DOCUMENT RESUME

ED 171 749 TM 008 838

AUTHOR	Wolfle, Lee M.
TITLE	Unmeasured Variables in Path Analysis.
INSTITUTION	Virginia Polytechnic Inst., Blacksburg.
SPONS AGENCY	National Center for Educational Statistics (DHEW/OE), Washington, D.C.
PUB DATE	Apr 79
CONTRACT	300-78-0561
NOTE	40p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Academic Achievement; Computer Programs; *Critical Path Method; Cultural Differences; Data Collection; *Mathematical Models; *Predictor Variables; Racial Differences; Research Problems; Senior High Schools; Socioeconomic Influences; *Standard Error of Measurement; *Statistical Bias; Statistical Studies
IDENTIFIERS	LISREL Computer Program

ABSTRACT

Structural equation models incorporating unmeasured variables make possible the rigorous testing of theories previously difficult to test adequately because of fallible measures of the theoretic variables. This paper first discusses a simple causal model; incorporating a single unmeasured variable for the purpose of exposition. A substantive example follows, incorporating several unmeasured variables for which multiple indicators were available. Joreskog's LISREL model for the analysis of linear structural relationships was used. The problem studied concerned the processes of educational attainment among several ethnic groups. The most important substantive conclusion was that measurement errors differ between ethnic groups and that ignoring errors leads to biased estimates of structural effects. However, application of LISREL depends on the collection of appropriate data. Unless the data have been collected by appropriate procedures, and unless the model is adequately specified, LISREL is unlikely to produce definitive results. (Author/CTM)

UNMEASURED VARIABLES IN PATH ANALYSIS*

Lee M. Wolfle

Virginia Polytechnic Institute and State University

* This paper was prepared for presentation at the annual meetings of the American Educational Research Association, San Francisco, April 8-12, 1979. Work on this project was in part supported by the National Center for Education Statistics, Department of Health, Education and Welfare (No. 300-78-0561). Susan Rothschild assisted in computations using facilities of Virginia Polytechnic Institute and State University Northern Virginia Campus. Maureen Moment provided helpful suggestions and criticisms which are gratefully acknowledged.

UNMEASURED VARIABLES IN PATH ANALYSIS*

Lee M. Wolfle

Structural equation models have been useful in estimating parameters of man substantive problems in educational research. Such models have been applied to study the effect f educational a tainment in intergenera onal o cupational mobility (Blau and Dunc . 1967), the social psycho ogical effects of one's best friend' coll e plans on the respondent's further ducation (Duncan, Haller, ortes, 1968), the effect o rents' an teachers' encouragement u educational attainment Sew and Hauser, 1975), and ethnic and al psychological effects academic achievement (Anderson and E ns, 1974; Anderson, 1978).

A the cited analyses (indeed, most studies incorporating regression structura uation rocedures ve rested upon the implicit, but more assumpt tha e inde endent ables were measured without see alon , 196 -9). In practice, measurement errors in ind variables have been ignored, because it was felt that ignoring random measurement errors merely to conclusions more conservative than would be the case. For example, it is well known that least-squares estimating procedures yield attenuated estimates of the regression slope and correlation coefficient in the bivariate case (see Appendix). Thus, it has been believed that such results understate the true relationships. In the

* This paper was prepared for presentation at the annual meetings of the American Educational Research Association, San Francisco, April 8-12, 1979. Work on this project was in part supported by the National Center for Education Statistics, Department of Health, Education and Welfare (No. 300-78-05). Susan Rothschild assisted in computations using facilities of irginia Polytechnic Institute and State University, Northern Virginia Campus. Maureen Moment provided helpful suggestions and criticisms which are grate y acknowledged.

case of multiple independent variables, however, the existence of measurement errors becomes a serious problem (Blalock, 1965; Bohrnstedt, 1969).

From a formal point of view, assuming measurement error in structural equation models is much the same as assuming variables to be unobserved. In the former case, one assumes that the true variable is observable, but only with error; in the latter, one assumes that the true variable is unobservable, and uses in its place one or more imperfectly measured indicators. Because true variables may never be measured exactly, in a strict sense all true variables are unobserved. In practice, then, observations are collected on manifest variables thought to be related to the latent variable of real theoretical interest.

Measurement errors and unobserved variables constitute a large topic. Indeed, the field of psychometrics addresses itself almost entirely to the problem of errors in variables. In sociology, substantial efforts are under way to estimate error in data collection (for example, Schuman and Presser, 197-), and in estimation procedures (for example, Blalock, Wells and Carter, 1970; Wiley, 1973). In path analysis, models containing unobserved variables have been a part of the literature for years (Hodge and Treiman, 1968; Siegel and Hodge, 1968; Duncan, 1969b; Hauser, 1969; Land, 1970; Wiley and Wiley, 1970; Hauser and Goldberger, 1971; Duncan, Featherman and Duncan, 1972; Hauser, 1972; Otto and Featherman, 1975; Bielby, Hauser and Featherman, 1977). Indeed, one of the earliest substantive applications of path analysis was by Sewell Wright (1925) to the interaction between corn crops and hog prices, and included hog breeding variables which were unobserved.

The application of structural models incorporating unmeasured variables may influence the explanation of educational phenomena.

As Kerlinger (1977) pointed out, models including unmeasured variables make possible the rigorous testing of theories previously difficult to test adequately because of fallible measures of the theoretical variables. And as Cooley (1978) noted, such models now define the "state of the art" in educational research. Unfortunately, both expository articles and reports of substantive applications of structural equation models incorporating unmeasured variables have been rare in educational research literature. This paper first discusses a simple causal model, incorporating a single unmeasured variable for the purpose of exposition. A substantive example will follow, incorporating several unmeasured variables for which multiple indicators were available. This paper thus extends the work of Wolfle (1977) and Williams (1978), who provided introductions to the subject of path analysis from the perspective of regression analysis, Wolfle (1978), who exposited path analysis as a means of substantive interpretation of data, and Anderson (1978), who exposited a nonrecursive equation model.

Let us begin with a simple example. Consider a simple causal chain of the process of intergenerational occupational mobility from father's socioeconomic status (X_3), to respondent's educational attainment (X_2), to respondent's socioeconomic status (X_1). However, let us revise the model such that true educational attainment is not directly observed. Instead, its observed indicator, educational attainment, is contaminated with errors of measurement. We assume that the amount of education actually recorded is caused by the respondent's true educational attainment, in addition to several other factors. For example, the respondent may be more or less ignorant of the number of years of regular school or college he or she completed and got credit for." The respondent may tend to round off educational attainment to even years, or multiples of four (such as 12, 16, or 20). Some respondents may wish to appear to have acquire more (or less) schooling than was actually the case.

The complete model consists of three equations, the first of which describes the fallible measurement of observed education while the other two represent the causal model as such. The three equations may be written:

$$x_2 = b_{2\eta}\eta + e,$$

$$\eta = b_{\eta 3}x_3 + u,$$

$$x_1 = b_{1\eta}\eta + v,$$

where x_1 is the respondent's Duncan (1961) socioeconomic index score as revised by Hauser and Featherman (1977), x_2 is the respondent's recorded educational attainment, x_3 is respondent's father's Duncan socioeconomic index score, η is true educational attainment, and e, u, and v are residual errors. All of these are measured as deviations from their means. The b's are, therefore, regression coefficients, and $b_{2\eta}$ = 1.00. The structural relationships may be diagrammed, as shown in Figure 1.

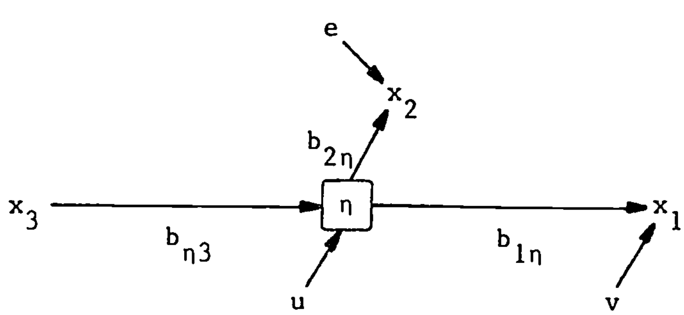

Figure 1. A Causal Chain of Intergenerational Status Attainment

A remarkable property of this model is that unbiased estimates of the coefficients may be obtained for both the causal parameters and the measurement process. This occurs because the model is overidentified, and because of some restrictive assumptions imposed on the expected associations of the residual errors with other variables in the model. The usual specification in such models is that the residual errors are uncorrelated in the population with other, predetermined variables in the equation. Thus

$$E(x_3u) = E(\eta v) = E(x_3v) = 0.$$

In addition, it is assumed that the residual error of measurement is uncorrelated with the true score, η, and also with both x_1 and x_3:

$$E(\eta e) = E(x_1e) = E(x_3e) = 0.$$

These strong assumptions are roughly equivalent to assuming the error of measurement is random, and not systematic. These oversimplifying assumptions are properties of the model, not necessarily of what the world is really like. In any realistic context, these assumptions are problematic, and must be assessed against the researcher's knowledge of the topic under investigation. For example, it is possible that respondents whose fathers are employed in occupations of low socioeconomic status, or who themselves are employed in such occupations, tend to overstate their educational attainment. Complex models can be constructed which permit the intercorrelation of residuals, but the simple alternatives for this simple example are either to abandon the exercise or to accept the restrictive assumptions.

The three equations in the model may be reduced to three normal equations with three unknowns. The details of these computations are shown in the Appendix. For purposes of illustration, the model was estimated with data taken from the 1977 general social survey of the National Opinion Research Center (1977), for whites and blacks. The correlations, means and standard deviations are shown in Table 1, and the results in Table 2.

The upper panel of Table 2 shows the results one would obtain with ordinary least squares (OLS). The structural coefficients (regression coefficients in their original metric) are shown, and below each in parenthesis are the standardized (path) coefficients. When comparing socioeconomic returns across groups, the structural coefficients should be used (see Kim and Mueller, 1976). The coefficients of determination are shown the right-hand column. For whites, one would conclude that one point of father's socioeconomic index (SEI) returns about .05 years of education, and that one year of education was converted into about 4.5 points of respondent's own SEI. For blacks also, one would conclude that one point of father's SEI yielded about .05 years of education, but that blacks were able to convert one additional year of education into only 3.4 points of their own SEI.

The lower panel of Table 2 shows the results that one would obtain from the model diagramed in Figure 1. Note that measurement errors seem to be larger for blacks than for whites; that is, the standardized coefficient relating true education to observed education is larger for whites than for blacks. Comparing the OLS estimates to the corrected estimates for the regression of education of father's SEI, one should note that the OLS estimate is identical to the corrected estimates. Random measurement error in the dependent variable does not bias the OLS estimate. However, random error of measurement in the independent variable imparts a downward bias to the OLS estimate. And the lower the precision of measurement,

Table 1. Correlations, Means, and Standard Deviations of Status Variables; Whites Above Diagonal (N = 1[illegible]3), Blacks Below Diagonal (N = 172).

	Father's SEI X_3	Education X_2	SEI X_1	White Mean	White S.D.
Father's SEI (X_3)	1.00	.370	.263	28.05	23.57
Education (X_2)	.255	1.00	.570	11.83	3.14
SEI (X_1)	.279	.560	1.00	37.48	24.53
Black Mean	14.48	10.43	23.39		
Black S.D.	17.90	3.66	21.85		

Table 2. Ordinary Least Square and Corrected Estimates of Parameters in a Causal Chain Model of Intergenerational Status Attainment

Dependent Variable	Predetermined Variables: Father's SEI (X_3)	Observed Education (X_2)	True Education (η)	R^2
	Ordinary Least Square Estimates			
	Whites			
Observed Education (X_2)	.0493 (.370)	. . .	. . .	.137
SEI (X_1)	. . .	4.451 (.570)	. . .	.325
	Blacks			
Observed Education (X_2)	.0521 (.255)	. . .	. . .	.065
SEI (X_1)	. . .	3.346 (.560)	. . .	.314
	Corrected Estimates			
	Whites			
True Education (η)	.0493 (.413)	. . .	. . .	.171
Observed Education (X_2)	. . .	. . .	1.000 (.895)	.801
SEI (X_1)	. . .	. . .	5.549 (.637)	.406
	Blacks			
True Education (η)	.0521 (.356)	. . .	. . .	.127
Observed Education (X_2)	. . .	. . .	1.000 (.715)	.511
SEI (X_1)	. . .	. . .	6.546 (.783)	.613

Note: Standardized (path) coefficients appear in parentheses.

the greater the downward bias. For whites, the corrected regression coefficient leads one to conclude that one year of true educational attainment was converted into 5.5 points of SEI. When measurement errors were ignored for whites, SEI returns to education were underestimated by about 20 percent, assuming the measurement errors were purely random. For blacks, the corrected regression coefficient leads one to conclude that one year of true educational attainment was converted into 6.5 points of SEI. When measurement errors were ignored for blacks, SEI returns to education were underestimated by about 49 percent, again assuming the measurement errors were well behaved.

This analysis was based on the assumption that the only kinds of measurement error for both blacks and whites were random, and not systematic. Our conclusions are, therefore, contingent on the correctness of those assumptions, but suggest that previous studies (for example, Duncan, 1969a) comparing the processes of status attainment for whites and blacks have exaggerated racial differences in returns to education by failing to account for measurement error. Bielby, Hauser and Featherman (1977) estimated status attainment models for whites and blacks incorporating both structural and response components. They found that response errors for whites were random, but were not for blacks. Nonetheless, the substantive consequences were the same as here: ignoring measurement errors exaggerated racial differences.

STRUCTURAL EQUATION MODELS WITH MEASUREMENT ERRORS

Unmeasured variables may also be included in causal models in the more general case in which unmeasured variables appear as underlying causes of several observed variables. This type of model translates into a confirmatory factor-analysis model with an assumed structural order among the factors. A general method has only recently been developed which incorporates recursive path analysis, interdependent econometric models, factor analysis, and analysis of covariance structures. This method, the analysis of linear structural relationships, or LISREL, was introduced by Jöreskog's (1973) technical paper, and has been updated in Jöreskog (1977). Less technical introductions are available in Long (1976) and Jöreskog and Sörbom (1978). A computer program is available (Jöreskog and Sörbom, 1978).

The LISREL model assumes a causal structure among a set of unmeasured, latent variables, some designated as exogenous and others as endogenous. These unmeasured variables are also related to a set of observed variables such that (in the example to follow) the latent variables appear as causes of the observed variables. The LISREL model, therefore, consists of two parts: the measurement model, and the structural equation model (Jöreskog and Sörbom, 1978: 3-4).

By way of example, Lichtman and Wolfle (1978) are studying the processes of educational attainment among several ethnic groups, including whites, blacks, and Hispanics. They proposed to compare structural equation models among ethnic groups in order to determine the extent to which differences exist in the educational returns to socioeconomic background and within-school variables. The population under study is the high school graduating class of 1972, described in detail by Levinsohn, et al. (1978). One should expect that members of this high school cohort have not all completed their education as of 1976 (the latest followup), so the results pertain to educational returns as attained four years after graduation. The respondents were initially surveyed in 1972, and followed up in 1973, 1974, and 1976. Because some questions were repeated in various questionnaires, or because alternative means exist for constructing manifest variables, in many cases multiple indicators exist for latent variables. This becomes important in light of Bohrnstedt and Carter's (1971: 142) admonition that, "sociologists seem to be blatantly unconcerned with the problems of measurement error." Moreover, Bielby, Hauser, and Featherman (1977) showed that differential measurement errors existed between blacks and whites in the 1973 replication of Blau and Duncan (1967), thus leading to exaggerated racial differences in models ignoring measurement error.

For the expository purposes of this paper, a preliminary model incorporating structural associations among several latent variables, and components of measurement errors, has been constructed. This model, shown in Figure 2, includes two manifest measures for father's educational attainment, mother's educational attainment, and the respondent's high school curriculum. Three manifest measures are included for respondent's educational attainment. Single manifest variables measure father's socioeconomic index, the number of siblings, and high school class ranking. The LISREL model specifies that each manifest variable is generated by a latent factor for that variable, plus a response error which is independent of the latent factor. In LISREL terminology:

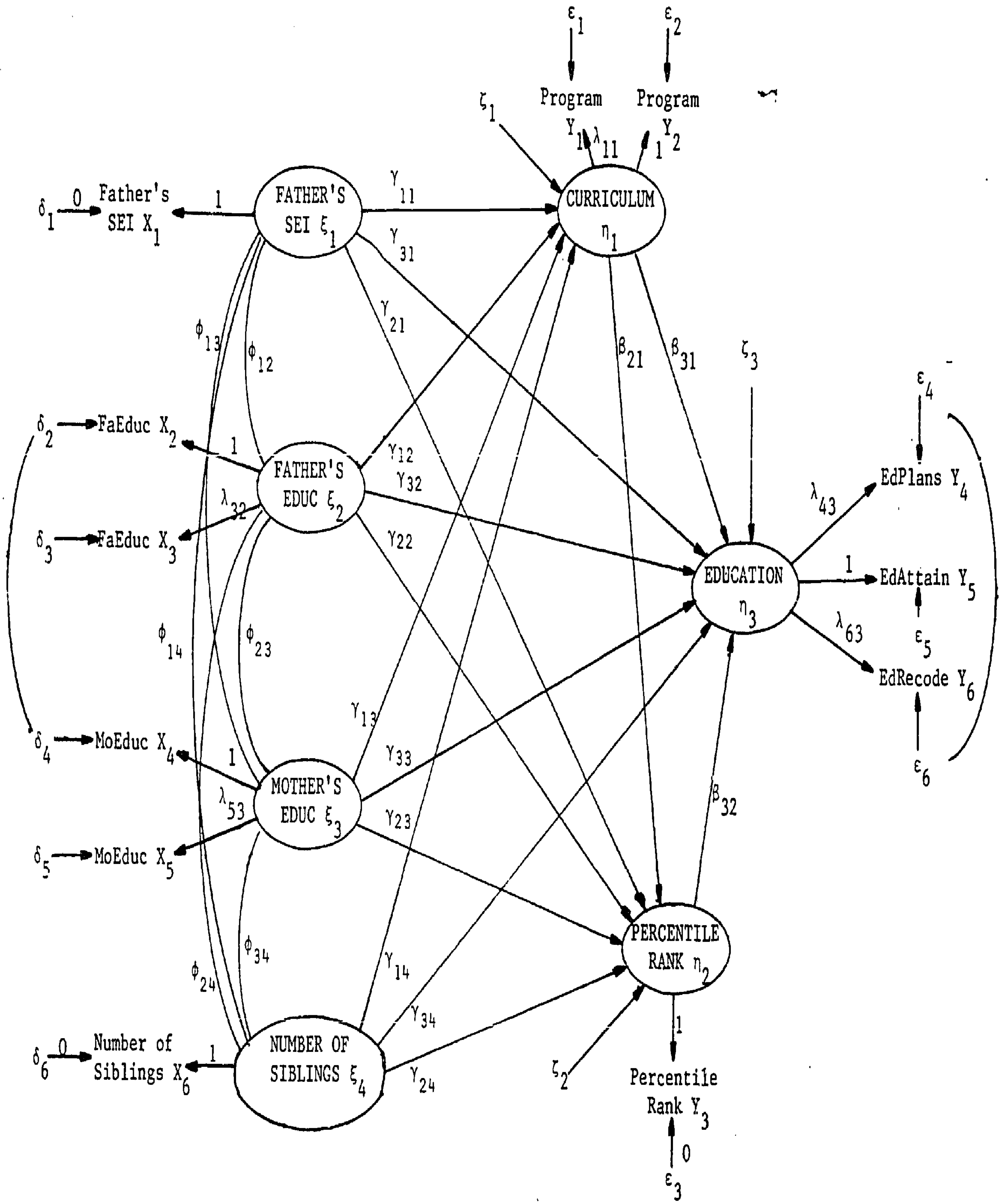

Figure 2. Structural Equation and Measurement Models of Educational Attainment, 1972 High School Graduates

two random vectors $\eta' = (\eta_1, \eta_2, \eta_3)$, and $\xi' = (\xi_1, \xi_2, \xi_3, \xi_4)$ represent the latent endogenous and latent exogenous variables, respectively. The model specifies a fully recursive causal structure among the latent variables, such that:

$$\beta \eta = \Gamma \xi + \zeta$$

where $\beta(3 \times 3)$ and $\Gamma(3 \times 4)$ are matrices of structural coefficients in which Γ is a full matrix relating the exogenous vector to each of the endogenous latent variables, and β is a matrix relating the endogenous variables to each other. $\zeta' = (\zeta_1, \zeta_2, \zeta_3)$ is a random vector of residuals uncorrelated with ξ.

The vectors η and ξ are not observed, but $y' = (y_1,\ldots,y_6)$ and $x' = (x_1,\ldots,x_6)$ are observed, such that:

$$y = \Lambda_y \eta + \varepsilon$$

and

$$x = \Lambda_x \xi + \delta$$

where ε and δ are vectors of errors of measurement in y and x, respectively. These errors of measurement represent both specific and random components of variation (see Alwin and Jackson, forthcoming). They are assumed to be uncorrelated with η, ξ, and ζ, but may be correlated among themselves. The matrices $\Lambda_y(6 \times 3)$, and $\Lambda_x(6 \times 4)$ are regression matrices of y on η and of x on ξ, respectively.

Let $\Phi(4 \times 4)$ be the covariance matrix of ξ. Let $\Psi(3 \times 3) = \text{diag}(\psi_{11}, \psi_{22}, \psi_{33})$ be the variance matrix of ζ. Let Θ_ε and Θ_δ be the covariance matrices of ε and δ, respectively. In application, some of the elements of the four regression matrices, and the four covariance matrices, are fixed and equal to assigned values. Other elements are free parameters to be estimated by the method of maximum likelihood. This defines the LISREL model.

The structural model is presented in the path diagram of Figure 2. The variables enclosed in ellipses are unobserved, latent variables. The manifest variables included in the model are as follows, in which the number in parentheses refers to the variable number as given in Levinsohn, et al. (1978):

X_1 = father's socioeconomic index (V2468),

X_2 = composite measure of father's education (V1627),

X_3 = father's education (V1009),

X_4 = composite measure of mother's education (V1628),

X_5 = mother's education (V1010),

X_6 = sum of older and younger brothers and sisters (V1460 + V1461 + V1462 + V1463),

Y_1 = high school program as reported by respondent (= 1 if academic, = 0 otherwise) (V209),

Y_2 = high school program as reported by school record (= 1 if academic, = 0 otherwise) (V196),

Y_3 = percentile rank in class (V631),

Y_4 = educational plans as of 10/1/76 (V1855),

Y_5 = educational attainment as of 10/1/76 (V1854), and

Y_6 = educational recode (Melone, personal correspondence).

The substantive portion of Figure 2 is a fully recursive model among the latent variables, represented by the following structural equations:

$$\eta_1 = \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \gamma_{13}\xi_3 + \gamma_{14}\xi_4 + \zeta_1$$

$$\eta_2 = \gamma_{21}\xi_1 + \gamma_{22}\xi_2 + \gamma_{23}\xi_3 + \gamma_{24}\xi_4 + \beta_{21}\eta_1 + \zeta_2$$

$$\eta_3 = \gamma_{31}\xi_1 + \gamma_{32}\xi_2 + \gamma_{33}\xi_3 + \gamma_{34}\xi_4 + \beta_{31}\eta_1 + \beta_{32}\eta_2 + \zeta_3.$$

In algebraic form, the measurement portion of Figure 2 is:

$$x_1 = \xi_1$$

$$x_2 = \lambda_{22}\xi_2 + \delta_2$$

$$x_3 = \lambda_{32}\xi_2 + \delta_3$$

$$x_4 = \lambda_{43}\xi_3 + \delta_4$$

$$x_5 = \lambda_{53}\xi_3 + \delta_5$$

$$x_6 = \xi_4$$

$$y_1 = \lambda_{11}\eta_1 + \varepsilon_1$$

$$y_2 = \lambda_{21}\eta_1 + \varepsilon_2$$

$$y_3 = \eta_2$$

$$y_4 = \lambda_{43}\eta_3 + \varepsilon_4$$

$$y_5 = \lambda_{53}\eta_3 + \varepsilon_5$$

$$y_6 = \lambda_{63}\eta_3 + \varepsilon_6$$

A metric for the latent variables is established by fixing $\lambda_{22} = \lambda_{43} = \lambda_{21} = \lambda_{53} = 1.0$. That is, the metric of the latent variables father's education, mother's education, curriculum, and respondent's education are fixed to be the same as that of the composite measures of education for father and mother, respectively, the school report of curriculum, and educational attainment as of 10/1/76. The metrics of father's socioeconomic status, siblings, and class rank have already been fixed by the algebra of the measurement model. Normalizations of this kind are necessary because the metric of an unobserved variable is arbitrary. Consequently, the regression slopes of manifest variables on latent variables are identifiable only relative to each other.

The model was estimated with data for white male 1972 high school graduates (N = 2955) with the specification that the response errors were uncorrelated. The resulting $\chi^2 = 150.75$, with df = 38, indicated that the model did not do a very good job of reproducing the observed variance-covariance matrix. Examination of the first-order derivatives indicated the possibility that the specification of uncorrelated response errors may have been untenable. Specifically, the response errors of x_2 and x_4 may be correlated. These variables are the composite measures of father's and mother's education, and apparently systematic errors of construction exist in both variables. Re-estimating the model allowing for correlated response errors between x_2 and x_4 resulted in a $\chi^2 = 95.00$, with df = 37. Because the difference in these chi-squares is itself distributed according to chi-square with one degree of freedom, it is obvious that the correlated response error was statistically significant. Yet once again the model does not do a very good job of reproducing the variance-covariance matrix. Re-examination of the first-order derivatives suggested that y_4 and y_6 had correlated response errors. Re-estimation yielded a $\chi^2 = 73.85$, with df = 36, which became the final model because the addition of the next most likely correlated response error did not significantly reduce the value of χ^2 (see Sörbom, 1975).

Identical models were also estimated for black males (N = 257) with $\chi^2 = 43.45$, with df = 36, and for Hispanic males (N = 125) with $\chi^2 = 51.59$, with df = 36. Estimates for the measurement model are shown in Table 3 for white males, Table 4 for black males, and Table 5 for Hispanic males. Shown in column 3 of these tables are the standard deviations of manifest variables; column 4 contains the standard deviations of response errors not accounted for by the underlying latent variables; column 5 shows the standard deviations of the latent variables; column 6 contains the relative slopes of the manifest variables regressed on the latent variables; and column 7 shows estimates of the reliability coefficients.

Among white males in the NLS sample, different reports of the same underlying variable were likely to have different slope coefficients. For some variables, such as curriculum track, these different slopes indicate different fits between the manifest and latent variables. For example, among whites the school record measure of curriculum track was a more reliable indicator of the true variable than was the student's own report. For other variables, different slopes reflect different scales of the manifest variables. For example, the composite measures of parental education were scaled from 1 (less than high school) to 5 (MA, or PhD), while the first followup questions of parental education were scaled from 1 (none, or grade school only) to 9 (PhD or equivalent). However, the reliability coefficients for these variables indicate that the composite measures of parental education, which were based on responses to baseyear, first followup, and activity state questionnaires, were more reliable measures of the underlying latent variables than were the first followup questions alone. However, caution should be exercised in generalizing from these preliminary results. The fact that the two measures of parental education differ in their scales of measurement may

Table 3. Measurement Model Parameter Estimates for White Male 1972 High School Graduates (N = 295

Variable True (τ_j)	Observed	Observed SD σ_i	SD of Error σ_{e_i}	SD of True Score σ_{τ_j}	Relative Slope λ_{ij}	Reliability Coefficient $(\sigma^2_{\tau_j}/\sigma^2_i)\lambda_{ij}$
FAEDUC ξ_2	X_2(V1627)	1.28	.44 (.11)	1.20	1.00	.88
	X_3(V1009)	2.22	.49 (.19)		1.81 (.02)	.53
MOEDUC ξ_3	X_4(V1628)	1.01	.46 (.11)	.90	1.00	.79
	X_5(V1010)	1.73	.34 (.20)		1.87 (.03)	.51
CURRICULUM η_1	Y_1(V209)	.50	.33 (.06)	.41	.91 (.02)	.61
	Y_2(V196)	.50	.28 (.06)		1.00	.67
EDUCATION η_3	Y_4(V1855)	2.30	1.20 (.25)		1.10 (.02)	.67
	Y_5(V1854)	1.97	.82 (.20)	1.79	1.00	.83
	Y_6(recode)	.70	.29 (.07)		.36 (.01)	* * *

Note. -- Standard errors of parameter estimates appear in parentheses.

Table 4. Measurement Model Parameter Estimates for Black Male 1972 High School Graduates (N = 25

Variable: True (τ_j)	Variable: Observed	Observed SD σ_i	SD of Error σ_{e_i}	SD of True Score σ_{τ_j}	Relative Slope λ_{ij}	Reliability Coefficient $(\sigma^2_{\tau_j}/\sigma^2_i)\lambda_{ij}$
FAEDUC ξ_2	X_2(V1627)	.94	.48 (.19)	.82	1.00	.76
	X_3(V1009)	1.67	.13 (.35)		2.04 (.12)	.49
MOEDUC ξ_3	X_4(V1628)	.97	.67 (.23)	.71	1.00	.54
	X_5(V1010)	1.63	.37 (.45)		2.34 (.22)	.44
CURRICULUM η_1	Y_1(V209)	.47	.36 (.12)	.38	.79 (.10)	.52
	Y_2(V196)	.46	.26 (.13)		1.00	.68
EDUCATION η_3	Y_4(V1855)	2.16	1.16 (.51)	1.58	1.15 (.09)	.62
	Y_5(V1854)	1.88	1.01 (.41)		1.00	.71
	Y_6(recode)	.64	.19 (.15)		.38 (.03)	* * *

Note. -- Standard errors of parameter estimates appear in parentheses.

Table 5. Measurement Model Parameter Estimates for Hispanic Male 1972 High School Graduates (N =

Variable: True (τ_j)	Variable: Observed	Observed SD σ_i	SD of Error σ_{e_i}	SD of True Score σ_{τ_j}	Relative Slope λ_{ij}	Reliabilit Coefficien $(\sigma^2_{\tau_j} / \sigma^2_i)\lambda_i$
FAEDUC ξ_2	X_2(V1627)	1.08	.53 (.25)	.93	1.00	.74
	X_3(V1009)	2.05	.87 (.47)		1.99 (.17)	.41
MOEDUC ξ_3	X_4(V1628)	.82	.34 (.13)	.75	1.00	.84
	X_5(V1010)	1.51	.33 (.32)		1.98 (.13)	.49
CURRICULUM η_1	Y_1(V209)	.41	.33 (.13)	.32	.76 (.15)	.46
	Y_2(V196)	.46	.33 (.15)		1.00	.48
EDUCATION η_3	Y_4(V1855)	2.19	1.23 (.65)		1.33 (.18)	.51
	Y_5(V1854)	1.77	1.12 (.49)	1.36	1.00	.59
	Y_6(recode)	.54	.32 (.16)		.32 (.04)	* * *

Note. -- Standard errors of parameter estimates appear in parentheses.

indicate that the two observed variables are not manifestations of the same true score.

Caution is particularly recommended in interpreting the results of the measurement of respondent's education. The preliminary model estimated here included two measures of education measured on the same scale: educational plans as of 10/1/76 and educational attainment as of 10/1/76. A third composite measure was included, one suggested by the staff of NCES in which 1 = no higher education, 2 = some college, and 3 = BA and higher. The relative slope of this composite measure varies so much from the other two manifest variable regression slopes, and its error variance is so different, that it should probably not be viewed as a manifest component of the same latent factor that underlies the other two education variables. One result of the mismatch is a meaningless reliability estimate greater than unity.

One further caution of interpretation is worth noting. Classical true score models express an observed score in terms of two orthogonal components composed of a true score and an error score. As a result, errors based on true score models are uncorrelated with true scores and other error scores. However, the vectors of residual factors, $\underset{\sim}{\delta}$ and $\underset{\sim}{\varepsilon}$, contain both measurement error and reliable variation specific to each manifest variable (Alwin and Jackson, forthcoming). As a result, it is possible for some or all of the residual errors to be correlated even in the population, much as we have seen that the errors of x_2 and x_4, the two composite measures of parental education, were correlated. Apparently, whatever errors of measurement entered into the construction of one parent's education composite score also entered into the other parent's.

Comparison of the measurement-model results for white males to those of blacks and Hispanics indicates that within each population the most reliable measure of parental education was the constructed composite variable. The most reliable measure of curriculum membership was the school record. The most reliable measure of true education was the respondent's report of his educational attainment as of 10/1/76. Across populations, the reliability coefficients for blacks and Hispanics were lower than those for whites. Both blacks and Hispanics exhibited less variation in the observed measures than did whites. Blacks and Hispanics also exhibited less variation in the latent factor scores; proportionately, there were even greater disparities among the latent variances than among the observed. As a result, the reliability coefficients for blacks and Hispanics were substantially lower than those of whites.

Clearly these findings suggest caution in interpreting models of status attainment among minority groups that do not take account of response error, especially when comparing structural coefficients across groups. Table 6 presents ordinary least square (in parentheses) and corrected LISREL estimates for the structural equation portion of the model represented by Figure 2. Comparison of these estimates provides some indication of the biases encountered when measurement errors are ignored. (Another example has been offered by Bielby, Hauser, and Featherman, 1977.)

First, the ordinary least squares regression of educational attainment on four family background variables and two intervening measures of high school effects accounts for two-fifths of the variance in educational attainment for white males, but only one-fourth of the

Table 6. Corrected (LISREL) and OLS Estimates of Parameters of the Educational Attainment Process for 1972 High School Graduates

Dependent Variable	Predetermined Variables*						R^2
	FaSEI	FaEduc	MoEduc	NumSibs	Curriculum	H.S. Rank	
Whites							
Curriculum	.002	.084	.052	-.025	. . .	. . .	.165
	(.002)	(.068)	(.049)	(-.026)	. . .	. . .	(.109)
H.S. Rank	-.032	.587	-.362	-.160	37.787	. . .	.325
	(.004)	(1.303)	(-.038)	(-.482)	(25.041)	. . .	(.232)
Education	.002	.194	.095	-.063	2.271	.012	.558
	(.003)	(.237)	(.092)	(-.071)	(1.326)	(.020)	(.400)
Blacks							
Curriculum	.003	.063	-.008	-.021	. . .	. . .	.086
	(.002)	(.056)	(.004)	(-.018)	. . .	. . .	(.050)
H.S. Rank	-.074	-2.814	2.952	.723	34.045	. . .	.237
	(-.032)	(-2.306)	(3.115)	(.552)	(24.198)	. . .	(.193)
Education	.012	.010	.344	-.021	1.825	.009	.411
	(.015)	(.013)	(.162)	(-.043)	(1.068)	(.016)	(.249)

Table 6. (continued)

Dependent Variable	Predetermined Variables*						
	FaSEI	FaEduc	MoEduc	NumSibs	Curriculum	H.S. Rank	R^2
Hispanics							
Curriculum	.004	-.000	-.021	.002	. . .	. . .	.062
	(.003)	(-.018)	(.041)	(.006)	. . .	. . .	(.025)
H.S. Rank	.014	-.212	2.996	1.301	50.983	. . .	.407
	(.147)	(.341)	(.172)	(1.227)	(26.166)	. . .	(.255)
Education	-.012	-.053	.168	-.072	2.770	.010	.582
	(-.006)	(-.096)	(.018)	(-.110)	(.558)	(.028)	(.249)

Note. -- The ordinary least squares estimates appear in parentheses below the LISREL estimates. The variables used in the OLS regressions were Father's SEI (X_1: V2468), Father's Education Composite (X_2: V1627), Mother's Education Composite (X_4: V1628), Number of Siblings (X_6: V1460 - V1463), Curriculum {= 1 if Academic; = 0 otherwise} (Y_2: V196), High School Percentile Rank (Y_3: V631), and Educational Attainment as of 10/1/76 (Y_5: V1854).

variance for both blacks and Hispanics. However, these results confound measurement error with true variation and result in coefficients of determination that understate the ratio of explained to total true variance by nearly 30 percent for whites, but by nearly 40 percent for blacks and 60 percent for Hispanics. By taking measurement errors into account, over half the variance in true educational attainment can be explained for Hispanics (R^2 = .582) and for whites (R^2 = .558), but not for blacks (R^2 = .411).

Comparison of the OLS to corrected estimates in the regression of educational attainment again indicates the biases due to ignoring measurement errors. In particular, notice that ignoring measurement errors does not necessarily produce attenuated estimates. Indeed, some of the OLS estimates are substantially larger than the corrected estimates. For all three groups, corrections for measurement error produce nearly identical effects, at least in the direction of the bias. There appear to be downward biases (the OLS estimates underestimate the corrected estimates) for mother's education and curriculum track. There appear to be upward biases for father's education, the number of siblings, and high school class ranking. There is also an upward bias for father's socioeconomic index among whites and blacks, but for Hispanics the OLS estimate for father's SEI understates the corrected estimate.

Although the direction of bias is nearly uniform across ethnic groups, the magnitude is not. One example is provided by the effect of membership in an academic track on educational attainment. Ordinary least squares regressions would indicate that membership in an academic track results in about one and one-third additional years of education for whites (measured four years after high school graduation), and one additional year for blacks, but only one-half year for Hispanics. However, when the confounding effects of measurement errors are removed, Hispanics are estimated to convert membership in a high school academic track into 2.8 additional years of education, while whites have a comparable estimate of 2.3 years, and blacks 1.8 years.

Another example is provided by the effect of high school rank on educational attainment. Ranking the magnitude of the OLS estimates would lead one to conclude that Hispanics were best able to convert increased high school class ranking into educational attainment, followed by whites, then blacks. However, when corrected for measurement errors, all three groups were apparently equally able to convert class ranking into educational attainment.

Overall, the consequences of ignoring measurement error appear to be greater in the case of Hispanics than of either whites or blacks, and greater for blacks than for whites. Since the biases in structural estimates ignoring measurement error are larger among Hispanics and blacks than among whites, uncorrected ethnic comparisons show unrealistically large differences between ethnic groups in the effects of familial background and high school process effects.

CONCLUSIONS

Educational researchers have long known that ignoring measurement errors will lead to biased estimates of structural effects. However, until recently multivariate analytic procedures which correct for measurement errors were not generally available. Recent developments by Jöreskog and Sörbom (1978) have made available a general computer program that permits estimation of structural effects corrected for measurement errors. The application of these techniques to a substantive problem in education has indicated the advantages of the LISREL approach, along with several cautionary reservations.

The most important substantive conclusion inherent in this analysis supports the findings of Bielby, Hauser and Featherman (1977): measurement errors differ between blacks and whites; ignoring them leads to biased estimates of structural effects. Moreover, the present analysis shows that Hispanics also report data with inherent measurement errors, and ignoring them will lead to estimates even more biased than among either whites or blacks.

Another set of substantive conclusions could be drawn from the estimated parameters of the Hispanic model of educational attainment. To the best of my knowledge, these are the first estimates, unbiased by measurement error, of the process of status attainment for any ethnic group in America other than whites or blacks. However, I have refrained from discussing Hispanics because the model explicated in this paper was a preliminary construction, and is already outmoded. In particular, the model omits measures of ability, and as Scarr and Weinberg (1978) demonstrated, the omission of ability leads to spurious estimates of causal effects.

Finally, a cautionary note is in order. Kerlinger (1977) correctly pointed out that the LISREL approach toward multivariate analysis contains a great deal of promise for testing theories that have been difficult to test adequately with previously available analytic procedures.

However, application of LISREL (indeed, any analytic procedure) depends upon the collection of appropriate data. Specifically, measures of different variables must be ascertained on different occasions, or by different means, data collection procedures that can be considerably more expensive than the usual survey.

My own view of the utility of LISREL is more skeptical than Kerlinger's (1977). Unless the data to be analyzed have been collected by appropriate procedures, and unless the model is adequately specified, LISREL is unlikely to produce the definitive tests Kerlinger suggests are possible. The past decade has seen recursive path analytic procedures faddishly applied to implausibly constructed models. Except for the inherent difficulties in specifying the model of the computer program, the next decade may see implausible examples of substantive analyses based on LISREL. In the past six months I have twice had manuscripts returned to me with reviewer's naive suggestions that the problems they recognized could be solved by reanalyzing the data with LISREL. They could not. As Cooley (1978: 13) so insightfully pointed out last year, more important than number crunching is the careful measurement of a few "right" variables, variables that permit statistical controls for major alternative explanations. Data analysis may stimulate thinking, but it is not a substitute for it.

APPENDIX

The three equations which define the structural model are:

$$x_2 = b_{2\eta}\eta + e \qquad (1),$$

$$\eta = b_{\eta 3}x_3 + u \qquad (2),$$

$$x_1 = b_{1\eta}\eta + v \qquad (3).$$

The notation may be revised such that x_1, x_2, x_3, and η refer to the standardized values of these variables. Equations (1) through (3) may be rewritten using the using the usual equalities:

$$p_{yx} = b_{yx}(\sigma_x/\sigma_y) \qquad (4),$$

$$p_{yu} = \sigma_u/\sigma_y \qquad (5).$$

These coefficients were termed path coefficients by Sewell Wright (1921). Rewriting equations (1) through (3) in terms of path coefficients and standardized variables yields:

$$x_2 = p_{2\eta}\eta + p_{2e}e \qquad (6),$$

$$\eta = p_{\eta 3}x_3 + p_{\eta u}u \qquad (7),$$

$$x_1 = p_{1\eta}\eta + p_{1v}v \qquad (8),$$

with the specifications

$$E(x_3u) = E(\eta v) = E(x_3v) = 0 \qquad (9),$$

$$E(\eta e) = E(x_1e) = E(x_3e) = 0 \qquad (10).$$

To solve the path coefficients in equations (6) through (8), we will multiply through these equations by one or another of the variables, and be taking expectations. Because the covariance of two standardized variables is the coefficient of correlation, taking expectations of a covariance will yield the population correlation coefficient, ρ.

First, multiply equation (6) through by η, and we have:

$$\rho_{2\eta} = p_{2\eta} \qquad (11),$$

since $E(\eta\eta) = 1$, and $E(\eta e) = 0$ by assumption. In similar fashion, multiplying equation (7) through by x_3, and equation (8) through by η, yields:

$$\rho_{\eta 3} = p_{\eta 3} \qquad (12),$$

$$\rho_{1\eta} = p_{1\eta} \qquad (13).$$

Multiplying equation (6) through by x_3, and taking expectations, yields:

$$\rho_{23} = p_{2\eta}p_{\eta 3} \qquad (14),$$

because $E(x_3 e) = 0$, by assumption, and $p_{\eta 3} = \rho_{\eta 3}$ by equation (12). Multiplying equation (8) through by x_3, and taking expectations, yields:

$$\rho_{13} = p_{1\eta}\rho_{\eta 3} \qquad (15),$$

and multiplying equation (8) through by x_2 yields:

$$\rho_{12} = p_{1\eta}p_{2\eta} \qquad (16),$$

because $E(x_2 v) = 0$, an equality implied by equations (9) and (10).

Equations (14), (15), and (16) form three equations in three unknowns, and:

$$p_{1\eta} = \sqrt{\rho_{12}\rho_{13}/\rho_{23}} \qquad (17),$$

$$p_{2\eta} = \sqrt{\rho_{12}\rho_{23}/\rho_{13}} \qquad (18),$$

$$p_{\eta 3} = \sqrt{\rho_{13}\rho_{23}/\rho_{12}} \qquad (19).$$

Using the sample correlation coefficients given in Table 1 to estimate the population coefficients in equations (17) through (19) give the standardized results presented in Table 2.

The corrected regression coefficient $b_{\eta 3}$ in equation (2), as implied by equation (4), is given by:

$$b_{\eta 3} = p_{\eta 3}(\sigma_\eta/\sigma_3) \qquad (20),$$

in which

$$\sigma_\eta = p_{2\eta}\sigma_2 \qquad (21),$$

the usual association between the true and observed standard deviations (for example, Gulliksen, 1950: 23). Note therefore that:

$$b_{\eta 3} = p_{\eta 3}(p_{2\eta}\sigma_2/\sigma_3) \qquad (22),$$

but $p_{\eta 3}p_{2\eta} = \rho_{23} = p_{23}$, so that

$$b_{\eta 3} = p_{23}(\sigma_2/\sigma_3) \qquad (23).$$

That is, the corrected regression coefficient of true education regressed on father's SEI is equal to the OLS regression coefficient one would obtain from regressing observed education on father's SEI.

The corrected regression coefficient $b_{2\eta}$ in equation (1) is given by:

$$b_{2\eta} = p_{2\eta}(\sigma_2/\sigma_\eta) \qquad (24),$$

but substitution by equation (21) reveals that

$$b_{2\eta} = 1.00 \qquad (25).$$

Finally, let's consider the corrected regression coefficient $b_{1\eta}$ in equation (3). From equation (25), equation (1) may be rewritten $x_2 = \eta + e$, and by substitution, equation (3) becomes:

$$x_1 = b_{1\eta}(x_2 - e) + v \qquad (26),$$

$$x_1 = b_{1\eta}x_2 - b_{1\eta}e + v \qquad (27).$$

Multiplying equation (27) through by x_2 and taking expectations yields:

$$\sigma_{12} = b_{1\eta}\sigma_2^2 - b_{1\eta}\sigma_{2e} + \sigma_{2v} \qquad (28),$$

but $\sigma_{2v} = 0$ as above, and $\sigma_{2e} = \sigma_e^2$, which may be verified by multiplying equation (1) by e and taking expectations. Thus,

$$\sigma_{12} = b_{1\eta}\ (\sigma_2^2 - \sigma_e^2) \qquad (29);$$

$$b_{1\eta} = \sigma_{12}/(\sigma_2^2 - \sigma_e^2) \quad = \quad \sigma_{12}\sigma_2^2/(\sigma_2^2\{\sigma_2^2 - \sigma_e^2\}) \qquad (30),$$

and

$$b_{1\eta} = b_{12}\ (\sigma_2^2/\{\sigma_2^2 - \sigma_e^2\}) \qquad (31).$$

That is, the corrected regression coefficient of respondent's SEI on true education is equal to the OLS estimate only when the error of measurement variance is zero. Otherwise, the greater the variance of errors, the greater will be the downward bias in b_{12}. In this example, because the variance of errors was greater for blacks than for whites, the OLS estimates were more biased for blacks.

REFERENCES

Alwin, Duane F., and David J. Jackson
forthcoming "Measurement error models for survey data: Issues and applications." K. F. Schuessler (ed.), Sociological Methodology 1980. San Francisco: Jossey-Bass.

Anderson, James G.
1978 "Causal models in educational research: Nonrecursive models." American Educational Research Journal 15: 81-97.

Anderson, James G., and Evans, F. B.
1974 "Causal models in educational research: Recursive models." American Educational Research Journal 11: 29-39.

Bielby, William T., Robert M. Hauser, and David L. Featherman
1977 "Response errors of black and nonblack males in models of the intergenerational transmission of socioeconomic status." American Journal of Sociology 82: 1242-1288.

Blalock, Hubert M.
1964 Causal Inferences in Nonexperimental Research. Chapel Hill: University of North Carolina Press.

1965 "Some implications of random measurement error for causal inferences." American Journal of Sociology 71: 37-47.

Blalock, Hubert M., Caryll S. Wells, and Lewis F. Carter
1970 "Statistical estimation with random measurement error." Pp. 75-103 in Edgar F. Borgatta and George W. Bohrnstedt (eds.), Sociological Methodology 1970. San Francisco: Jossey-Bass.

Blau, Peter M., and Otis Dudley Duncan
1967 The American Occupational Structure. New York: John Wiley.

Bohrnstedt, George W.
1969 "Observations on the measurement of change." Pp. 113-133 in Edgar F. Borgatta and George W. Bohrnstedt (eds.), Sociological Methodology 1969. San Francisco: Jossey-Bass.

Bohrnstedt, George W., and T. Michael Carter
1971 "Robustness in regression analysis." Pp. 118-146 in Herbert L. Costner (ed.), Sociological Methodology 1971. San Francisco: Jossey-Bass.

Cooley, William W.
1978 "Explanatory observational studies." Educational Researcher 7 (October): 9-15.

Duncan, Otis Dudley
1961 "A socioeconomic index for all occupations." Pp. 109-138 in Albert J. Reiss, Jr., Occupations and Social Status. New York: The Free Press of Glencoe.

1969a "Inheritance of poverty of inheritance of race?" Pp. 85-110 in D. P. Moynihan (ed.), On Understanding Poverty. New York: Basic Books.

1969b "Some linear models for two-wave, two-variable panel analysis." Psychological Bulletin 72: 177-182.

Duncan, Otis Dudley, David L. Featherman, and Beverly Duncan
1972 Socioeconomic Background and Achievement. New York: Seminar Press.

Duncan, Otis Dudley, A. O. Haller, and A. Portes
1968 "Peer influences on aspirations: A reinterpretation." American Journal of Sociology 74: 119-137.

Gulliksen, Harold
1950 Theory of Mental Tests. New York: John Wiley.

Hauser, Robert M.
1969 "Schools and the stratification process." American Journal of Sociology 74: 587-611.

1973 "Disaggregating a social-psychological model of educational attainment." Pp. 255-284 in Arthur S. Goldberger and Otis Dudley Duncan (eds.), Structural Equation Models in the Social Sciences. New York: Seminar Press.

Hauser, Robert M., and David L. Featherman
1977 The Process of Stratification: Trends and Analyses. New York: Academic Press.

Hauser, Robert M., and Arthur S. Goldberger
1971 "The treatment of unobservable variables in path analysis." Pp. 81-117 in Herbert L. Costner (ed.), Sociological Methodology 1971. San Francisco: Jossey-Bass.

Hodge, Robert W., and Donald J. Treiman
1968 "Social participation and social status." American Sociological Review 33: 723-740.

Jöreskog, Karl G.
1973 "A general method for estimating a linear structural equation system." Pp. 85-112 in Arthur S. Goldberger and Otis Dudley Duncan (eds.) Structural Equation Models in the Social Sciences. New York: Seminar Press.

Jöreskog, Karl G.
1977 "Structural equation models in the social sciences: Specification, estimation and testing." Pp. 265-287 in P. R. Krishnaiah (ed.), Application of Statistics. Amsterdam: North-Holland Publishing Co.

Jöreskog, Karl G., and Dag Sörbom
1978 Analysis of Linear Structural Relationships by the Method of Maximum Likelihood. Chicago: National Educational Resources, Inc.

Kerlinger, Fred N.
1977 "The influence of research on education practice." Educational Researcher 6 (September): 5-12.

Kim, J., and Charles W. Mueller
1976 "Standardized and unstandardized coefficients in causal analysis: An expository note." Sociological Methods and Research 4: 423-438.

Land, Kenneth C.
1970 "On the estimation of path coefficients for unmeasured variables from correlations among observed variables." Social Forces 48: 506-511.

Levinsohn, Jay R., Louise B. Henderson, John A. Riccobono, and R. Paul Moore
1978 National Longitudinal Study Base Year, First, Second and Third Follow-Up Data File Users Manual, Volumes I and II. Wahsington, D.C.: National Center for Education Statistics.

Lichtman, Marilyn, and Lee M. Wolfle
1978 Analysis of the NLS Data: Processes of Educational and Status Attainment in Distinctive Subgroups: A Proposal in Response to RFP 78-76, Educational and Occupational Development of American Youths: Analyses of the National Longitudinal Study of the High School Class of 1972. Washington, D.C. National Center for Education Statistics.

Long, J. Scott
1976 "Estimation and hypothesis testing in linear models containing measurement error: A review of Joreskog's model for the analysis of covariance structures." Sociological Methods and Research 5: 157-206.

National Opinion Research Center
1977 Codebook for the Spring 1977 General Social Survey. Chicago: National Opinion Research Center.

Otto, Luther B., and David L. Featherman
1975 "Social structural and psychological antecedents of self-estrangement and powerlessness." American Sociological Review 40: 701-719.

Scarr, Sandra, and Richard A. Weinberg
1978 "The influence of 'family background' on intellectual attainment." American Sociological Review 43: 674-692.

Schuman, Howard, and Stanley Presser
1978 "The assessment of 'no opinion' in attitude surveys." Pp. 241-275 in Karl F. Schuessler (ed.), Sociological Methodology 1979. San Francisco: Jossey-Bass.

Sewell, William H., and Robert M. Hauser
1975 Education, Occupation, and Earnings: Achievement in the Early Career. New York: Academic Press.

Siegel, Paul M., and Robert W. Hodge
1968 "A causal approach to the study of measurement error." Pp. 28-59 in H. M. Blalock and A. B. Blalock (eds.), Methodology in Social Research. New York: McGraw-Hill.

Sörbom, Dag
1975 "Detection of correlated errors in longitudinal data." The British Journal of Mathematical and Statistical Psychology 28: 138-151.

Wiley, David E.
1973 "The identification problem for structural equation models with unmeasured variables." Pp. 69-83 in Arthur S. Goldberger and Otis Dudley Duncan (eds.), Structural Equation Models in the Social Sciences. New York: Seminar Press.

Wiley, David E., and James A. Wiley
1970 "The estimation of measurement error in panel data." American Sociological Review 35: 112-117.

Williams, John D.
1978 "Path analysis from a regression perspective." Multiple Linear Regression Viewpoints, Monograph Series No. 3, Volume 9.

Wolfle, Lee M.
1977 "An introduction to path analysis." Multiple Linear Regression Viewpoints 8: 36-61.

1978 "Strategies of path analysis." A paper presented to the American Educational Research Association, Toronto, March, 1978.

Wright, Sewell
1921 "Correlation and causation." Journal of Agricultural Research 20: 557-585.

1925 Corn and Hog Correlations. Washington, D.C.: U.S. Department of Agriculture Bulletin 1300, January 1925.