collection procedures included interviews with school personnel (e.g., principals, teachers, reading aides) and recording data from pupil records. The data collection process was completed in two weeks, (Kean, Summers, Raivetz and Parker, 1979).

Over 500 multiple regression runs were conducted to establish which of the 245 variables measured had the most impact on the reading achievement gain. Eighteen variables were identified as contributing to achievement gain. "WHAT WORKS IN READING?" points out the difficulties of analyzing this complex data base without a theory. The Philadelphia technical report consists almost entirely of descriptions of variables and correlation matrices. Cross-tabulations, path analysis, or modeling of the 245 variables were not undertaken, and neither the relations among the 18 variables nor their impact on third and fourth grade test scores were analyzed.

As mentioned above, "WHAT WORKS IN READING?" found 18 of the 245 variables studied to have a statistically significant beta weight in predicting the gain score. Table 1 lists definitions, means, and standard deviations for eleven of the independent variables, and for the three dependent variables: the gain score, and the third and fourth grade reading scores. The eleven independent variables include measures of student, teacher, and school organization.

Table 2 shows the 14x14 correlation matrix of the variables listed in Table 1. The impression obtained from Table 2 is that the matrix is thin. Of the 90 correlations in it, 17 or 19% are greater than .15, and 12 or 13% are greater than .25. Table 2 contains 55 correlations among pairs of the 11 independent variables and 5, or 9%, are greater than .25. The highest correlation of any variable with CATGAIN, the gain score, is .08.

TABLE 1

Code Names, Definitions, Means, and Standard Deviations for Eleven Independent Variables and Three Dependent Variables

Code Names	Definitions	Mean	Standard Deviation
$T_1 - T_2$	Difference between Grade 3 and Grade 4 scale score	28.43	52.50
T_1	California Achievement Test--Reading Comprehension Scale Score for Grade 3, 1975	385.06	67.74
X_1	Days Students were present	130.51	10.41
X_2	Student attended kindergarten 1=NO, 2=YES	1.80	0.40
X_3	Number of non-teaching supportive staff per school	----*	11.02
X_4	Percent of students scoring above 84th percentile California Achievement Test 1976--Total Reading	.20	.13
X_5	Percent of classroom teachers with less than 2 years experience	.197	.136
X_6	Number of teacher pay periods with no absence	13.89	3.79
X_7	Teacher attends outside professional conference meetings 1=NO, 2=YES.	1.17	.39
X_8	First year teaching grade 4 1=NO, 2=YES	1.17	.35
X_9	Minutes per week of individual independent reading	73.35	60.31
X_{10}	Teacher would select the same reading program again	1.54	.50
X_{11}	Times per week aide in room during reading	2.55	2.31
T_2	California Achievement Test, Reading Comprehension Scale Score for Grade 4, 1976	412.50	72.56

*Means for X_3 were not shown in the November 1979 technical report of "WHAT WORKS IN READING?"

TABLE 2

CORRELATION MATRIX FOR 14 VARIABLES IN PHILADELPHIA ACHIEVEMENT STUDY, N 1,363

	Difference between Gr 3 & Gr 4 scale score $T_2 - T_1$	CAT Read Comprehen Scale Score Gr. 3-1975 T_1	Days Student Present X_1	Student Attend Kindergtn X_2	# of Non-teaching Support Staff X_3	% Students Above 84th Percentile CAT - 1976 X_4	% Class Teachers With Less Than 2 yr. Experience X_5	# Teacher Pay Periods With no Absence X_6	Teacher attends outside conference X_7	1st year teaching GR. 4 X_8	Min per week individ. independ. reading X_9	Teacher select same reading program X_{10}	Aide time during reading each week X_{11}	CAT-Reading Comp. Scale Score - 1976 Grade 4 T_2
$T_2 - T_1$	1.000													
T_1	-.292	1.000												
X_1	.074	.161	1.000											
X_2	.020	123	.116	1.000										
X_3	-.004	-.290	-.134	-.145	1.000									
X_4	-.051	.386	134	.141	-.626	1.000								
X_5	.071	-.118	-.056	-.095	.327	-.154	1.000							
X_6	.042	076	-.017	-.006	-.135	-.040	078	1.000						
X_7	-.037	-.112	-.033	.004	-.027	-.082	.118	.005	1.000					
X_8	-.002	-.139	-.061	.013	.104	.007	.021	.021	.021	1.000				
X_9	083	.090	.086	.034	-.143	.113	.005	-.004	-.125	.021	1.000			
X_{10}	-.007	.187	-.029	-.006	-.142	.030	-.033	.080	.023	-.011	-.119	1.000		
X_{11}	.004	-.362	-.116	-.110	.527	-.427	.074	.010	.254	-.012	-.069	-.106	1.000	
T_2	N.A	.722	.197	.129	-.273	.382	-.065	.101	-.132	-.132	.144	.169	.335	1.000

DATA ANALYSIS

Gain Score Model Used by Philadelphia Researchers

The regression runs done by Federal Reserve Bank economists used the difference between the third and fourth grade reading achievement scores as a single dependent variable. The use of "difference", "change", or "gain" scores has been thoroughly examined (Thorndike and Hagen, 1955; Thorndike, 1966; Bohrnstedt, 1969; Cronbach and Furby, 1970; Alwin and Sullivan, 1975; Kim and Mueller, 1976; Kessler, 1977; Pendleton, Warren and Chang, 1979). These examinations have generally advised against using gain scores because: the difference between the two measures has lower reliability than the measures considered separately; their use requires low error variance and high reliability in variables; the calculation of the gain score reliability tends to be unstable because it depends on five other values--the three correlations and two variances; and the analysis of gain scores is complicated by the effects of regression toward the mean.

Similarly, Thorndike (1963:40, 1966:124), points out two characteristics of using gain scores. First, the gain score is almost certain to be negatively correlated with the initial achievement score. Second, the variance of the gain scores is in some cases no more than one-fourth the size of the variance of the Time 1 and Time 2 scores. Table 3 illustrates Thorndike's comments. Despite these disadvantages, gain scores continue to be used in applied research and evaluation work (Alwin and Sullivan, 1975).

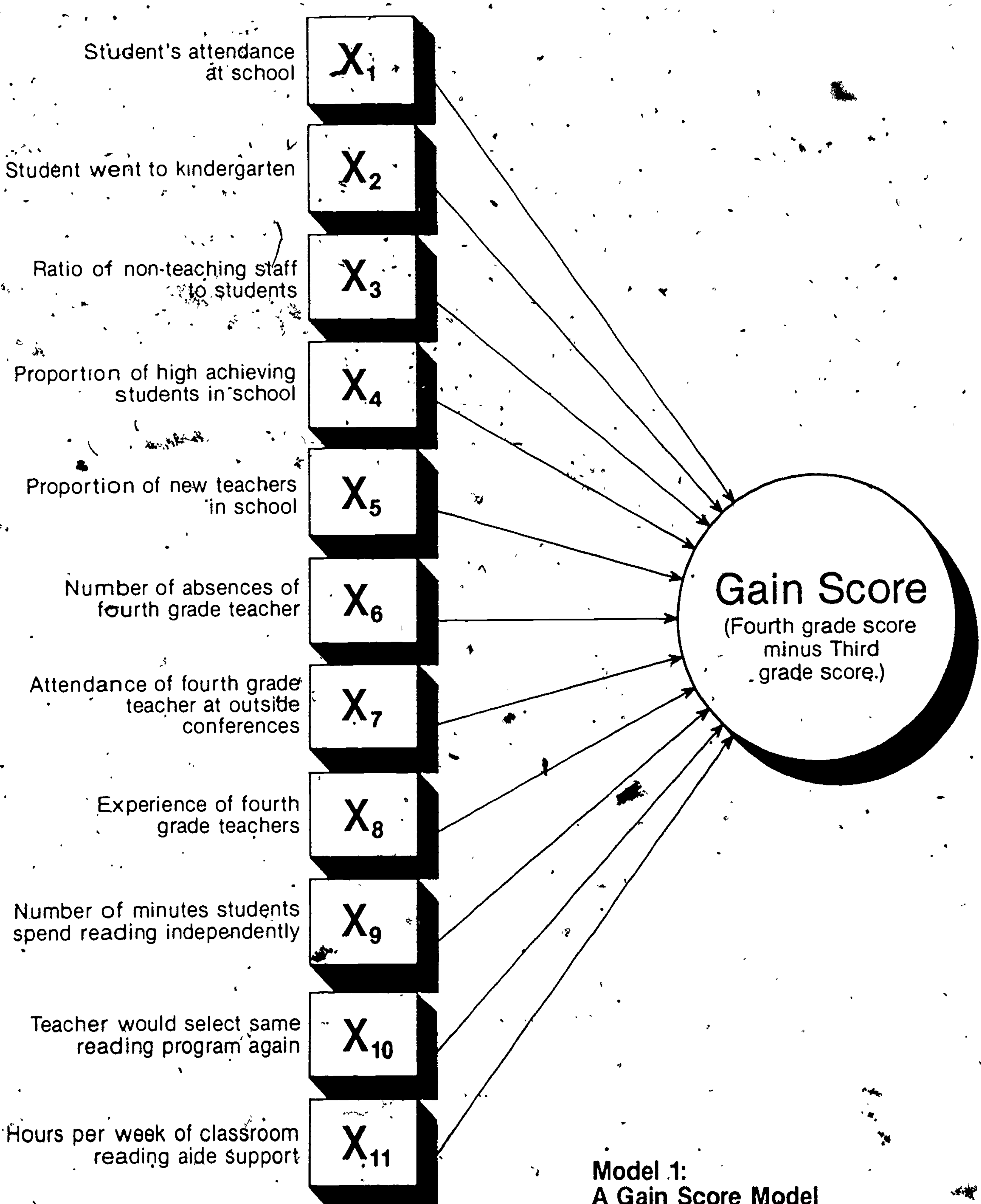
Table 3 shows the variance-covariance matrix of the gain score and the third and fourth grade achievement scores. Thorndike's comments (see Thorndike,

TABLE 3

Variance-Covariance Matrix of Gain Score with the Two Achievement Scores

N = 1,363

	GAINSCORE	GRADE 3 READING SCORE	GRADE 4 READING SCORE
GAINSCORE	2,756.25	------	------
GRADE 3 READING SCORE	-1,039.52	4,588.71	------
GRADE 4 READING SCORE	Not calculated in Philadelphia Study	3,548.78	5,264.95



Model 1:
A Gain Score Model
No time assumptions—all independent variables are assumed to affect a single dependent variable. All error terms are assumed to be zero.

1963:40, 1966:124) are appropriate here. As anticipated, the gain score is negatively correlated with the initial score and the covariance of the gain score is 60% the size of the Time 1 variance and 52% the size of the Time 2 variance.

An additional problem with the gain score is that it obliterates information. For example, in analysis of reading achievement scores, it could be hypothesized that reading level and reading gain are associated. However, computation of a gain score eliminates data on the student's reading level at either time. In this case, for example, we find a negative correlation between third grade reading level and reading gain from grade three to grade four, indicating that lower students tended to improve more rapidly than higher level students. Consequently, factors associated with high gain may also contribute to low overall achievement. This makes analysis of a gain score difficult to interpret.

Model I shows the gain score model. All 11 independent variables are assumed to influence the gain score and are assumed to be measured without error. The gain score model was analyzed using the just-identified multiple regression option of LISREL.

A Structural Equation Model of the Data

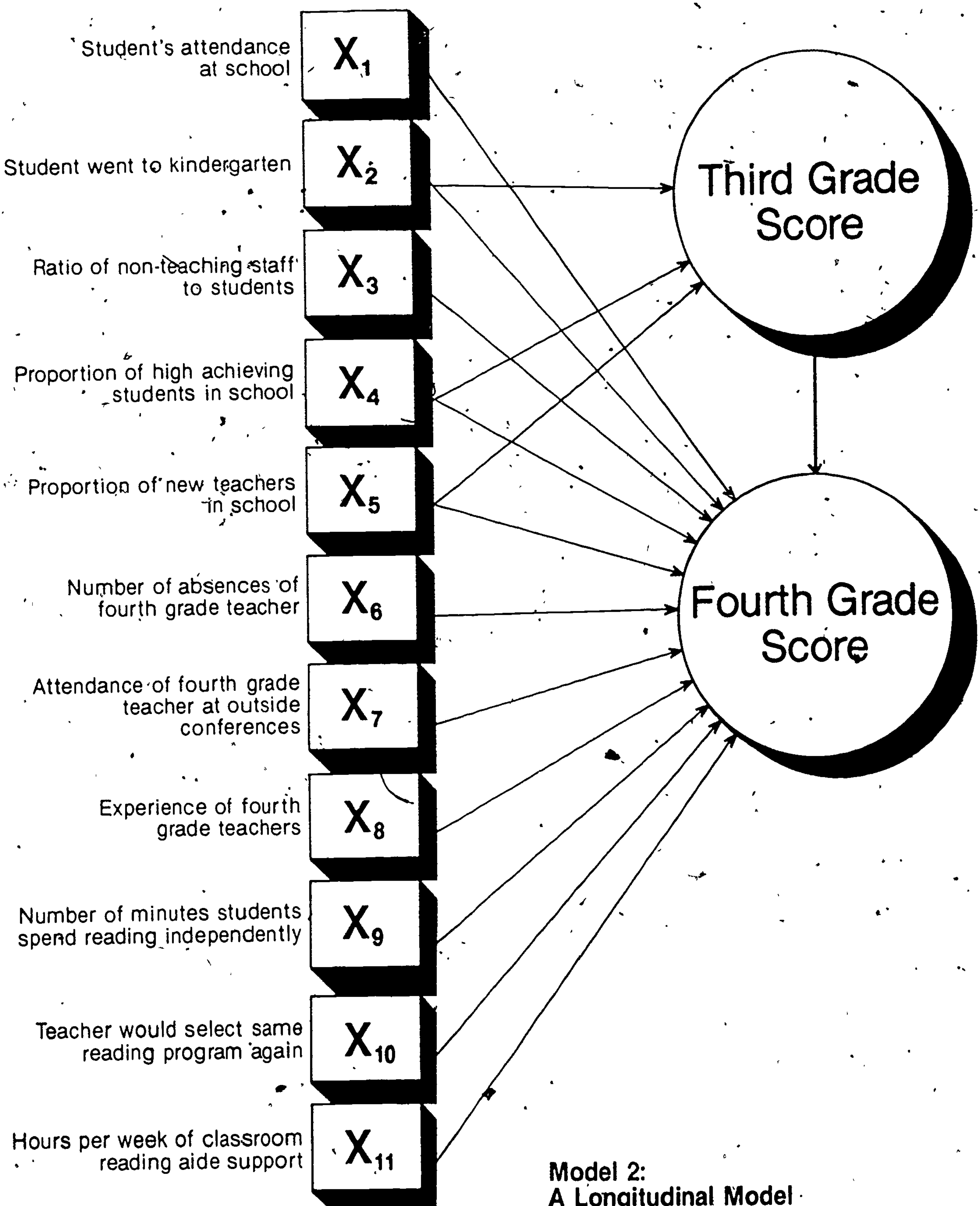
It is theoretically reasonable that the gain score analysis can be extended by considering alternative models that use data from both times rather than using the difference score as a single dependent variable. The analysis below used the maximum likelihood exploratory factor analysis (EFAP) and structural equation programs, LISREL IV and V, of Joreskog and Sorbom (1981).

Maximum likelihood estimation procedures, originally developed by the British statistician R. A. Fisher (1921), yield estimations which are efficient and consistent for large samples. These approaches were introduced to sociologists in the middle 1970's (Hauser and Goldberger, 1971; Burt, 1973). Two-, three-, and four-wave multi-variable models have been extensively studied with these approaches (See, for example, Duncan, 1972; Hannan and Young, 1977; Hargens, Reskin and Allison, 1976; Long, 1976; Joreskog and Sorbom, 1977; and Wheaton, Muthein, Alwin and Summers, 1977).

Model 2 shows one alternative model for analyzing the Philadelphia data using the two dependent variables, Time 1 and Time 2, together instead of analyzing their difference. Two multiple regression runs were made using the 11 variables; first the third grade achievement variable was used as the dependent variable then the fourth grade variable was used.

A review of "WHAT WORKS IN READING?" and conversations with School District of Philadelphia research and evaluation staff showed that 3 of the 11 variables can be hypothesized to influence both the third and fourth grade scores while the other 8 can be hypothesized to influence only the fourth grade score. The 3 variables influencing scores at both times were whether the student went to kindergarten (x_2), and the proportion of students in the school scoring well on the achievement test (x_4), and the proportion of new teachers ([illegible]).

Model 2 contains two structural equations. The first uses only the independent variables affecting the third grade score. The second uses all 11 independent variables plus the third grade score. In this model, error terms are also assumed to have expectations of 0.



Model 2:
A Longitudinal Model

Three independent variables are assumed to affect Time 1, the third grade score. All variables and the third grade score are assumed to affect Time 2, the fourth grade score. All error terms are assumed to be zero.

The LISREL IV computer program was used to analyze the covariance matrix. Briefly, the LISREL model consists of two parts: the measurement model and the structural model. The measurement model specifies how latent or hypothetical constructs are measured in terms of observed variables. There are two measurement models, one for dependent variables, and one for independent variables. Let $x' = (x_1, x\ , \ldots x_q)$ be a vector of observed independent variables and let $y' = (y\ , y\ , \ldots y_p)$ be a vector of observed dependent valuables. Then,

$$x = \Lambda_x \quad + \qquad (1)$$

$$y = \quad_y \quad + \qquad (2)$$

where and are random vectors of latent independent and dependent variables, $\quad' = (\quad_1 \ldots \quad)$ and $\eta' = (\eta_1 , \eta_2 \ldots \eta_m)$ respectively. The vectors and are errors of measurement in y and x respectively when y and x are measured as deviations from their means. The matrices Λ_x (q x n) and $\quad_y$(p x m) are regression matrices of (x on and y on η) respectively.

The structural model linking the two measurement models is given in (3).

$$\eta_1 = \quad + \qquad (3)$$

Where and are coefficient matrices and $\zeta = (\zeta_1 , \zeta_2 \ldots \zeta_m)$ is a random vector of residuals reflecting disturbance terms or errors in equations (Joreskog and Sorbom, 1978).

The two equations comprising the model in Figure 2 are given in (4) and (5).

$$\eta_1 = \Gamma_1 \xi + \zeta_1 \quad (4)$$

$$-\beta_{21} \eta_1 + \eta_2 = \Gamma_2 \xi + \zeta_2 \quad (5)$$

η_1 the third grade score is seen to be comprised of two parts, $\Gamma_1 \xi$ which represents the effects of the three independent factors and ζ_1 which is the residual variance unexplained by the three variables affecting the third grade score. Equation (5) can also be rewritten as (6).

$$\eta_2 = \beta_{21} \eta_1 + \Gamma_2 \xi + \zeta_2 \quad (6)$$

η_2 the fourth grade score is seen to be comprised of three parts: $\beta_{21} \eta_1$ which represents the effect of η_1 on η_2; $\Gamma_2 \xi$ which is the effect of the independent variables; and ζ_2 which is the unexplained residual variance of η_2.

This model is recursive in that the fourth grade score is assumed to have no effect on the third grade score. Identification problems in recursive models have been frequently commented on (Heise, 1969, 1970). In order to make recursive models identifiable, restrictive assumptions are generally made about error terms. For example, in order to identify the Model 2 the usual procedure is to assume that all error terms and covariances of error terms are zero and uncorrelated with each other. This includes assuming the covariance of the error terms, ζ_1 and ζ_2 in equations (4) and (5) is equal to zero.

The assumption of zero error terms and zero error covariances is equivalent to assuming that all variables are measured without error. On the independent variable side of the model, in LISREL terminology, this situation is called "fixed X"; Λ_x is assumed to be an identity matrix, the independent variable error matrix $\Theta_\delta = 0$ and $x = \xi$. Similar assumptions are made on the dependent variable side, i.e., $\Lambda_y = I$, $\Theta_\epsilon = 0$ and $y = \eta$. Given these assumptions, the structural equation model becomes the following.

$$\eta_y = \Gamma x + \zeta \tag{7}$$

Analysis and Discussion of Models 1 and 2

Table 4A presents estimates for Model 1 and Model 2. There are dramatic differences in the estimates obtained in magnitude and sign. Note also, that with the exception of the effect of T_1 in equation 2 of Model 2, most effects are small. In Model 1, the gain score model, eleven independent variables explain just over 2.5 percent of the gain score variance. With effects of this magnitude it is hopeless to draw substantive conclusions of consequence. However, in order to increase efficiency and remove clutter, a second set of estimates (shown in Table 4B), were calculated, fixing insignificant effects at 0. In terms of goodness of fit and explained variance, dropping insignificant effects had little adverse consequence, indicating that information loss was trivial. At the same time, the reduced complexity of the models makes them easier to compare.

Estimates were prepared, suppressing to 0 the effects of X_4, X_5 and X_7 in equation 2. In this set of estimates, the effect of X_5 in equation 1

TABLE 4A

LISREL ESTIMATES OF MODEL 1 AND MODEL 2

	Dependent Variables		
	MODEL 1	MODEL 2	
Independent Variables	$T_2 - T_1$	T_1	T_2
X_1	.373 (.074)		.579* (.083)
X_2	2.587* (.020)	10.445*(.062)	4.183 (.024)
X_3	-.407 (-.085)		.434* (.066)
X_4	-44.753 (-.111)	190.846*(.366)	63.114* (.113)
X_5	31.111 (.081)	- 24.782*(-.050)	3.733 (.007)
X_6	.635* (.046)		1.064* (.056)
X_7	-3.829* (.028)		-4.075 (-.022)
X_8	1.803* (.012)		14.180* (-.068)
X_9	.062 (.071)		.123* (.102)
X_{10}	-.625* (-.006)		12.267* (.085)
X_{11}	+.010* (.000)		-3.642* (-.116)
T_1			.653* (.610)
$(1 - \Psi)$	.0264	.195	.469
χ^2/d.f.	0003/0	277.81/8	

* Significance at less than .05

Figures in parentheses are standardized estimates was also found to be nonsignificant, so a second set of estimates were prepared, setting at 0, X_5 in equation 1. In this second set of estimates, the X in equation 2 shifted just out of the critical region.

First, the gain score model yields quite a different picture than the 2 wave model. The variables X_2 and X_7 are non-significant in Model 1 and in equation 2 of Model 2. X_7, the teachers attendance at outside conferences, is an ambiguous measure. It may measure level of professional interest and awareness, but it may also measure teacher absence from the classroom, or a desire for upward professional mobility, i.e., to get out of the classroom.

X, Kindergarten, is an interesting variable since it is non-significant in Model 1 and in equation 2 of Model 2, but significant in equation 1 of Model 2. Kindergarten experience has an indirect effect which is missed altogether in the gain score model.

The variable, X_5, Teacher experience, is significant in Model 1 but not in Model 2. In other words, students of experienced teachers show more improvement than students of inexperienced teachers, but when we control for reading competence at Time 1, the teacher experience makes no difference in reading competence at Time 2. The effect of Model 1 could represent a difference in assignment. It would seem reasonable that the school would take teacher experience into account in making classroom assignments.

Four teacher and classroom variables, X_6, X_8, X_{10}, and X_{11}, are non-significant in Model 1, but are significant in Model 2. It seems their effects show when assignment is taken into account.

Three remaining variables, X_1, X_4, and X_9 are significant in both models. However, only X_1 and X_9 agree in both models. It is interesting to note that X_1 and X_9, along with X_2, are the only independent variables

TABLE 4B

LISREL ESTIMATES OF MODEL 1 AND MODEL 2
WITH INSIGNIFICANT PATHS SUPPRESSED

	Dependent Variable					
	MODEL 1		MODEL 2			
Independent Variable	$T_2 - T_1$		T_1		T_2	
X_1	.383	(.075)			.469	(.067)
X_2	0†	0†	10.999	(.065)	0†	0†
X_3	-.344	(-.071)			.303	(.045)(not significant)
X_4	-40.351	(-.100)	198.813	(.381)	64.051	(.115)
X_5	32.349	(.084)	0†	0†	0†	0†
X_6	0†	0†			.860	(.045)
X_7	0†	0†			0†	0†
X_8	0†	0†			-8.139	(-.039)
X_9	.067	(.076)			.102	(.085)
X_{10}	0†	0†			7.604	(.052)
X_{11}	0†	0†			-2.065	(-.065)
T_1					.680	(.635)
$1 - \Psi$	.024		.195		.463	
χ^2/d.f.	4.407/6		272.712/12			

† Fixed at 0

Figures in parentheses are Standardized estimates measured at the student level. All others are at the classroom and school level. The interpretation of X_1, student attendance, and X_9, time in the classroom spent reading independently, is straight forward. Students who come to school more, and spend more time reading while at school, can read better at the end of the year. Not very profound but a finding nontheless.

The variables, X_3 and X_4, supplementary staff and proportion of high scoring students in the school, share a positive sign in Model 2 and a negative sign in Model 1. Looking first at X_4, students in schools including a high proportion of high scoring students do not improve as much as students in schools with a lower proportion of high scoring students; but when reading level at time 1 is controlled, students in schools having a high proportion of high scoring students score higher at time 2.

This is a complex association. The variables, X_3 and X_4 are highly correlated negatively, -.626. Considering just Model 2, they have opposite signed correlations with the dependent variable, but their effects in Model 2 have the same sign. Substantively, it seems that X_4 is measuring the level of reading competence in the school. It is arguable that what is being measured is the socioeconomic level of the school; middle and upper middle class students tend to have higher levels of scholastic success than working and lower class students.

In either case, Model 2 suggests that since supplemental staff persons are assigned on the basis of need, schools with low general levels of competence will receive more staffing resources, accounting for the high negative correlation between X_3 and X_4. One could call this an allocation effect. Consequently, X_3 has a negative correlation with T_2, because of this allocation effect; but when X_3 and X_4 are entered in the same equation, the partial effect of X_4 is positive, suggesting that when the allocation effect of staffing is controlled, the effect on reading levels is positive.

In Model 1 the effects of both X_4 and X_3 on the gain score are negative, leading to the conclusion that supplementary staffing has a detrimental influence. Instead, what apparently is operating is a negative association between gain and initial competence level. Low students have higher gains, perhaps because of a ceiling effect, i.e., less room for improvement and perhaps also because of the allocation effect of supplementary staff, i.e., less concentrated instruction.

The foregoing interpretation seems satisfactory except that it is contradicted by the effect of X_{11}. The variable X_{11}, time in classroom of reading aides, should be expected to parallel the effects of X_4, to the extent that both measure supplementary staffing. The only major difference between the two is that X_{11} is measured at the classroom level. There are two possible interpretations: first, it may be that a collinearity effect is distorting the effect of X_{11}, since it is correlated with both X_3, .527, and X_4, -.427. This may also explain why X_{11} becomes nonsignificant (see Table 4B). Second, it may be that there is a true negative component in X_{11}. For example, it has been suggested (Conant, 1971) that classroom aides may be misused, supplanting rather than supplementing instruction from more highly trained and qualified teachers. In other words, teachers who make the highest use of aides may be over-relying on aides.

In summary, Model 2 tends to produce a pattern of effects which come closer to matching reasonable expectations about reading achievement.

consistently, reversals in signs of effects of independent variables suggest that performance of a particular student depends largely on that student's starting point, so that when a student's starting point is taken into account a clearer picture is obtained of the factors contributing to his or her progress.

Model 2 accounts for approximately 20 percent of the Time 1 variance and 45 percent of the Time 2 variance. This is an improvement over the miniscule amount of gain score variance accounted for by Model 1. At the same time it must be emphasized that effects are small in both models and may be, although statistically significant, substantively trivial. For example, Model 2 indicates that each day of absence from the classroom results in an expected loss of half a point on the CAT reading achievement test, when the standard deviation of that test is 72.56. Model 2 also indicates that each minute spent per week at independent reading results in an expected increase of one-tenth point on the CAT (.102). Increasing that time by an hour per week would amount to a six point improvement, not a very large payoff.

These findings must be viewed in the context of specification. We have seen how strongly the sign and magnitude of effects can be altered when new information is added. The addition of other variables would probably alter estimates. The low percentage of variance explained suggests that there must be other major influences on reading abilities which have not been taken into consideration.

A STRUCTURAL EQUATION MODEL WITH NON-ZERO ERROR ASSUMPTIONS

Model 2 leaves a major substantive issue unresolved concerning the lack of symmetry of the effects of X_3 and X_{11}. Both relate to staffing variables and should have parallel effects. Three highly intercorrelated variables, X_3 X_4 and X_{11}, were analyzed using confirmatory factor analyses to explore an exhaustive range of factor structures. A two factor structure as in Model 3 produced the most satisfactory fit.

Initially, specification of the measurement model attempted to load all three variables onto one factor. These attempts resulted in unsatisfactory goodness of fit diagnostics, leading to the present loading of three variables on two factors. Identification of these correlated disturbances was accomplished by placing arbitrary constraints on Phi, the matrix of correlations among factors. Model 3 had excellent goodness of fit indicators, e.g. a probability level of .15. The goodness of fit was also improved in increments by specifying small, correlated error terms among independent variables.

The Model 3 structure suggests that X_3 and X_{11} have different but overlapping underlying effects. The association of X_4 with both factors indicates that the allocation effect of supplementary staffing discussed in the previous section, applies to both staffing variables. The correlation between the two factors is .754, indicating that while there is substantial overlap, each factor is to a degree unique. If the lack of symmetry between X_3 and X_{11} had been due to a distortion from collinearity, it should have been possible to load all three variables on one factor. The resistance of the covariance structure to a single factor structure is taken as empirical evidence that lack of symmetry is not due to collinearity, but that the effect of X_{11} is indeed unique from that of X_3.

Employment of confirmatory factor analysis in this way creates a "measurement model" relating all eleven independent variables to ten independent factors.

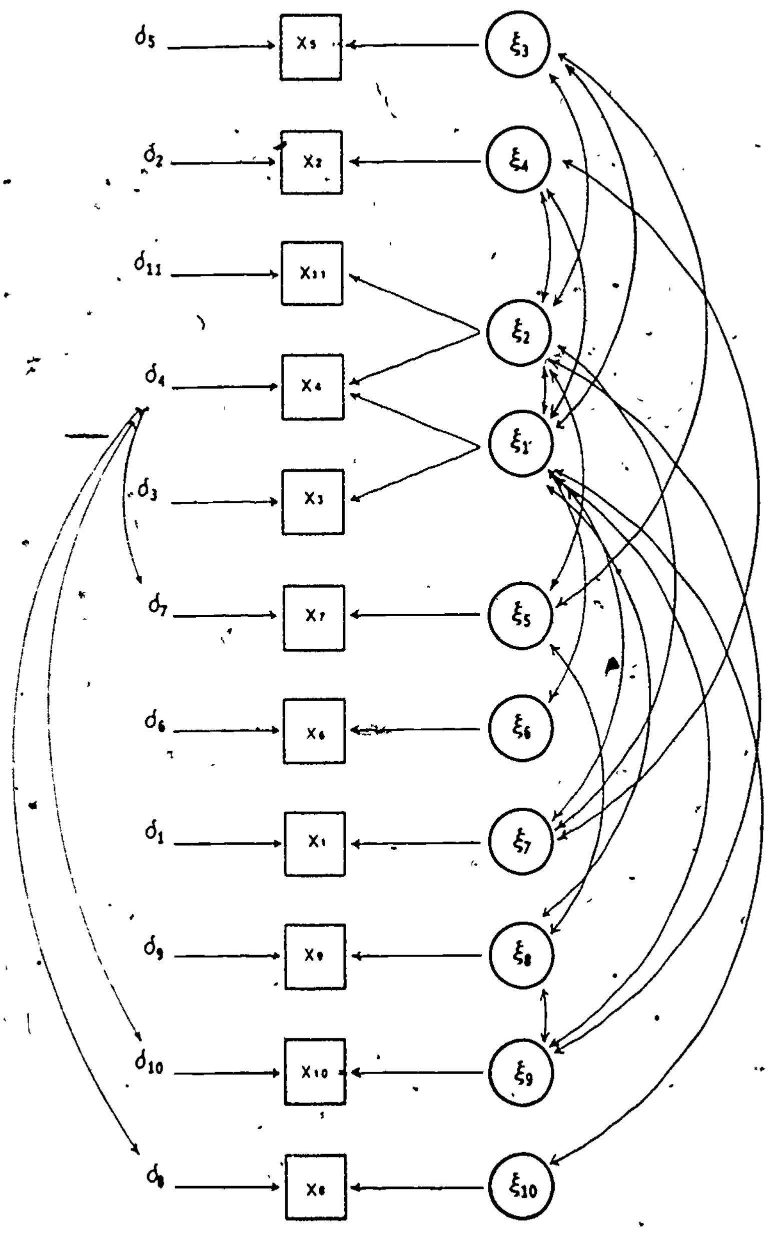
The measurement model hypothesizes an observation to be composed of a true value attributable to the thing observed, and measurement error as described earlier in equation 1, $\underset{\sim}{X} = \underset{\sim}{\Lambda}_X \underset{\sim}{\xi} + \underset{\sim}{\delta}$ where $\underset{\sim}{X}$ is a matrix of independent observed variables; $\underset{\sim}{\Lambda}_X$ is a matrix of factor loadings showing how much of each variables' variance loads on each factor; $\underset{\sim}{\Phi}$ is a variance-covariance matrix of the unobserved independent factors and $\underset{\sim}{\delta}$ is a matrix of the errors in measurement.

Equation (8) shows the unstated independent measurement model of a typical multiple regression approach.

$$\underset{\sim}{X} = \underset{\sim}{\xi} \qquad (8)$$

Since the errors are assumed to be zero, $\underset{\sim}{\delta}$ drops out of equation (1). Since multiple regression programs usually assume the number of variables equals the number of factors, $\underset{\sim}{\Lambda} X$ is an identity matrix. Thus ξ is shown to be equal to the variance-covariance matrix of the independent variables.

Models estimating correlated measurement error are useful especially given a set of theoretically linked and collinear multiple-indicators. The logic of procedures used to obtain models with realistic error estimates can also be inverted by fixing measurement error along a range of hypothetical levels in order to observe the sensitivity of other parameters to measurement errors. Examples of this will be shown later in the paper.



Model 3:
Independent Measurement Model
Does not assume errors are zero or uncorrelated. Shows detail of factor structure.

Discussion of Model 4

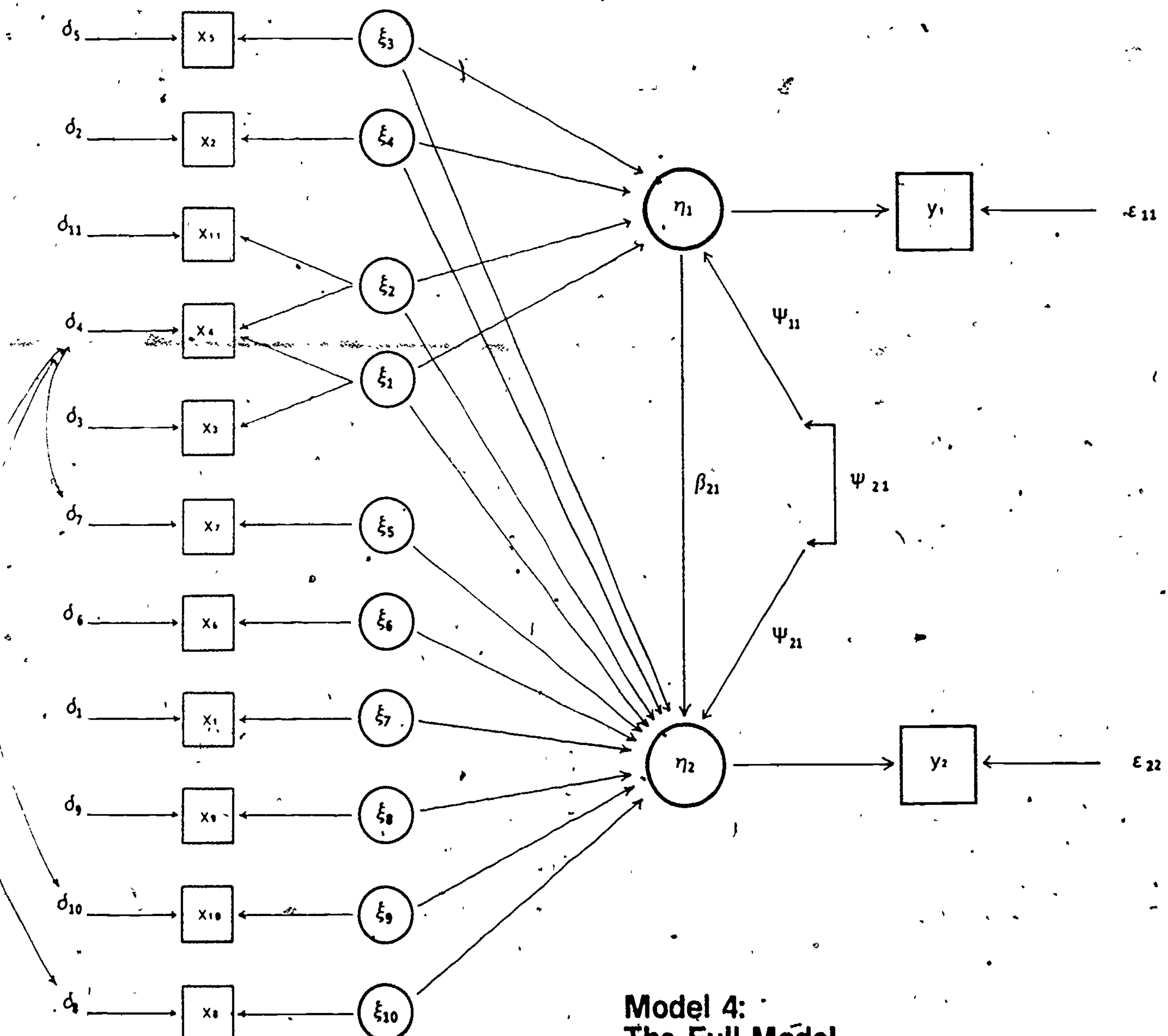
Model 4 includes the independent measurement model shown in Model 3 with the exception that the links among the factors shown in Model 3 are not shown in Model 4 to simplify its presentation. The full model, as described in equation 3 and exemplified in Model 4, links the independent measurement model to the dependent measurement model.

Five additional matrices are now used. One of them, gamma (Γ), links the independent and dependent models. Three matrices, lambda (Λ_y), beta (β), and theta epsilon (Θ_ϵ), describe the dependent measurement model and one matrix, psi (Ψ), gives the errors made in predicting values of the dependent factors.

Row 1 of gamma represents the effects of the independent factors on Time 1, and row two is their effects on Time 2. Three of the eleven independent variables were hypothesized to have an effect on the time one reading score, but one of those loaded on two factors. Consequently, four of the ten parameters in row 1 were estimated, and the remaining six were assumed to be zero. Since all eleven variables, and therefore all ten factors are hypothesized to affect the Time 2 reading score, all ten gamma coefficients were estimated in row 2.

Beta is a 2x2 square matrix of coefficients relating the two dependent factors. The diagonal elements represent the effect of each factor on itself. β_{12} represents the effect of Time 2 on Time 1, and is, therefore assumed to be 0. β_{21} represents the effect of Time 1 on Time 2, and is estimated.

Psi is a 2x2 symmetric matrix of errors in equations, the diagonal elements represent the unexplained variance in each dependent factor after all



Model 4:
The Full Model

Details of effects between independent factors omitted for clarity. See Model 3 for details.

the independent variables have their effect calculated. The single off diagonal element, ψ_{21}, is the correlation between the unexplained variance of each factor. This is usually assumed to be zero, however, if it is reasonable to assume that omitted variables are operating (Zellner 1963), then ψ_{21} will not equal zero and will upwardly bias the estimation of beta.

Goodness of Fit and Longitudinal Assumptions

The LISREL IV program provides an χ^2 measure of goodness of fit, as an indicator of distortion introduced by over-identifying restrictions. However, the χ^2 value is misleading for large samples because its magnitude is a function of sample size. There is a serious danger of over-fit when restrictions too diligently refine a model to minimize χ^2.

Some researchers prefer the Tucker-Lewis statistic which is based on the ratio of the sigma matrix to the covariance matrix, and is therefore independent of sample size (see Tucker and Lewis, 1973; Knoke, 1979).

In Model 3, the χ^2 value obtained was quite low, 20.79 with 15 degrees of freedom and a probability of occurrence of .143. This is very low and may suggest overfit. However, in Model 4, the χ^2 value takes a large step upward, to 212.34. This suggests that over identifying restrictions in the gamma or beta matrices are at fault. Freeing all parameters in the gamma matrix results in an χ^2 value of 43.73 indicating that over-identifying restrictions in equation 1 are responsible for most of the distortion picked up by χ^2.

It is not important to minimize χ^2, however it is important to balance goodness of fit against over-identifying restrictions. In this case it would appear that the structure of the model does not allow for a cross-legged effect of reading achievement at Time 1 on classroom assignment in the coming year.

However, these effects, when estimated, are trivial in magnitude. Consequently, it seems unwarranted to revise the model in order to incorporate them.

Comparison of Models 2 and 4

Table 6 compares estimates obtained from Model 2, to those obtained from Model 4. The more complex factor structure has increased the amount of variance explained in the Time 1 and Time 2 variables. Of major interest are the effects of the two factors associated with X_3, X_4, and X_{11}. Factor 1, which is influenced by X_3 but not X_{11}, has a positive effect greater than the effect of either X_3 or X_4 in Model 2. Factor 2, influenced by X_{11} but not X_3, has a very large negative effect which is considerably larger than the negative effect of X_{11} in Model 2. These changes are also reflected in a large change in the proportion of variance explained in fourth grade reading, .46 in Model 2, and .55 in Model 4. These results indicate that there may well be a detrimental effect resulting from inappropriate use of reading aides.

There are other notable changes in estimates. Teacher attendance, X_6, has a small but significant effect in Model 2 but an insignificant effect in Model 4. Teacher's attendance at outside conferences, X_7, is non-significant in Model 2, but has a small, significant effect in Model 4. Teachers approval of the reading program, X_{10}, is significant in Model 2 but not in Model 4. Since these are substantively miniscule effects in either case, one hesitates to draw conclusions, but it is consistent to suggest that a teacher's performance may have a great deal to do with his or her effectiveness in using aides. By drawing out the effect of over-reliance on aides, effects of other teacher and classroom variables were bound to shift.

TABLE 5

COMPARISON OF MODEL 2 WITH MODEL 4

Independent Variable	Model 2 EQ.1	Model 2 EQ.2	Model 4 EQ.1	Model 4 EQ.2
X_1		.083*		.044*
X_2	.062	.024	.023*	.012
X_3		.066*		
Factor 1			.311*	.221*
X_4	.366	.113*		
Factor 2			-.737*	-.396*
X_{11}		-.116*		
X_5	.050	.007	-.125*	-.012
X_6		.056*		.020
X_7		-.022*		.060*
X_8		-.068*		-.043*
X_9		.102*		.111*
X_{10}		.085*		.025
T_r		.610*		.553*
1 -	.195	.469	.31	.55
χ^2/d.f.	277.81/8		212.338/18	

*Significance less than .05

All figures are standardized estimates

THE ANALYSIS OF MEASUREMENT AND SPECIFICATION ERROR

To this point we have considered three different models. Regardless of which is considered the best, our tinkering has resulted in changes in sizes and magnitudes of effect coefficients, and also led to changes in interpretation. The kinds of applied policy recommendations derived would depend on which model was considered.

However, specification of the causal arrangement of variables is only one of many choices that can influence the magnitude of effect coefficients. A major area of skepticism concerning multivariate models is in regard to the disturbances on error terms.

Two major classes of error concern multivariate models: errors in variables and errors in equations. Errors in variables generally refers to validity and reliability of observation. Errors in equations, or specification error, refers to the empirical adequacy of a model. Omission of a variable which should be included, would result in specification error, as would treatment of a non-linear relationship as linear.

It is difficult to assure the elimination of error. However, as a safeguard, it is possible to examine the behavior of parameter estimates under hypothetical error conditions. Land and Felson (1978) describe a set of techniques which they refer to as sensitivity analysis. Applied to analysis of error, one of their suggestions is to hypothesize a range of error conditions, and examine the consequences of those error conditions on parameter estimates.

As described above, typical identification of two-wave one-variable models assume no errors exist in the measurement of variables, no error covariances exist and the disturbances of the equations are uncorrelated. Bohrnstedt and Carter's (1971) and Alwin and Jackson's (1980) admonitions that path analysis is based on very restrictive assumptions and that such assumptions reflect a blatant unconcern with measurement error problems are also pertinent here.

The error assumptions typically made in order to identify two-wave one-variable models are unrealistic. For example, in the Philadelphia data it is reasonable to assume there is some measurement error in the measurement of third and fourth grade reading achievement. Since a similar measuring instrument was used at both times, it may be reasonable to hypothesize correlated measurement error (see Wiley and Wiley, 1974). The Philadelphia analysis implicitly assumes the gain score was without error. Moreover, their independent variables probably contain some measurement error.

Specification error, due to the omission of independent variables, is also a major concern, especially because the Philadelphia study collected but did not report on variables over which the school had no control, e.g., race, sex, and socioeconomic background. The absence of these variables probably creates an upward bias, overstating the effect of Time 1 on Time 2 reading achievement. Variables autocorrelated over time tend to be so because the same set of independent variables tend to be operative at both times, such that exclusion of an important independent variable results in a spurious serial effect.

If tandem measurement error is present in the dependent variables, suppression on the longitudinal effect should be anticipated. On the other hand, specification error should lead to an upward bias in the longitudinal effect. Using LISREL, it is possible to deal with these separately. The matrix θ_ε, a 2 x 2 symmetric matrix, represents errors in dependent variables. $\theta\varepsilon_{11}$ represents error in y_1, $\theta\varepsilon_{22}$ error in y_2 and $\theta\varepsilon_{21}$ is the correlation between the errors, in this case perhaps a test-retest bias. To examine the effect of measurement error on β_{21} the elements $\theta\varepsilon_{11}$ and $\theta\varepsilon_{22}$ were fixed at 0%, .05 and .10 of the variance of their respective y's. Examining all combinations, nine hypothetical measurement error conditions were examined, as shown in Figure 1.

AMOUNT OF ERROR IN y_1

AMOUNT OF ERROR IN y_2		0	.05	.10
	0	0, 0	0, .05	0, .10
	.05	.05, 0	.05, .05	.05, .10
	.10	.10, 0	.10, .05	.10, .10

FIGURE 1

The matrix Ψ is also a 2 x 2 symmetric matrix of errors in equations where Ψ_{11} represents the variance of Eta $_1$, the unexplained variance in equation 1, and Ψ_{22}, the variance of Eta $_2$, the unexplained variance in equation 2. If specification error is present the residuals of the two structural equations will be correlated, perhaps reflecting the presence of omitted variables.

To investigate the potential for specification error to influence β_{21}, Ψ_{21} was fixed at 0, .05, .10 and .15. A sensitivity analysis was performed on the data (Land and Felson, 1978: 289). Each of the four levels of specification error was applied to each combination of measurement error, resulting in 36 error models.

Results of the Sensitivity Analysis

Table 6 shows the results of analyzing the thirty-six models containing combinations of error levels in Y_1, Y_2 and Ψ_{21}.

First, that the inclusion of these error estimates did not disturb the goodness of fit of the basic model, as X^2 remains constant throughout the table. Second, there are trends with changes in measurement error. Third, there is yet another set of trends to be interpreted in relation to changes in specification error, Ψ_{21}.

These are maximum likelihood estimates but these trends are identical to what would be expected from least-squares estimates, given an assumption from measurement theory that for large samples, the estimated variance of a variable measured with error will always be higher than the true variance, but that in cross product expressions, random error will tend to cancel, resulting in unbiased cross product estimates (Siegel and Hodge, 1968).

Table 6 shows three trends relating to measurement error. First, the unstandardized Beta estimates increase with removal of error from Y_1, but remain constant with removal of error from Y_2. In other words, measurement error in the antecedent time 1 variable will result in downward bias in the unstandardized beta.

TABLE 6

RESULTING PARAMETER ESTIMATES
GIVEN ERROR ASSUMPTIONS IN Y_1, Y_2, AND Ψ_{21}

MAGNITUDE OF ERROR ASSUMPTION			ESTIMATES OF:				
Y_1	Y_2	Ψ_{21}	β	b	δ_β	Ψ_{22}	χ^2
0	0	0	.592	.553	.029	.419	212.3
0	0	.05	.515	.481	.029	.423	212.3
0	0	.10	.437	.408	.031	.434	212.3
0	0	.15	.359	.335	.032	.452	212.3
0	.05	0	.593	.568	.029	.389	212.3
0	.05	.05	.516	.495	.030	.392	212.3
0	.05	.10	.440	.422	.031	.403	212.3
0	.05	.15	.364	.349	.033	.421	212.3
0	.10	0	.592	.583	.030	.355	212.3
0	.10	.05	.519	.511	.031	.358	212.3
0	.10	.10	.445	.438	.032	.369	212.3
0	.10	.15	.371	.365	.034	.387	212.3

MAGNITUDE OF ERROR ASSUMPTION			ESTIMATES OF:				
Y_1	Y_2	Ψ_{21}	β	b	δ_β	Ψ_{22}	χ^2
.05	0	0	.639	.582	.029	.403	212.3
.05	0	.05	.557	.507	.030	.406	212.3
.05	0	.10	.476	.433	.031	.418	212.3
.05	0	.15	.393	.350	.033	.436	212.3
.05	.05	0	.639	.597	.030	.371	212.3
.05	.05	.05	.559	.522	.031	.375	212.3
.05	.05	.10	.479	.448	.032	.386	212.3
.05	.05	.15	.399	.373	.034	.405	212.3
.05	.10	0	.639	.613	.031	.336	212.3
.05	.10	.05	.561	.539	.032	.340	212.3
.05	.10	.10	.483	.464	.033	.351	212.3
.05	.10	.15	.405	.389	.035	.370	212.3

MAGNITUDE OF ERROR ASSUMPTION			ESTIMATES OF:				
Y_1	Y_2	Ψ_{21}	β	b	δ_β	Ψ_{22}	χ^2
.10	0	0	.694	.615	.031	.383	212.3
.10	0	.05	.607	.538	.031	.387	212.3
.10	0	.10	.520	.461	.032	.399	212.3
.10	0	.15	.434	.384	.034	.418	212.3
.10	.05	0	.694	.630	.032	.351	212.3
.10	.05	.05	.609	.554	.032	.355	212.3
.10	.05	.10	.524	.477	.033	.366	212.3
.10	.05	.15	.440	.400	.035	.385	212.3
.10	.10	0	.694	.648	.032	.315	212.3
.10	.10	.05	.611	.571	.033	.319	212.3
.10	.10	.10	.529	.494	.034	.330	212.3
.10	.10	.15	.447	.418	.036	.349	212.3

Second, the value of the standardized beta estimate increases with removal of error from either Y_1, or Y_2. This can be explained by recalling that unstandardized and standardized estimates are related by equation (9), where b = standardized beta, β = unstandardized beta, and S = standard deviation:

$$b = \beta \frac{S_{y_1}}{S_{y_2}} \qquad (9)$$

If remains constant, and error is removed from Y , S_Y, will shift downward, requiring that b shift upward. Consequently, with S_{Y_1} and β constant, standardized estimates will be downwardly biased by random measurement error in Y_2.

Third, Table 6 also shows the standard error of Beta to increase with removal of error from either the independent or the dependent variable, despite the tendency for estimates of B to increase. Ordinarily, one would expect σ_β to decrease as β increases. A bothersome aspect of measurement error is that it produces downwardly biased estimates of the standard error of beta, even given the downward bias of β. Note, for example, that where error in Y_1 = 0, Y_2 = 0, Ψ_{21} = 0; we find β = .592 and σ_β = .029, while where error in Y_1 = .10, Y_2 = .10, Ψ_{21} = 0; we find β = .694 and σ_β = .032; the standard error increases despite an increase in β.

In this example, the bias in the standard error is not critical only because the sample size is large, 1363. In smaller samples, this bias could easily result in an inappropriate inference.

In addition to an analysis of measurement error, Table 6 contains an analysis of specification error. The matrix Ψ contains, in addition to the unexplained variances of η_1 and η_2 a cross product term, Ψ_{21}, representing the covariance between η_1 and η_2. The explained variance in η_2 is accounted for by gamma estimates in equation 2, representing effects of the independent variables, and by β, representing the effect of η_1, the grade 3 reading scores.

β represents a peculiar effect because, unlike the independent variables, y_1, a test score, is not a measure of an attribute of a causal agent, but an indicator of the student's previous accomplishment. It is arguable that β does not represent a causal influence at all, but a spurious result of the influence of unobserved variables operating ~~at~~ both Time 1 and Time 2. Consequently, the covariance between η_1 and η_2 which is accounted for by β, could just as well be allocated to Ψ_{21}.

The consequences of adding increasing proportions of covariance to Ψ_{21} is summarized in Table 6. Note that as the error in Ψ_{21} increases β decreases, and as the standard error of β increases the unexplained variance in η_2 increases, and this trend hold regardless of measurement error in y_1 and y_2.

This analytical procedure shows that problems associated with measurement error can be studied by setting bounds on parameter estimates under a range of realistic error conditions. In the analysis of measurement error, β ranged from a low of .553 to a high of .648, with a standard error from .029 to .032 under what actually are optimistic assumptions about the low magnitudes of measurement error.

Results of Testing a Model with Realistic Error Assumption

What are realistic error assumptions to use in analyzing the Philadelphia data? Reliability coefficients between alternate forms of the same test are characteristically in the range of .85 to .94 (Thorndike and Hagen, 1977: 92), suggesting that measurement error in achievement tests could realistically be set at 25%. A separate analysis was done setting measurement error at .25 on both Y_1 and Y_2, and including a 5% correlation in $\theta\varepsilon_{21}$ to allow test-retest bias. Additionally, the off diagonal psy, Ψ_{21}, was set at .05 to allow for specification errors due to omitted variables. The resulting β was .788, the σ_β was .044, and Ψ_{22} was .144.

Including higher rates of error in the longitudinal variables had the effect of removing a downward bias in the beta linking Time 1 and Time 2. An increase in beta occurred even though the setting of Ψ_{21} at .05 removed an upward bias in beta due to specification error.

CONCLUDING COMMENTS

The original report, What Works in Reading? (Kean et al, 1979) bases conclusions on questionable methodological practices. Among these are the procedure by which eighteen "significant" independent variables (out of 245) were selected, the uncritical use of gain scores, and disregard for problems of measurement error.

While it appears futile to salvage substantive conclusions, important methodological lessons can be drawn from the secondary analysis above.

The procedure by which independent variables were selected reflected a lack of theoretical guidance although theoretical guidance is essential in multivariate analysis. With a large number of variables, statistical significance is not a useful criterion. At the 95% confidence level, 12 zero-order associations can be expected to be "significant" by chance. Here we do not begin to count an astronomical number of partials, of which five percent will also be large enough to pass a significance test by chance alone. One quasi-theoretical decision guiding the analysis was the decision to drop variables outside the control of the school district, including race, socioeconomic status and sex. This ill advised step appeared motivated by an understandable desire to minimize controversy, but it was counter productive to an understanding of the data base.

The secondary analysis reported here entailed successive refinements. This is not to say that other approaches would not be equally appropriate. For example, the gain score model could also be refined by "residualizing" the gain score variable, as recommended by Bohrnstedt (1969).

The gain score model (Model 1) yields results which are virtually uninterpretable. Effects contradict long standing principals of educational practice. The longitudinal model (Model 2) results in dramatic changes in magnitude and sign of effects in contrast to Model 1. Effects in Model 2 are also more in agreement with prior expectations (see Rankin, 1980).

Subsequent refinements, including the introduction of a 10 factor measurement model on the independent side (Model 3), and the analysis of sensitivity to measurement error on the dependent side (Model 4), do not result in dramatic shifts in parameter estimates, but they do illustrate techniques which can and should be applied as new generations of software make them not only practical, but easily accessable to sociological and educational researchers everywhere.

REFERENCES

ALWIN, D. F. and D. J. JACKSON (1980) "Measurement Models for Response Errors in Surveys: Issues and Applications," pp. 68-119 in K. F. Schuessler (ed.) Soc. Methodology 1980. San Francisco: Jossey-Bass.

ALWIN, D. F. and M. J. SULLIVAN (1975) "Issues of Design Design and Analysis in Evaluation Research." Soc. Methods and Research 4 (August): 77-100.

BOHRNSTEDT, G. W. (1969) "Observations on the Measurement of Change," pp. 113-133 in E. F. Borgatta and G. W. Bohrnstedt (eds.) Soc. Methodology 1969. San Francisco: Jossey-Bass.

___ and T. M. CARTER (1971) "Robustness in Regression Analysis," pp. 118-146 in H. L. Costner (ed.) Sociological Methodology 1971. San Francisco: Jossey-Bass

BURT, R. S. (1973) "Confirmatory Factor-Analytic Structures and the Theory Construction Process." Soc. Methods and Research 2: 131-190.

CONANT, E. H. (1971) "A Cost-Effectiveness Study of Employement of Non-Professional Teaching Aides in Public Schools" Washington, D.C.: U.S. Office of Education, Bureau of Research

CRONBACH, L. J. and L. FURBY (1970) "How We Should Measure 'Change'--or Should We?" Psychological Bulletin 74: 68-80.

DUNCAN O.D. (1975 Introduction to Structural Equation Models. Academic Press, New York: p. 52

- - - (1972) "Unmeasured Variables in Linear Models" pp. 36-82 in H. L. Costner (ed.) Sociological Methodology 1972, pp. 36-82. San Francisco: Jossey-Bass.

_ _ _ (1969) "Some linear models for two-wave, two-variable panel analysis." Psychological Bulletin 77: 177-182.

_ _ _ (1966) "Path Analysis: sociological examples." Amer. J. of Sociology 72: 1-16.

FISHER, R. A. (1921) "On the mathematical foundations of theoretical statistics." Phil. Transactions A. 222: 309.

GILLESPIE, M. G. and J. FOX (1980) "Specification Error and Negatively Correlated Disturbances in "Parallel Simultaneous-Equation Models." Soc. Methods and Research 8: 273-308.

GOODMAN, L. A. (1968) "The Analysis of Cross-Classified Data: Independence, Quasi-independence, and Interactions in Contingency Tables With or Without Missing Entries" Journal American Statistics Assoc. 63 (December): 1091-1131

HANNAN, M. T. and A. A. YOUNG (1977) "Estimation in Panel Models: Results on Pooling Cross-Sections and Time Series," pp. 52-83 in D. R. Heise (ed.) Sociological Methodology 1977. San Francisco: Jossey-Bass.

HARGENS, L., B. F. RESKIN, and P. D. ALLISON (1976) "Problems in Estimating Measurement Error from Panel Data," Soc. Methods and Research 4 (May): 439-458.

HAUSER, R. M. and A. GOLDBERGER (1971) "The Treatment of Unobservable Variables in Path Analysis," pp. 81-117 in H. Costner (ed.) Sociological Methodology 1971. San Francisco: Jossey-Bass.

HEISE, D. R. (1970). "Causal Inference from Panel Data," pp. 3-27 in E. B. Borgatta (ed.) Sociological Methodology 1970. San Francisco: Jossey-Bass.

IVERSON, G. R. (1979) "Decomposing Chi-Square: A Forgotten Technique" Soc. Methods and Research 8 (November): 143-157

JORESKOG, K.G. and D. SORBOM (1981) LISREL V Analysis of Linear Structural Relationships by Maximum Likelihood and Least Squares Methods. Preliminary Edition, University of Uppsala, Department of Statistics

___ and D. SORBOM (1978). EFAP-II Exploratory Factor Analysis Program. Uppsala, Sweden: University of Uppsala.

___ and D. SORBOM (1977) "Statistical Models and Methods for Analysis of Longitudinal Data," pp 285-325 in D. V. Aigner and A. S. Goldberger (eds.) Latent Variables in Socioeconomic Models. Amsterdam: No. Holland Publishing Co.

KEAN, M. H., A. S. SUMMERS, M. J. RAIVETZ, and I. J. FARBER (1979) WHAT WORKS IN READING?: The Results of a Joint School District/Federal Reserve Bank Empirical Study in Philadelphia (May) Office of Research and Evaluation, The School District of Philadelphia.

___ (1979) WHAT WORKS IN READING?: Technical Supplement of a Joint School District/Federal Reserve Bank Empirical Study in Philadelphia. (November) Office of Research and Evaluation, The School Distict of Philadelphia.

KESSLER, R. C. (1977) "The use of change scores as criteria in longitudinal survey research." Quality and Quantity 11: 43-66.

KIM, J. and C.W. Mueller (1976) "Standardized and Unstandardized Coefficients in Causal Analysis: An Expository Note." Soc. Methods and Research, (May): 423-437.

KNOKE, David (1979) "Stratification and the Dimensions of American Political Orientations" American Journal of Political Science Vol. 23, No. 4 (November)

LAND, K.C. (1969) "Principals of Path Analysis" pp.3-37 in E.F. Borgatta (ed.) Sociological Methodology. San Francisco: Jossey-Bass.

- - - and M. FELSON (1978) "Sensitivity Analysis of Arbitrarily Identified Simultaneous-Equation Models" Soc. Methods and Research 6 (February): pp.283-307.

LONG, J.S. (1976) "Estimation and Hypothesis Testing in Linear Models Containing Measurement Error: A Review of Joreskog's Model for the Analysis of Covariance Structures" Soc. Methods and Research 5 (November): 157-206

PENDLETON, B.F., R.D. WARREN, and H.C. CHANG (1979) "Correlated Denominators in Multiple Regression and Change Analysis" Soc. Methods and Research 7: 451-474.

RANKIN, Richard (1980) "What Works in Reading" unpublished paper

SIEGEL, P. M. and R. W. HODGE (1968) " A Causal Approach to the Study of Measurement Error" in Hubert M. Blalock and Ann B. Blalock (eds) Methodology in Social Research, pp 28-59. New York: McGraw-Hill

THORNDIKE, R. L. and E. HAGEN (1977) "Measurement and Evaluation in Psychology and Education" New York: J. Wiley

___ (1966) "Intellectual Status and Intellectual Growth." Journal of Educational Psychology 57: 121-127.

___ (1963) The Concepts of Over and Under Achievement. New York: Bureau of Publications, Columbia University.

___ and E. HAGEN (1955) Measurement and Evaluation in Psychology and Education. New York: John Wiley.

TUCKER, L.R. and LEWIS, C. (1973) "A Reliability Coefficient for Maximum Likelihood Factor Analysis" Psychometrika, Vol. 38 No. 1, (March)

WHEATON, B., B. MULTHEN, D. ALWIN, and G. F. SUMMERS (1977) "Assessing Reliability and Stability in Panel Models," pp. 84-136 in D. R. Heise (ed.) Sociology Methodology 1977. San Francisco: Jossey-Bass.

WILEY, J. A. and M. G. WILEY (1974) "A Note on Correlated Errors in Repeated Measurement." Soc. Methods and Research 3: 172-188.

WON, E. (1982) "Incomplete Corrections for Regression Unreliabilities: Effects on the Coefficient Estimates" Soc. Methods and Research 10: 271-284

ZELLNER, A. (1963) "Estimates for Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results" Journal of American Statistical Assoc. 58: 977-992

BIOGRAPHIES OF AUTHORS

Leslie Hendrickson is the Evaluation Specialist of the Eugene Public Schools. His current research interests involve the application of time-series procedures to econometric and causal models.

Barnie Jones is a doctoral candidate in sociology at the University of Oregon. His major research interests are in quantitative methodology and social stratification.