MS-831: Jack, Joseph and Morton Mandel Foundation Records, 1980–2008.
Series D: Adam Gamoran Papers. 1991–2008.
Subseries 4: The Jewish Indicators Project, 1996–2000.

Box	Folder
66	9

Kaplan, David. Consultation proposal. Curriculum vitae, 1997-1999.

Mandel Foundation

July 7, 1999

Dr. David Kaplan
School of Education
University of Delaware
Newark, DE 19716

Dear David,

I am delighted that you have agreed to work with us on the Indicators project. At this point, we have asked you to do the following work:

1. Cull GSS, SASS, and possible other data sets (e.g. NLSY) for data on Jewish education/ Jewish identity.

2. Refine (if possible) the earlier CIJE report-- essentially, look at the data from different perspectives (e.g. adjusted means analysis).

3. Produce a report that summarizes the findings of the data review and demonstrates what can be learned from an indicators study in order to make the case for future such studies.

This work is to be done by the middle of August. At that point we will re-group with our advisory committee and decide how to go forward. If we continue on the track set at our February meeting, the next steps would include:

- Producing a second report based on the suggestions you make based on your work in Step 1 above.
- Create a dissemination plan for the first report that you create.

Based on the amount of work that you and I estimated, we are assuming that, between now and mid-August, you will work approximately 10 days at the rate of $500 a day. At that point, we will estimate how much time it will take to do the second report.

We are looking forward to working together.

Sincerely,

Gail Dorph
Senior Education Officer

C.C. Adam Gamoran
Mark Gurvis
Annette Hochstein

Vita

David Kaplan

Work:

School of Education
University of Delaware
Newark, DE. 19716
Voice Mail: 302-831-8696
Fax: 302-831-4445
Internet: dkaplan @ udel.edu
Web page: http://www.udel.edu/dkaplan

Home:

727 Colgate Lane
Newark, DE 19711
Ph: 302-733-0512

Education

1987 Ph. D., Education, University of California, Los Angeles
Major area: Educational Statistics and Psychometrics.
Cognate area: Econometrics
Minor area: Sociology of Education

1983 M. A., Education, University of California, Los Angeles.

1978 B. A. (Cum Laude), Psychology, California State University, Northridge.

Present Position Professor, School of Education and Department of Psychology, University of Delaware. Associate Professor, 1992- 1998; Assistant Professor, 1987-1992, University of Delaware.

Refereed Journal Publications

In press

Kaplan, D. On the extension of the propensity score adjustment method for the analysis of group differences in MIMIC models. *Multivariate Behavioral Research.*

1999

Kaplan, D. & Ferguson, A. J. On the utilization of sample weights in latent variable models. *Structural Equation Modeling*, *6*, 305-321.

1998

Kaplan, D., & George, R. Evaluating latent variable growth models through ex post simulation. *Journal of Educational and Behavioral Statistics*, *23*, 216-235.

George, R., & Kaplan, D. A structural model of parent and teacher influences on the science attitudes of eighth graders: Evidence from NELS:88. *Science Education, 82,* 3-17.

1997

Kaplan, D. & Elliott. P. R. A model-based approach to validating education indicators using multilevel structural equation modeling. *Journal of Educational and Behavioral Statistics, 22,* 323-348.

Kaplan, D., & Elliott, P. R. A didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling: A Multidisciplinary Journal, 4,* 1-24.

1995

Kaplan, D. The impact of BIB-spiralling induced missing data patterns on goodness-of-fit tests in factor analysis. *Journal of Educational and Behavioral Statistics*, *20*, 69-82.

Kaplan, D., & George, R. A study of the power associated with testing factor mean differences under violations of factorial invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 2*, 101-118.

1994

Kaplan, D., & Venezky, R. L. Literacy and voting behavior: A bivariate probit model with sample selection. *Social Science Research, 23*, 350-367.

Kaplan, D. Estimator conditioning diagnostics for covariance structure models. *Sociological Methods and Research*, *23*, 200-229.

1993

Kaplan, D., & Wenger, R. N. Asymptotic independence and separability in covariance structure models: Implications for specification error, power, and model modification. *Multivariate Behavioral Research*, *28,* 483-498.

Fromme, K., Stroot, E., & Kaplan, D. The comprehensive effects of alcohol: Development and psychometric assessment of a new expectancy questionnaire. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 5*, 19-26.

1992

Muthén, B., & Kaplan, D. A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology, 45,* 19-30.

1991

Kaplan, D. The behaviour of three weighted least squares estimators for structured means analysis with non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 44*, 333-346.

Kaplan, D. On the modification and predictive validity of covariance structure models. *Quality and Quantity, 25*, 307-314.

1990

Kaplan, D. Evaluation and modification of covariance structure models: A review and recommendation. *Multivariate Behavioral Research, 25*, 137-155.

Kaplan, D. Rejoinder on evaluating and modifying covariance structure models. *Multivariate Behavioral Research, 25*, 197-204

Kaplan, D. Contributions to structural modeling of mathematics achievement: Application of categorical variable structural equation methodology. *International Journal of Educational Research, 14*, 175-192.

Lapan, R. T., McGrath, E., & Kaplan, D. Factor structure of the Basic Interest Scales by gender across time. *Journal of Counseling Psychology, 37*, 216-222.

1989

Kaplan, D. Model modification in covariance structure analysis: Application of the expected parameter change statistic. *Multivariate Behavioral Research, 24,* 285-305.

Kaplan, D. Power of the likelihood ratio test in multiple group confirmatory factor analysis under partial measurement invariance. *Educational and Psychological Measurement, 49*, 579-586.

Kaplan, D. The problem of error rate inflation in covariance structure models. *Educational and Psychological Measurement, 49*, 333-337.

Kaplan, D. A study of the sampling variability and z-values of parameter estimates from misspecified structural equation models. *Multivariate Behavioral Research, 24*, 41-57.

1988

Kaplan, D. The impact of specification error on the estimation, testing, and improvement of structural equation models. *Multivariate Behavioral Research, 23,* 69-86.

1987

Muthén, B., Kaplan, D., & Hollis, M. On structural equation modeling with data that are not missing completely at random. *Psychometrika, 51*, 431-462.

1985

Muthén, B., & Kaplan, D. A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38,* 171-189.

Manuscripts Submitted for Publication

Ferguson, A. J., & Kaplan, D. (1998). The influence of track placement and classroom context on self-concept and locus-of-control: A propensity score analysis. Submitted to *American Educational Research Journal.*

Book Chapters and Annual Volumes

Venezky, R. L., & Kaplan, D. (1998). Literacy habits and political participation. In M. Cecil Smith (ed.), *Literacy for the 21st Century.* Westport, CN: Greenwood Publishing Group.

Kaplan, D. (1998). Methods for multilevel data analysis. In. G. A. Marcoulides (ed.), *Modern Methods for Business Research.* Mahwah, NJ: Lawrence Erlbaum and Associates.

Kaplan, D. (1996). An overview of concepts and issues in multilevel structural equation modeling. In H. Ernste (ed.), *Multilevel Analysis with Structural Equation Models.* (pp. 1-18). Zurich, Switzerland: Department of Geography, Swiss Federal Institute of Technology (ETH).

Kaplan, D. (1995). Statistical power in structural equation modeling. In R. H. Hoyle (ed.), *Structural Equation Modeling: Concepts, Issues, and Applications* (pp. 100-117). Newbury Park, CA: Sage Publications, Inc.

Kaplan, D. (1992). Structural equation modeling. In M. C. Alkin (ed.), *Encyclopedia of Educational Research 6th edition.* New York: Macmillian.

Glutting, J. J. & Kaplan, D. (1990). Stanford-Binet Intelligence Scale: Fourth Edition: Making the case for reasonable interpretations. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of Psychological and Educational Assessment of Children: Volume 1. Intelligence and Achievement.* (pp. 277-295). New York: The Guilford Press.

Conference Proceedings and Technical Reports

Kaplan, D. & Venezky, R. L. (1995). Literacy and voting behavior: A statistical analysis based on the 1985 Young Adult Literacy Survey. *NCAL Technical Report TR94-14.* Philadelphia: National Center on Adult Literacy.

Kaplan, D., & Elliott, P. R. (1994). A Multilevel Structural Model of Science Achievement From an Indicator System Perspective: Implications for Educational Policy Analysis. *Final Report to the AERA Grants Program Committee*

Kaplan, D.& Venezky, R. L. (1993). What can employers assume about the literacy skills of GED graduates? *NCAL Technical Report TR93-5.* Philadelphia: National Center on Adult Literacy.

Kaplan, D. & Wenger, R. N. (1993). Asymptotic independence and separability in covariance structure models. In R. Steyer, K. Wender, & K. Widaman (Eds.), *Psychometric Methodology: Proceedings of the 7th European Meeting of the Psychometric Society in Trier* (pp. 203-208). Stutgaart and New York: Gustav Fischer Verlag.

Kaplan, D. (1992). The Analysis of Adult Literacy Survey Data: Problems in Factor Analysis with BIB-Spiralled Item Administration. *NCAL Occasional Paper OP92-2.* Philadelphia: National Center on Adult Literacy.

Book Reviews

Kaplan, D. (1994). [Review of *Structural Equation Modeling with EQS and EQS/Windows: Basic Concepts, Applications, and Programming*]. *Applied Psychological Measurement., 18,* 191-192.

Kaplan, D. (1993). [Review of *Testing Structural Equation Models*]. *Structural Equation Modeling, 1,* 98-99.

Kaplan, D. (1990). [Review of *Multivariate Statistics: A Practical Approach*]. *Journal of Educational Statistics, 15,* 171-174.

Conference Presentations (refereed)

1999

Kaplan, D. *Dynamic multipliers in single level and multilevel models.* Paper presented at the annual meeting of the American Educational Research Association. Montreal, Canada.

Kaplan, D. *Statistical models applied to national data sets for informing education policy.* Symposium paper presented at the annual meeting of the American Educational Research Association. Montreal, Canada

Kaplan, D. Discussant for AERA paper session. *Topics in Structural Equation Modeling.* Montreal, Canada

1998

Kaplan, D. *On the assumptions associated with the application of the propensity score adjustment method to latent variable models.* Paper presented the annual meeting of the American Educational Research Association. San Diego, CA.

Kaplan, D. & Ferguson. *On the utilization of sampling weights in latent variable models.* Paper presented at the annual meeting of the American Educational Research Association. San Diego, CA.

Kaplan, D. Discussant for AERA symposium. *Modeling Non-Normal Data.* San Diego, CA.

1996

Kaplan, D. Evaluating latent variable models using ex post simulation. Paper presented at the North American Meeting of the Psychometric Society, Banff, Alberta, Canada.

Kaplan, D. *On the extension of the propensity score adjustment method to covariance structure modeling.* Paper presented at the annual meeting of the American Educational Research Association. New York, New York.

Kaplan, D. Discussant for AERA paper session. *Latent Trait Model Fit.* New York, New York.

Kaplan, D., Zuzovsky, R., & Tamir, P. *Parental involvement as perceived by Israeli pupils and their parents: A comparison of urban and kibbutz families.* Symposium paper presented at the annual meeting of the American Educational Research Association. New York, New York.

1995

Kaplan, D. *Literacy and voting behavior: A bivariate probit model with sample selection.* Symposium paper presented at the annual meeting of the American Educational Research Association. San Francisco, CA.

Kaplan, D. & Elliott, P. R. *Centering problems in multilevel covariance structure modeling.* Paper presented at the annual meeting of the American Educational Research Association. San Francisco, CA.

Kaplan, D., & Elliott, P. R. *Validating science education indicators through quantitative policy modeling: Evidence from NELS:88.* Paper presented at the annual meeting of the American Educational Research Association. San Francisco, CA.

1994

Kaplan, D., & Elliott, P. R. *A policy guidance system for science achievement: An application of multilevel structural equation modeling*. Paper presented at the annual meeting of the American Educational Research Association. New Orleans, Louisiana.

Kaplan, D., & Elliott, P. R. *On the utility of multilevel structural equation modeling for building an educational policy guidance system.* Paper presented at the annual meeting of the American Educational Research Association. New Orleans, Louisiana.

Kaplan, D. Discussant for AERA paper session: *Theoretical developments and educational applications of covariance structure analysis*. New Orleans, Louisiana.

1993

Kaplan, D., & Wenger, R. N. *Asymptotic independence and separability in covariance structure models: Implications for specification error, power, and model modification*. Paper presented at the meeting of the American Educational Research Association. Atlanta, Georgia.

1992

Kaplan, D. *Collinearity diagnostics for covariance structure models*. Presented at the meeting of the Psychometric Society. Columbus, OH.

Kaplan, D. *Assessing the factor structure of data arising from balanced incomplete block spiralling designs.* Presented at the meeting of the American Educational Research Association. San Francisco, CA.

Kaplan, D. *Collinearity diagnostics for covariance structure models*. Presented at the meeting of the American Educational Research Association. San Francisco, CA.

1991

Kaplan, D., & Wenger, R. N. *Asymptotic independence and separability in covariance structure models.* Presented at the European meeting of the Psychometric Society. Trier, Germany.

Kaplan, D., & Wenger, R. N. *Asymptotic independence and separability in covariance structure models.* Presented at the joint meeting of the Classification Society and Psychometric Society. New Brunswick, NJ.

Kaplan, D. *A Monte Carlo study of three weighted least squares estimators for structured means analysis with non-normal Likert variables.* Presented at the meeting of the American Educational Research Association. Chicago, IL

Kaplan, D. *A study of the power associated with testing factor mean differences under violations of factorial invariance.* Presented at the meeting of the American Educational Research Association. Chicago, IL.

1990

Kaplan, D. *The behavior of three weighted least squares estimators for structured means analysis with non-normal Likert variables.* Presented at the meeting of the Psychometric Society. Princeton, NJ.

Kaplan, D. *A multistage method for studying mean structures in multiple group higher order confirmatory factor analysis.* Presented at the meeting of the American Educational Research Association. Boston, MA.

Kaplan, D. *Alternative fit indices in covariance structure modeling: Just say whoa!* Presented at the meeting of the American Educational Research Association. Boston, MA.

Kaplan, D. Discussant for AERA/NCME symposium: *The assessment of test anxiety: Applications of covariance modeling to issues of construct validation.* Boston, MA

1989

Kaplan, D. *Power of the likelihood ratio test in multiple group confirmatory factor analysis under partial measurement invariance.* Presented at the 6th European meeting of the Psychometric Society. Leuven, Belgium.

Kaplan, D. *On the modification and selection of competing covariance structure models.* Presented at the meeting of the American Educational Research Association. San Francisco, CA.

Kaplan, D. *On the utility of classical statistical theory for building and evaluating covariance structure models.* Presented at the meeting of the American Educational Research Association. San Francisco, CA.

1988

Kaplan, D. *On specification error problems in covariance structure models.* Presented at the meeting of the Psychometric Society. Los Angeles, CA.

Kaplan, D. *Modification of structural equation models: Application of the expected parameter change statistic.* Presented at the meeting of the American Educational Research Association. New Orleans, LA.

Kaplan, D. *A Monte Carlo study of the sampling variability and z-values of parameter estimates for misspecified structural equation models.* Presented at the meeting of the American Educational Research Association. New Orleans, LA.

1987

Kaplan, D. *The impact of specification error on the estimation, testing, and improvement of structural equation models.* Presented at the meeting of the American Educational Research Association. Washington, D.C.

1985

Muthén, B., & Kaplan, D. *On comparing item characteristic curve parameters across groups.* Presented at the meeting of the American Educational Research Association. Chicago, IL.

Muthén, B., Kaplan, D., & Hollis, M. *Latent variable modeling with missing data: Attrition in longitudinal studies.* Presented at the meeting of the American Educational Research Association. Chicago, IL.

1984

Muthén, B., & Kaplan, D. *A comparison of some methodologies for the factor analysis of non-normal Likert variables*. Presented at the meeting of the American Educational Research Association. New Orleans, LA.

Invited Addresses, Workshops and Conferences

1999

Kaplan, D. *Invited Workshop on Hierarchical Linear Modeling*. Presented to the School of Education, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem, Israel.

Kaplan, D. *A Multilevel Model of the Effects of School Choice on Academic Achievement.* Presented to the School of Education, Hebrew University of Jerusalem, Israel.

1998

Kaplan, D. *Invited Workshop on Structural Equation Modeling*. Presented to the School of Education, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem, Israel.

Kaplan, D. *Elements of univariate and multivariate growth curve modeling*. Presented to the School of Education, Hebrew University of Jerusalem, Israel, and the School of Education, Tel-Aviv University, Israel.

Kaplan, D. *On the extension of the propensity score adjustment method for the analysis of group differences in latent variable models*. Presented at the 1998 meeting of the Israeli Sociological Association, Haifa, Israel.

1997

Kaplan, D. *On the use of latent variable growth modeling for monitoring change in multiple achievement domains.* Department of Psychology, College of William and Mary, Williamsburg, VA.

Kaplan, D. Invited Workshop on Structural Equation Modeling. Presented to the Faculty of Social Sciences, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem, Israel.

Kaplan, D. *Structural Equation Modeling*. Invited AERA Graduate Student Seminar Roundtable. Annual meeting of the American Educational Research Association. Chicago, IL.

Kaplan, D. *Statistical modeling of hierarchy, structure, and temporality in complex organizations: An example from education.* Presented to the Faculty of Social Sciences, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem, Israel.

1995

Kaplan, D. *Modeling science education indicators: An application of multilevel structural equation modeling*. Presented to the Delaware Chapter of the American Statistical Association. Newark, DE.

Kaplan, D. *Recent developments and future directions in structural equation modeling*. Presented to the Department of Statistics, Tel Aviv University, Tel Aviv, Israel.

Kaplan, D. *Modeling and validating science education indicators*. Presented to the School of Education, Tel Aviv University, Tel-Aviv, Israel.

Kaplan, D. Invited participant in *Conference on Analytic Uses of Longitudinal Databases*. Washington, DC.

1994

Kaplan, D. *The utility of multilevel structural equation modeling for organizational policy studies: The case of education*. Presented to the RMD Conference on Causal Modeling. Purdue University.

1990

Kaplan, D. *Covariance structure modeling*. Presented to the Delaware Chapter of the American Statistical Association. Newark, Delaware.

Grants and Awards

Jordan, N. & Kaplan, D. National Institute of Child Health and Human Development. *A developmental study of mathematics disabilities.* 1999-2002. Amount: $449,216

Kaplan, D. The Spencer Foundation. *Developing longitudinal statistical models for education policy.* 1998-2001. Amount: $125,000.

D. Archbald, Kaplan, D., & Y. Nakib. U. S. Department of Education, OERI National Institute on Educational Governance, Finance, Policy-Making and Management (OERI # R308F60010). *A National Study of the Effects of School Choice on Achievement and Opportunity.* 1996-1998. Amount: $419,926.

Kaplan, D. National Science Foundation (# REC--9550472). *Model-based indicator systems for informing science education policy.* 1995-1997. Amount: $168,516.

Venezky, R. L, & Kaplan, D. U.S. Department of Education, OERI National Center on Adult Literacy. Project title: *Studies of adult literacy skills and assessment.* 1994-1995. Amount: $260,000.

Kaplan, D. University of Delaware International Programs and Special Sessions International Travel Grant. Project Title: *Modeling School Effectiveness in Israeli Schools.* 1995. Amount:: $2,200.

Kaplan, D. American Educational Research Association (NSF # RED-9255347). Project title: *Quantitative approaches to educational policy analysis utilizing multilevel structural equation modeling.* 1993-1994. Amount: $15,000.

Kaplan, D. U.S. Department of Education, OERI National Center on Adult Literacy. Project title: *Models of literacy and literacy related behaviors.* 1991-1992. Amount: $53,242.

Kaplan, D. University of Delaware General University Research Grant. Project title: *Specification error issues in multiple populations.* 1988-1989. Amount: $5,000.

Professional Affiliations

Member:
American Educational Research Association (Division D and SIG/Educational Statisticians); Delaware Chapter of the American Statistical Association; National Council on Measurement in Education, Psychometric Society

Professional Activities

Editorial Boards:
Educational and Psychological Measurement; Journal of Educational and Behavioral Statistics (including Management Committee); Journal of Educational Research; Multivariate Behavioral Research; Structural Equation Modeling

Program Chair:
American Educational Research Association (Division D, Section 3a), 1994.

Conference Session Chair:
American Educational Research Association; Psychometric Society

Ad hoc Reviewer:
American Educational Research Journal; British Journal of Mathematical and Statistical Psychology; Child Development; Computational Statistics and Data Analysis; Educational Evaluation and Policy Analysis; Journal of Educational Measurement; Journal of Educational Research; Journal of Studies on Alcohol; Measurement and Evaluation in Counseling and Development; Multivariate Behavioral Research; Organizational Research Methods; Psychological Bulletin; Psychometrika; Structural Equation Modeling; AERA Division D; APA Division 5.

Grant Reviewer:
NAEP Data Reporting Grant Program; The Spencer Foundation

University Activities

University:
Advisory Council on Graduate Studies
Senate Committee on Graduate Studies
Senate Committee on Instructional, Computing and Research Support Services
University Statistical Laboratory Advisory Committee
Jewish Studies Program Advisory Committee

College:
Acting Director, Center for Educational Leadership and Policy
Dean's Research Advisory Council
College Committee on Graduate Studies in Education
Delaware Educational Research and Development Center Advisory Board
Member: Center for Educational Leadership and Policy

Department:
Ad hoc Committee on Departmental Computer Networking
Chair, Curriculum Committee
Faculty Development Committee
Promotion and Tenure Committee
Search Committees (as needed)
Coordinator of Doctoral Program in Measurement, Statistics, and Evaluation (rotating)
Chairs Advisory Committee

Journal of Educational and Behavioral Statistics
Fall 1997, Vol. 22, No. 3, pp. 323–347

A Model-Based Approach to Validating Education Indicators Using Multilevel Structural Equation Modeling

David Kaplan
University of Delaware

Pamela R. Elliott
University of Rochester

Keywords: *education indicators, education policy, multilevel modeling, structural equation modeling*

This article considers an approach to validating the selection of education indicators by incorporating them into a multilevel structural model and using the estimates from that model to engage in policy-relevant simulations. Multilevel structural equation modeling was applied to a subsample of the first follow-up of the National Education Longitudinal Study of 1988 (National Center for Education Statistics, 1988) to demonstrate the potential of this approach. Focus of attention was on science education indicators. A within-school model of science achievement was linked to a between-school model of the academic press of the school. Separate estimation of these models revealed adequate fit to the data after minor modifications. The multilevel model also showed adequate fit to the data. Utilizing the reduced form of the full multilevel model, predictive validity of the model was studied by gauging movements in various outcome indicators as a function of changes in policy-relevant input indicators. The article closes with a discussion of the limitations of the proposed modeling approach, the potential for future model development, and the implications of this approach for quantitative modeling within the domain of education policy.

Over the past several years, national attention has focused on the educational performance of U.S. students—particularly in the areas of science and mathematics. Indeed, the often lamented poor performance of U.S. students in

This research was supported by a grant from the American Educational Research Association, which receives funds for its grants program from the National Science Foundation and the National Center for Education Statistics (U.S. Department of Education) under NSF Grant No. RED-9255347. Additional work was supported by the National Science Foundation under Grant REC-9550472. Opinions reflect those of the authors and do not necessarily reflect those of the granting agencies.

The authors wish to acknowledge the valuable advice and assistance of Bengt Muthén and Ginger Nelson Goff.

science and mathematics compared to our key economic competitors Germany and Japan led to the Education Goals 2000 legislation, which stated, among other things, that "by the year 2000, U.S. students will be first in the world in science and mathematics achievement" (National Education Goals Panel, 1994). Unfortunately, precious few years remain to implement the necessary changes in our educational system to meet this goal. Nevertheless, if the United States is to succeed in attaining the target set by the National Education Goals Panel, it is essential that policymakers have a clear understanding of salient determinants of science and mathematics achievement. Past discussions on the development of education indicators (see, e.g., Murnane & Raizen, 1988; Shavelson, McDonnell, & Oakes, 1989) represent an attempt at understanding those determinants.

When reviewing the extant literature on education indicators it is not uncommon to find, as a starting point, the theoretical model of schooling shown in Figure 1. This model derives from the organizational paradigm in educational research (see, e.g., Oakes, 1986). There are a number of features of the organizational model displayed in Figure 1 worth considering in light of the purposes of this study. First, while this model does not represent a casual path diagram in any formal sense, the organizational model does suggest potentially testable structural relationships among the variables comprising the inputs, processes, and outputs of schooling. Second, the model is inherently multilevel in form, with a subset of the inputs and processes occurring at higher levels of the educational system—for example, teachers and classrooms, and/or schools. To take one example, the model suggests that the fiscal resources of the school have an indirect effect on student achievement through structurally related teacher-level variables such as teaching quality and instructional quality.

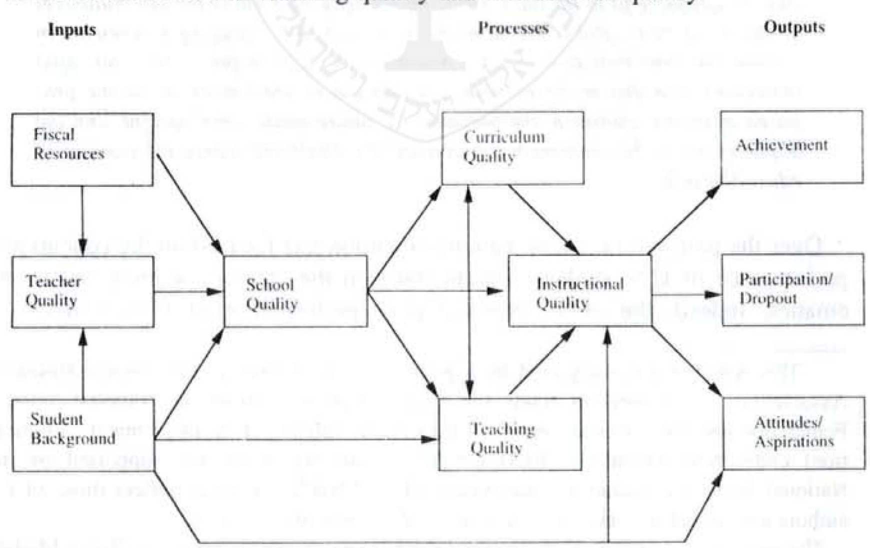

FIGURE 1. *A theoretical model of the U.S. educational system*
Note. From Oakes (1986)

The purpose of this article is to propose a model-based approach for validating education indicators that explicitly takes into account the organizational features of schooling. This approach attempts to make explicit the philosophical position taken by de Neufville (1978), who wrote, "Theoretical indicators can also be validated by looking at their movements in relation to indicators of other variables when an interrelationship is presumed" (p. 177). The approach advocated in this article has two distinguishing components. First, this approach requires the selection of indicators that (a) have been theoretically and, in some cases, empirically linked to an outcome of interest, (b) can serve the function of monitoring real or simulated changes in education outcomes, and (c) can act as proxies for policy instruments. Second, this approach recognizes and explicitly accounts for the fact that indicator data are perhaps best modeled as systems of structural equations operating at different levels of the educational organization. Specifically, indicator data are viewed as being generated from clustered sampling schemes wherein student-level variables that predict outcomes of interest, such as performance in science, may vary as a function of school-level differences in school climate, policies, and resources that perhaps follow their own set of structural relations. Thus, we will argue that the appropriate statistical methodology to be utilized in specifying and estimating this type of model is multilevel structural equation modeling.

The organization of this article is as follows. The next section describes the statistical model employed in this study. A brief description of model estimation will be included, and references to relevant technical papers on the subject will be provided. Following the description of the statistical model, the data source and variables will be described. This will be followed by the results of a multilevel structural model of science achievement. Next, a series of policy-relevant simulations will be conducted in order to examine the utility of the model for validating education indicators. This section of the article is for demonstration purposes only and should not be construed as an attempt at serious policy analysis in the domain of science education. The article will conclude with a discussion of how the methodological approach developed in this article is consistent with a general philosophy of indicator validation.

Statistical Modeling of Education Indicators

Preliminary Background

In order for indicator systems to accurately gauge the health science education they must reasonably reflect the complexities of the educational system (see, e.g., Darling-Hammond, 1992; Oakes, 1986; Porter, 1991). One feature of the educational system that must be considered in a properly specified model is the multilevel organizational structure of schooling and the related fact that data generated from schooling research often arise from clustered sampling schemes. For example, in the National Education Longitudinal Study of 1988 (National Center for Education Statistics [NCES], 1988) students are sampled in classrooms that are, in turn, sampled from schools. Extraordinary advances in school process research have resulted from the application of *multilevel linear regression* methods to account for these clustered sampling schemes (see, e.g., Aitkin & Longford, 1986; Bock, 1989; Bryk & Raudenbush, 1992; Raudenbush & Willms, 1991; Willms, 1992). The benefit of applying multilevel regression methods to educational data is that they mitigate many of the statistical problems that are typically encountered when clustered sampling is ignored—specifically, biases arising from aggregation or disaggregation of the data (see, e.g., Bryk & Raudenbush, 1992). Moreover, and of great relevance to this study, multilevel models allow for the estimation of cross-level effects.

An inspection of the organizational model in Figure 1 suggests, however, that indicators are also structurally (i.e., causally) related and that modeling only the multilevel features of education would ignore the structural relationships within and between levels of the educational system. To handle these types of relationships, structural equation modeling is required. Structural equation modeling represents a major methodological breakthrough in the study of complex interrelationships among variables (see, e.g., Bollen, 1989; Jöreskog, 1977). Structural equation modeling embodies the unification of two methodological traditions: Thurstonian factor analysis originating from psychology and psychometrics, and simultaneous equations (path analytic) modeling originating from econometrics and finding its way to educational research through sociology (see, e.g., Goldberger & Duncan, 1973).

There is no question that structural equation modeling has been, and continues to be, a popular tool in the domain of quantitative educational research. Indeed, with respect to testing organizational theories of schooling, early work utilizing structural modeling methods includes Levin (1970); Michelson (1970); Boardman, Davis, and Sanday (1977); Bidwell and Kasarda (1975); and Cohn and Millman (1975). Moreover, Glasman and Biniaminov (1981), in their review of input-output schooling studies, proposed, but did not estimate, a structural model of the schooling process and alluded to the potential of such a model for policy studies.

Despite the ubiquitous application of structural equation modeling in education research, some policy analysts have argued that research on schooling may not be sufficiently developed to admit strong causal inferences (e.g., McDonnell, 1989). Others (e.g., Darling-Hammond, 1992) have expressed concern that attempting to statistically assess causal relationships in cross-sectional indicator data may lead to inappropriate inferences. We argue, on the other hand, that it is possible to statistically capture the salient multilevel and structural relationships of the educational process and incorporate them into a model-based indicator system through the judicious and progressive application of *multilevel structural equation modeling* (Muthén, 1989, 1994; Muthén & Satorra, 1989). Multilevel structural equation modeling affords analysts the ability to specify a system of student-level equations for predicting student-level outcomes, while at the same time modeling variation in the student-level indicators as a function of possibly different systems of equations for the between-classroom or between-school indicators. We will attempt to show that multilevel structural modeling provides a rational quantitative strategy which, when applied to theory-generated collections of indicators, can be used for indicator system validation.

A Multilevel Structural Equation Model

To simplify the discussion and prepare the groundwork for the application of multilevel structural equation modeling to science education indicators we will first assume that the indicators possess reasonably good validity and reliability. We recognize that this assumption may be unreasonable in most cases, but we hasten to point out that the model outlined below can be extended to handle the case where multiple measures of each construct are obtained and integrated into the analysis by making use of multilevel-based measurement models (see Muthén, 1991, 1994).

Consider first the within-school model. It is assumed that the intercepts and means of the student-level indicators vary over schools and that there exists a model which holds at the school level that explains variation in the intercepts and means of the student-level indicators. We will also assume that the slopes are fixed. We begin by writing the within-school model as

$$\mathbf{y}_{ig} = \boldsymbol{\alpha}_g + \mathbf{B}_y \mathbf{y}_{ig} + \boldsymbol{\epsilon}_{ig}, \tag{1}$$

where $\mathbf{y}_{ig}$ is a vector of student-level indicators (e.g., science achievement, time spent on homework, etc.), some of which are exogenous for the *i*th student ($i = 1, \ldots, N$) in the *g*th ($g = 1, \ldots, G$) school, $\boldsymbol{\alpha}_g$ is a vector of parameters containing average values (intercepts and means) of the student-level indicators which are assumed to vary over schools, $\mathbf{B}_y$ in a matrix of regression coefficients relating the student-level indicators to each other, and $\boldsymbol{\epsilon}_{ig}$ is a disturbance term for the student-level equation.

In the language of structural equation modeling via LISREL (Jöreskog & Sörbom, 1991), (1) is referred to as an *all-y model*. In the all-y specification, all variables are treated as endogenous variables. Thus, $\mathbf{y}_{ig}$ is a *p*-dimensional vector of variables, where the first, say, $p - q$ variables are endogenous variables, while the last q variables are exogenous variables. The remaining matrices are given appropriate dimensions. So, for example, the first $p - q$ elements of $\boldsymbol{\alpha}_g$ are intercepts in the usual sense of the word, while the last q elements of $\boldsymbol{\alpha}_g$ are the means of the exogenous variables. For the purposes of this study, the all-y specification is used to simplify the notation and does not result in any loss of generality.

The model expressed in (1) is referred to as the *structural form* of the within-school part of the model and represents the statistical relationships among the indicators as they are deduced by the organizational model for student-level relationships. The direct effects are contained in the $\mathbf{B}_y$ matrix and are those that are typically displayed in path diagrams. For the purposes of this study we wish to model variation in the intercepts and means of the student-level indicators. To accomplish this, it is useful to reexpress (1) in its reduced form as

$$\mathbf{y}_{ig} = (\mathbf{I} - \mathbf{B}_y)^{-1}\boldsymbol{\alpha}_g + (\mathbf{I} - \mathbf{B}_y)^{-1}\boldsymbol{\epsilon}_{ig}\,, \tag{2}$$

where it is assumed that the inverse of $(\mathbf{I} - \mathbf{B}_y)$ exists (see Muthén, 1994).

As stated earlier, we assume that the levels of the student-level indicators (contained in $\boldsymbol{\alpha}_g$) vary across the G schools and that this variation can be explained by school-level indicators. Thus, we write a between-school model for the intercepts and means as

$$\boldsymbol{\alpha}_g = \boldsymbol{\alpha} + \mathbf{B}_\alpha \mathbf{z}_g + \boldsymbol{\delta}_g\,, \tag{3}$$

where $\boldsymbol{\alpha}$ is the grand mean vector across the G schools, $\mathbf{z}_g$ are school-level endogenous and exogenous indicators (such as the academic press of the school, professional teaching conditions, etc.), $\mathbf{B}_\alpha$ is a matrix of regression coefficients relating $\mathbf{z}_g$ to the intercepts of the student indicators, and $\boldsymbol{\delta}_g$ is a vector of disturbances for the intercept equation.

Up to this point the model described in (1), (2), and (3) allows intercepts and means to be expressed as a function of school-level indicators. This model has been discussed in detail in Muthén (1994). A unique feature of the model presented here is that the between-school indicators $\mathbf{z}_g$ are allowed to follow a separate between-school simultaneous equation model that can be written as

$$\mathbf{z}_g = \boldsymbol{\tau} + \mathbf{B}_z \mathbf{z}_g + \mathbf{u}_g \tag{4}$$

(see Muthén, 1994). Assuming that the inverse of $(\mathbf{I} - \mathbf{B}_z)$ exists, (4) can be reexpressed in reduced form as

$$\mathbf{z}_g = (\mathbf{I} - \mathbf{B}_z)^{-1}\boldsymbol{\tau} + (\mathbf{I} - \mathbf{B}_z)^{-1}\mathbf{u}_g\,, \tag{5}$$

where $\boldsymbol{\tau}$ is a vector of intercepts for the school-level endogenous indicators, $\mathbf{B}_z$ is a matrix of coefficients relating school-level indicators to each other, and $\mathbf{u}_g$ is a vector of disturbances for the school-level indicator equation.

After a series of substitutions we arrive at the expression for the ith student's score in the gth school, taking into account the structural relationships within as well as between schools. This final model can be written as

$$\mathbf{y}_{ig} = (\mathbf{I} - \mathbf{B}_y)^{-1}\boldsymbol{\alpha} + \boldsymbol{\Pi}\boldsymbol{\tau} + \boldsymbol{\Pi}\mathbf{u}_g + (\mathbf{I} - \mathbf{B}_y)^{-1}\boldsymbol{\delta}_g + (\mathbf{I} - \mathbf{B}_y)^{-1}\boldsymbol{\epsilon}_{ig}\,, \tag{6}$$

where $\boldsymbol{\Pi} \equiv (\mathbf{I} - \mathbf{B}_y)^{-1}\mathbf{B}_\alpha(\mathbf{I} - \mathbf{B}_z)^{-1}$ contains the regression coefficients relating school-level exogenous variables to student-level endogenous variables taking into account between-school structure and within-school structure. In structural equation modeling terminology, the matrix $\boldsymbol{\Pi}$ can be considered a *multilevel total effects matrix*.

Specific details of the estimation of the parameters of this model, as well as software considerations, are beyond the scope of this article (see Muthén, 1994, for details). Generally, though, Muthén (1994) considers the following three sample covariance matrices: (a) the total sample covariance matrix ($\mathbf{S}_T$) used in conventional structural equation modeling, which estimates $\boldsymbol{\Sigma}_W + \boldsymbol{\Sigma}_B$ in the multilevel case, where $\boldsymbol{\Sigma}_W$ and $\boldsymbol{\Sigma}_B$ are within- and between-group population covariance matrices, respectively; (b) the pooled within-group sample covariance matrix ($\mathbf{S}_{PW}$), which is a consistent and unbiased estimator of $\mathbf{\Sigma}_W$; and (c) the sample between-group covariance matrix ($\mathbf{S}_B$), which is a consistent and unbiased estimator of $\mathbf{\Sigma}_W + c\mathbf{\Sigma}_B$, where c reflects group size.

In the case of equal within-group sample sizes, a full information maximum likelihood (FIML) estimator can be obtained. However, in cases of unbalanced data, the FIML estimator would require a term for each distinct group size. To remedy this, Muthén (1994) developed a quasi-likelihood estimator, referred to as MUML (Muthén's ML-based estimator), that can be written as

$$\begin{aligned} F_{\mathrm{MUML}} = {} & G\{\ln|\mathbf{\Sigma}_W + c\mathbf{\Sigma}_B| + \mathrm{trace}[(\mathbf{\Sigma}_W + c\mathbf{\Sigma}_B)^{-1}\mathbf{S}_B] - \ln|\mathbf{S}_B| - p\} \\ & + (N - G)\{\ln|\mathbf{\Sigma}_W| + \mathrm{trace}[\mathbf{\Sigma}_W^{-1}\mathbf{S}_{PW}] - \ln[\mathbf{S}_{PW}] - p\}, \end{aligned} \tag{7}$$

where p is the total number of variables. In the case of balanced data MUML is equivalent to FIML. In the unbalanced case MUML uses less information but has been shown to give very similar results to FIML (Muthén, 1994). The MUML estimator has been shown to yield approximate chi-square distributions and standard errors. It should be noted that $\mathbf{S}_{PW}$ and $\mathbf{S}_B$ can be obtained through standard software or through a special program written by Muthén (see Nelson & Muthén, 1991) which also gives intraclass correlations and an estimate of the constant c, the average group size.[1] This estimation procedure can be implemented in any structural equation modeling software package that allows for multiple-group estimation. For this article, LISREL (Jöreskog & Sörbom, 1991) will be used. A didactic discussion specific to the model described above is given in Kaplan and Elliott (1997).

For the purposes of indicator validation it is required that we obtain a prediction function from (6). Specifically the predicted values of the endogenous indicators for the ith student in the gth school are given by

$$\hat{\mathbf{y}}_{ig} = (\mathbf{I} - \hat{\mathbf{B}}_y)^{-1}\hat{\boldsymbol{\alpha}} + \hat{\mathbf{\Pi}}\hat{\boldsymbol{\tau}}, \tag{8}$$

where $\hat{\mathbf{B}}_y$, $\hat{\mathbf{\Pi}}$, $\hat{\boldsymbol{\alpha}}$, and $\hat{\boldsymbol{\tau}}$ are obtained from fitting the multilevel model. This prediction equation allows one to observe simultaneous changes in all endogenous indicators at both levels of the system as a function of changes in one or more exogenous indicators.

Data Source, Selection of Indicators, and Scaling

As noted earlier, the organizational model shown in Figure 1 was never intended to be a causal model capable of estimation and testing in its totality. Instead, Figure 1 represents the conceptual links among vaguely defined groups of indicators. However, Figure 1 does suggest relationships and directions of influence between subsets of indicators that can be incorporated into a statistical model for purposes of estimation, testing, and prediction. Many of these relationships and directions of influence have been explicitly discussed in, for example, Shavelson, McDonnell, and Oakes (1989). One purpose of this study is to begin to establish statistically appropriate empirical connections between measures of the indicators that represent the relationships implied by the organizational model. It is not necessary that every indicator that has ever been suggested for collection actually appear in the statistical model (though that does not mean that one would not want to collect as many indicators as possible). Indeed, such a model would be neither parsimonious nor useful.

The decision regarding which indicators to include in a statistical model depends partly on the goals of the model. If the goal is solely that of explanation, then one might want to include as many indicators as possible and estimate the innumerable relationships implied by the organizational framework. In this way, one mitigates biases that are known to occur when variables are omitted from a model (see, e.g., Kaplan, 1988, 1989). If the goal, however, is predictive utility, it might be sufficient to incorporate only those indicators that have been suggested by the organizational model as the most influential and policy-relevant predictors of schooling outcomes. In this case omitted variable problems are likely, and, therefore, one would not expect to obtain a statistically well-fitting model. In fact, it has been shown that the incorporation of small and nonsignificant paths in a structural model can degrade the predictive utility of the model when measured by single-sample cross-validation indexes (Kaplan, 1991). Given the tendency of analysts applying structural equation modeling to ignore the potential usefulness of engaging in prediction and simulation (see Elliott & Kaplan, 1995), and given the role that indicator systems can play in informing education policy, we have chosen to give priority to policy relevance, attempting to make clear our rationale for the choice of a particular model. However, as will be seen below, our initial specifications have undergone minor modifications that bring them more in line with the data.

Data Source

Data for this study come from the first follow-up of the National Education Longitudinal Study of 1988 (NELS:88; *NCES*, 1988). NELS:88 is an ongoing study that aims to provide trend data about critical transitions experienced by students as they leave elementary school and progress through high school and beyond. NELS:88 was designed to collect policy-relevant data about educational processes and outcomes, especially as they pertain to student learning, dropping out, and school effects on students' access to programs and equal opportunity to learn. Base-year data were collected in 1988, with planned follow-ups at 2-year intervals.

The subset of students used in this study was obtained as follows. Of the 27,994 students in the total NELS:88 sample, those students who were in 10th grade during the first follow-up were selected. Of those students, only those whose science teachers and school administrators filled out surveys were retained. Next, a subset of variables (described below) was chosen from the student survey, the teacher survey, and the school survey to form science education indicators. Any missing data or multiple responses led to listwise deletion of the subject from the data set. This led to a final student-level sample of $N = 1{,}165$ and a school-level sample of $G = 356$.

Because there were a limited number of students within classrooms, we were unable to estimate a three-level student within-classroom within-school model. However, it was important to include some teacher-level variables in the model. Therefore, a set of teacher-level variables appear in the student-level model. We recognize that because the within-school model contains data at both the student level and the teacher level, our results will be somewhat biased. The potential effects of small within-school sample sizes will be discussed in the conclusion section of this article.

A detailed discussion of the rationale behind the choice of the specific variables used in this study can be found in Kaplan and Elliott (1994) and will not be presented here in the interest of space. Suffice it to say that the choice of the following variables rests on an extensive review of the science education indicators literature—particularly the work of Shavelson, McDonnell, and Oakes (1989) and Murnane and Raizen (1988). We clearly recognize that the adequacy of this model, both in terms of classical notions of statistical fit as well as predictive validity, rests on the choice of variables, their measurement properties, and their location in a set of simultaneous equations. With that in mind, what is described below is not to be construed as assessing the validity of this model per se, but rather to suggest an approach to validating any set of indicators for which empirical and theoretical associations are presumed.

Within-School Indicators

Exogenous indicators. The following variables were used in the within-school model as exogenous variables. TIME was a teacher-level variable measuring the number of minutes per week the science class and lab meet. Three scales related to teachers' goals for the class were developed on the basis of maximum likelihood exploratory factor analysis.[2] These were SURVIVE, measured by responses to statements such as "I am happy just to get through the day"; EMPLOY = teacher focus on employable skills for the students; and TGOAL = teacher's goals for student understanding. In addition, a measure of teacher background was also included as BA = a dummy code for whether a teacher had a bachelor's degree in a science field (1 = yes).

Of the exogenous indicators listed above, TIME and BA are considered policy relevant. The indicator TIME measures the amount of time spent in science class and science lab. Below, we will alter this value and consider how other science education indicators change when the number of hours spent in science instruction is increased. The indicator BA reflects the science training of the teacher. Although this indicator may not be directly manipulable in the same sense that TIME is, it may still be of policy relevance to consider the influence of teacher training in science (as measured by this indicator) on other indicators of science education.

Endogenous indicators. The following variables were used in the within-school model as endogenous variables. Two scales related to teachers' objectives were formed on the basis of maximum likelihood exploratory factor analysis. These were AWARE = promoting awareness of use of science in everyday life and SKILLS = developing science skills. Two scales related to teachers' activities were formed on the basis of maximum likelihood exploratory factor analysis. These were DISCUSS = emphasis on science discussion, and EXPERS = emphasis on experiments.

Student-level endogenous indicators included in the within-school model were also formed on the basis of maximum likelihood exploratory factor analysis. These were PERCEP = student's perception of teacher's emphasis on problem-solving skills, CHALL = how often the student feels challenged in class, and UNDERST = how often the student is asked to show understanding in class. A measure of how well students did in their science classes was also obtained as GRADES = grades in science classes. Finally, the outcome of interest was SCIACH = IRT-estimated number right on the NELS:88 science achievement test. Note that the amount of homework assigned and the time spent on homework were not included in this model. Preliminary investigations revealed that there was virtually no variance in these variables with respect to the subsample used in this analysis.

Between-School Indicators

Exogenous indicators. Seven variables were included in the between-school model as exogenous variables. Three variables were selected as reflecting the school-level resources available for quality science education (see Catterall, 1989). These were LUNCH = percentage of students in the school on free or reduced lunch programs; SALARY = the average difference between the highest and lowest reported teacher salary in the school, recoded to a 1–8 scale; and CLASSGRP = dummy variable response to whether the school used homogeneous ability grouping in its science classes (1 = yes).

Of the exogenous indicators listed above, LUNCH and CLASSGRP are considered of policy relevance. The indicator LUNCH is a proxy for the socioeconomic status of the school. Below, we will set this indicator to zero, to reflect only schools where there are no children on free or reduced lunch—that is, schools of upper socioeconomic status (including private schools)—and examine values of science education indicators for these schools. The indicator CLASSGRP is a dummy variable reflecting the school's policy on class grouping. Manipulation of this variable allows us to examine the role of class grouping on values of other indicators of science education.

Endogenous indicators. The resource indicators were followed by three scales derived from a maximum likelihood exploratory factor analysis reflecting endogenous indicators of the professional teaching conditions of the school. These were STAFCLIM = principals' response to the extent to which the staff explores new ideas, cooperate amongst themselves, help out with additional duties, and share beliefs about the mission of the school; ACAEMPH = principals' response to the extent to which the school publicizes academic achievement and monitors student academic progress; and CONFLICT = principals' response to the extent to which teachers have a negative attitude toward students, cannot motivate students, and have conflict with administrators.

The final school-level endogenous indicator was ACAPRESS = academic press of the school, measured by the principals' responses to items asking if teachers at the school (a) press students to achieve academically, (b) expect students to do homework, and (c) encourage students to enroll in academic classes.

Results

In what follows, the separate initial and final specifications of the within-school model will be presented first. This will be followed by the intraclass correlations. The separate initial and final specification of the between-school model will be presented. Finally, the two separate models will be combined into the multilevel model. Models will be modified by using a combination of modification indexes and expected parameter change statistics (Saris, Satorra, & Sörbom, 1987; Sörbom, 1989). This method of model modification has been suggested by Kaplan (1989, 1990a, 1990b) as a way of guiding specification error searches taking into account statistical power and substantive considerations.

Assumptions underlying the use of structural equation modeling were also examined. In particular, the variables used for the within- and between-school model did not deviate in any important way from the assumption of multivariate normality, and thus it was felt that maximum likelihood estimation was robust in this case (Muthén & Kaplan, 1985, 1992). Listwise deletion was used in this study despite the fact that this method is known to present problems for tests of goodness of fit unless the missing data are missing completely at random (see, e.g., Muthén, Kaplan, & Hollis, 1987). This article does not attempt to explicitly model the problem of missing data.

Results of Within-School Model

Figure 2 presents the path diagram of the initial within-school model. This initial specification was found not to fit the data ($\chi^2(57, N = 1{,}165) = 214.302$, $p < .05$, $GFI = 0.975$, $AGFI = 0.954$, $RMSEA = 0.049$, $p = .599$).[3] A series of modifications were made on the basis of the modification index and the expected parameter change statistic. These modifications included freeing paths from TIME to EXPERS, EMPLOY to DISCUSS, SKILLS to AWARE, CHALL to SCIACH, and CHALL to GRADES. Once these paths were free, nonsignificant paths were removed if they maintained the integrity of the initial model. These removals included the path from SURVIVE to AWARE and from TIME to AWARE. The final model, which is shown in Figure 3, was found to adequately fit the data ($\chi^2(54,$ $N = 1{,}165) = 69.949$, $p = .071$, $GFI = 0.991$, $AGFI = 0.983$, $RMSEA = .016$, $p =$

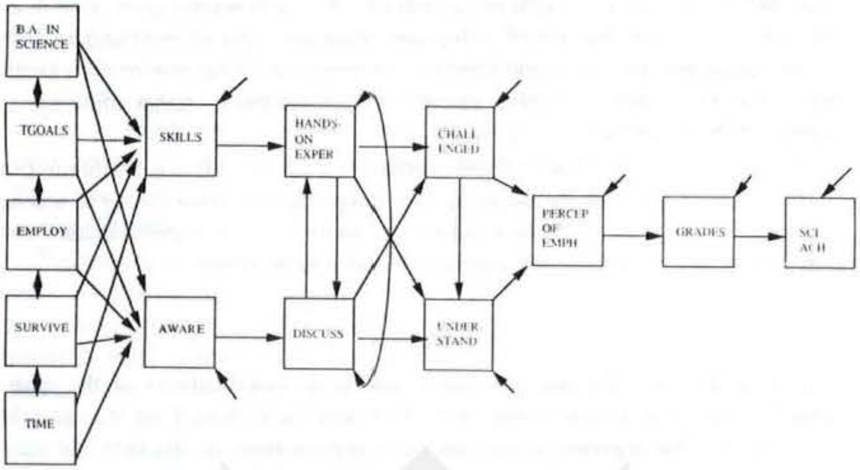

FIGURE 2. *Path diagram of initial within-school model*

1.0). The standardized direct, indirect, and total effects are presented in Table 1. The system of equations describing the final within-school model are given in the Appendix.

Between-School Variation: The Intraclass Correlations

As suggested by Muthén (1994) the decision to proceed with a multilevel analysis depends, in part, on the extent to which there are substantively large intraclass correlations among the within-school variables. Intraclass correlations provide a measure of the degree of between-school variation in the within-school variables and, in the multilevel structural equation modeling case, are

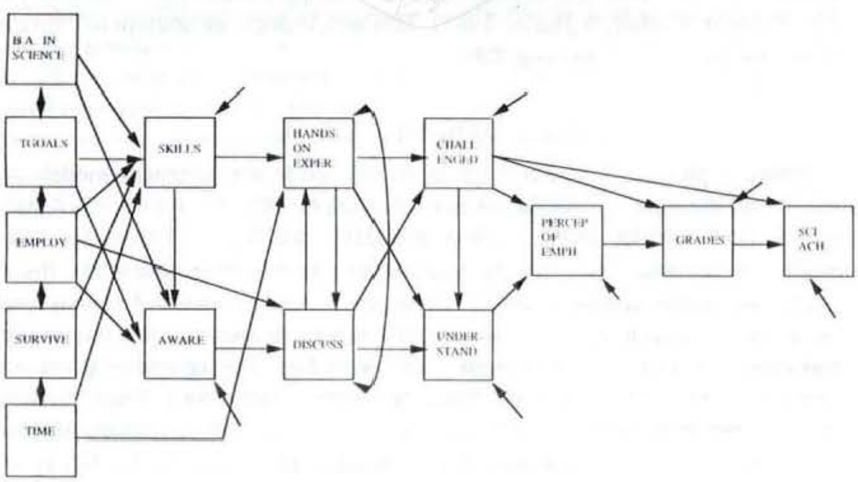

FIGURE 3. *Path diagram of final within-school model*

TABLE 1

Direct, indirect, and total effects of within-school model: Standardized solution

	GRADES	PERCEP	UNDERST	CHALL	EXPERS	DISCUSS	SKILLS	AWARE	BA	TGOAL	EMPLOY	SURVIVE	TIME
SCIACH	0.370*		0.123*										
		0.065*	0.061*	0.009*	−0.003	−0.007	−0.002	−0.002	0.000	0.000	−0.002	0.000	−0.001
	0.370*	0.065*	0.184*	0.009*	−0.003	−0.007	−0.002	−0.002	0.000	0.000	−0.002	0.000	−0.001
GRADES		0.175*	0.099*										
			0.066*	0.024*	−0.003	−0.006	−0.002	−0.001	0.000	0.000	−0.001	0.000	−0.001
		0.175*	0.165*	0.024*	−0.003	−0.006	−0.002	−0.001	0.000	0.000	−0.001	0.000	−0.001
PERCEP			0.304*	0.135*									
			0.071*		−0.009	−0.010	−0.005	−0.003	−0.000	−0.000	−0.003	0.000	−0.002
			0.375*	0.135*	−0.009	−0.010	−0.005	−0.003	−0.000	−0.000	−0.003	0.000	−0.002
UNDERST					−0.005	−0.042							
					−0.009	0.003	−0.009	−0.010	0.000	0.000	−0.009	0.001	−0.003
					−0.014	−0.039	−0.009	−0.010	0.000	0.000	−0.009	0.001	−0.003
CHALL			0.527*		−0.036	0.029							
					0.001	−0.014	−0.016	0.004	−0.001	−0.001	−0.001	0.001	−0.007
			0.527		−0.035	−0.015	−0.016	0.004	−0.001	−0.001	−0.001	0.001	−0.007
EXPERS						−0.238*	0.465*						0.169*
					−0.051*	0.012	−0.037*	−0.059*	0.033*	0.025	0.056*	−0.029*	0.022
					−0.051*	−0.226*	0.427*	−0.059*	0.033*	0.025	0.056*	−0.029*	0.191*
DISCUSS					0.223*			0.263*			0.105*		
					−0.011	−0.051*	0.155*	−0.013*	−0.009	−0.011	0.109*	−0.010*	0.047*
					0.212*	−0.051*	0.155*	0.250*	−0.009	−0.011	0.215*	−0.010*	0.047*
SKILLS									0.065*	0.049	0.230*	−0.067*	0.071*
									0.065*	0.049	0.230*	−0.067*	0.071*
AWARE							0.228*		−0.078*	−0.076*	0.317*		
									0.015*	0.011	0.052*	−0.015*	0.016*
							0.228*		−0.064*	−0.065*	0.369*	−0.015*	0.016*

Note. For each variable in the first column, the first row of values represents direct effects, the second row of values represents indirect effects, and the third row of values represents total effects.

*$p < .05$.

calculated as the ratio of the between-group variances to the sum of the between- and within-group variances (see Muthén, 1991). The combination of the within-school data and between-school data resulted in a slight loss of data, bringing the within-school sample size to 1,069 and the between-school sample size to 321. For these new data, we found that the intraclass correlations for the variables of the within-school model varied between 8% and 64%. Thus, it was felt that the intraclass correlations were large enough to warrant a multilevel analysis.

Results of Between-School Model

Before formally combining the within- and between-school models in the multilevel analysis, it was felt that the between-school model should be specified, estimated, and, if need be, modified separately. Figure 4 shows the path diagram of the initial between-school model. This model was found not to fit the data ($\chi^2(6, N = 365) = 74.20$, $p = .000$, $GFI = 0.946$, $AGFI = 0.748$, $RMSEA = 0.180$, $p < .001$). Two modifications were made to the initial model. These included freeing the paths from STAFCLIM to ACAEMPH and from CLASSGRP to ACAEMPH. These modifications resulted in a much improved fit ($\chi^2(4, N = 365) = 11.71$, $p = .024$, $GFI = 0.991$, $AGFI = 0.935$, $RMSEA = 0.074$, $p = .171$). No other modifications were warranted on the basis of the modification index or the expected parameter change. No parameters were fixed on the basis of lack of statistical significance. The final between-school model is shown in Figure 5, and the standardized direct, indirect, and total effects are displayed in Table 2. The system of equations describing the final between-school model are shown in the Appendix.

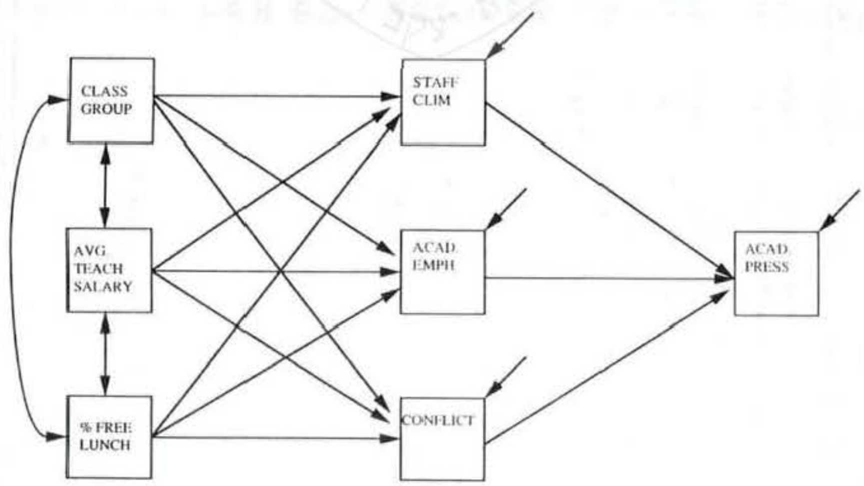

FIGURE 4. *Initial between-school model*

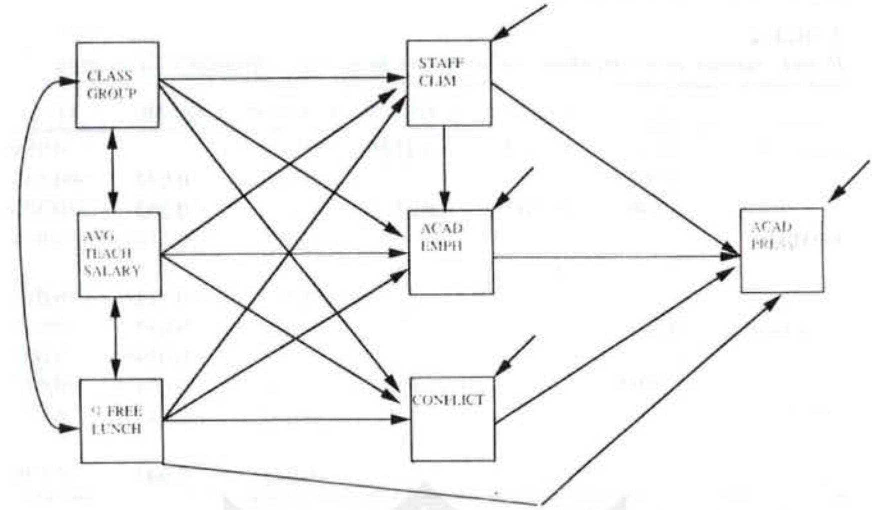

FIGURE 5. *Final between-school model*

Results of Multilevel Analysis

The final stage involved combining the two models into a single multilevel model. Again, we are assuming that the intercepts and means of the within-school model vary across schools and that a separate school model holds which can account for the intercept variation. The multilevel model combines the two separate final models and allows paths to be estimated between the school-level variables and the intercepts of the within-school equations. For this model, we estimated the paths between ACAPRESS and the intercepts and means of the within-school model, with the exception of BA and TIME. This was done because we felt that BA and TIME were truly exogenous variables at the within-school level and outcomes of between-school variables chosen for this model. It should be noted that the intercepts in this model are interpreted as the expected values of the student-level variables given that the predictors in their respective equations are zero. Since the value of zero may not be sensible for many of these variables, the intercepts are not directly meaningful. Therefore, the coefficients relating the intercepts to the school-level variables may not be interpretable.

The initial model was estimated in LISREL (Jöreskog & Sörbom, 1991). Note that the sample size for the within-school model is $N - G = 809$, compared to $N = 1{,}165$ when the within-school model is analyzed separately. The initial model was found to fit the data as evidenced by the likelihood ratio chi-square (χ^2(383, $N_B = 356$, $N_W = 809$) = 381.413, $p = .513$, $GFI = .987$, $RMSEA = 0.00$, $p = 1.0$). Given the fact that the overall model shows adequate fit, and given the lack of substantive justification for removing these across-level paths, it was decided not to simplify the model. The standardized regression coefficients of the within-school intercepts on ACAPRESS are shown in Table 3. In addition to many paths being small and not statistically significant, many have signs that do not

TABLE 2
Direct, indirect, and total effects of between-school model: Standardized solution

	STAFCLIM	ACAEMPH	CONFLICT	CLASSGRP	SALARY	LUNCH
ACAPRESS	−0.014	0.341*	−0.212*			−0.258*
	0.103			0.011	0.032	−0.042
	0.088	0.341*	−0.212*	0.011	0.032	−0.300*
STAFCLIM				0.071	0.114*	0.014
				0.071	0.114*	0.014
ACAEMPH	0.301*			0.025	0.037	−0.017
				0.021	0.034*	0.004
	0.301*			0.046	0.071	−0.013
CONFLICT				0.017	−0.043	0.176*
				0.017	−0.043	0.176*

Note. For each variable in the first column, the first row of values represents direct effects, the second row of values represents indirect effects, and the third row of values represents total effects.
*$p < .05$.

make good theoretical sense. We speculate as to the reasons for this finding in the Summary and Conclusions section.

Indicator Validation via Policy-Relevant Simulations

In this section we explore the utility of the final multilevel model for indicator validation. Specifically, utilizing Equation 8, which gives the prediction of the endogenous variables from the full multilevel specification, we simulate various policy-relevant situations and study their effects on the NELS:88 science achievement test scores. Simulations were accomplished by reading appropriate matrices from the output of LISREL into the SAS Interactive Matrix Language (PROC IML; SAS Institute Inc., 1989).

We first consider simulations involving a change in one indicator at a time. This is then followed by the simultaneous change in a number of indicators at both levels of the system. Two points should be noted when considering the following results. First, as discussed above, the model contains certain specification errors, as can be seen in some of the theoretically incorrect signs in Table 3. Second, it is important to reiterate that these simulations are for demonstrating the utility of a model-based approach to indicator validation and should not be read as suggesting serious policy alternatives. Considerably more methodological and substantive work needs to occur before this (or any) model based on this methodology is used for policy analysis. Given the noted caveats, the results presented here should be interpreted as predicted values of a hypothetical student, teacher, or school possessing particular within-school and between-school characteristics.

TABLE 3
Regression of within-school intercepts on ACAPRESS*: Standardized solution*

	ACAPRESS
SCIACH	0.118
GRADES	0.261
PERCEP	−0.026
UNDERST	−0.661
CHALL	−0.239
EXPERS	−0.033
DISCUSS	0.093
SKILLS	−0.019
AWARE	0.101
BA[a]	
TGOAL	0.101
EMPLOY	−0.023
SURVIVE	0.150*
TIME[a]	

Note. [a]Variables treated as exogenous in the within-school model.
*$p < .05$.

Simulation 0: Baseline Predictions

The heading *Baseline* in Table 4 gives the fitted values of the *i*th student in the *g*th school at the estimated point. These are considered the baseline values and should be used for comparative purposes. It should be noted that all predicted values are within the admissible ranges of their scales. This is an important piece of information regarding the quality of a model for indicator validation. Had the baseline values been predicted to lie outside admissible ranges then this particular model would have needed to be reexamined and perhaps further modified to bring predicted values in line with known scale properties. The ability to check on the reasonableness of the initial predicted values is one advantage of the approach advocated in this article and confers a degree of confidence in future predictions derived from this model.

Simulation 1: Homogeneous Versus Heterogeneous Class Grouping

This simulation yields a predicted achievement score for a student from a school that uses homogeneous grouping on the basis of science ability with the effects of all other indicators held constant. We contrast this with a predicted score of a student from a school with heterogeneous class grouping. Simulations 1a and 1b of Table 4 show the results of this simulation. The results of this simulation show that slightly higher levels of academic press are predicted for schools where homogeneous rather than heterogeneous class grouping is used. This appears to result in slightly higher levels of teacher-reported activities such as emphasis on experiments and so on, but the effect on achievement is not noticeably different when compared to heterogeneous class grouping.

TABLE 4

Predicted indicator scores based on policy simulations from the multilevel model

Indicator	Baseline	Sim 1a	Sim 1b	Sim 2a	Sim 2b	Sim 3	Sim 4	Sim 5
SCIACH	13.161	13.161	13.159	13.160	13.161	13.157	13.228	13.225
GRADES	2.023	2.023	2.022	2.023	2.023	2.022	2.034	2.034
PERCEP	7.741	7.740	7.741	7.740	7.743	7.731	7.714	7.029
CHALL	2.000	1.999	2.000	2.000	2.000	2.000	1.983	1.978
UNDERST	3.549	3.549	3.550	3.547	3.557	3.532	3.526	3.506
EXPERS	8.765	8.765	8.766	8.827	8.588	9.355	8.747	9.399
DISCUSS	18.177	18.177	18.176	18.208	18.087	18.565	18.224	18.644
SKILLS	10.237	10.237	10.237	10.377	9.833	10.697	10.227	10.828
AWARE	2.749	2.749	2.749	2.747	2.755	2.778	2.755	2.784
BA	0.741	0.741	0.741	1[a]	0[a]	0.741	0.741	1[a]
TGOALS	3.601	3.601	3.600	3.601	3.600	3.601	3.617	3.617
EMPLOY	1.848	1.848	1.848	1.848	1.847	1.848	1.845	1.845
SURVIVE	0.866	0.866	0.866	0.866	0.866	0.866	0.887	0.887
TIME	4.712	4.712	4.712	4.712	4.712	6.696[a]	4.712	6.709[a]
ACAPRESS	7.099	7.104	7.088	7.099	7.099	7.099	7.549	7.554
STAFCLIM	4.572	4.620	4.478	4.572	4.572	4.572	4.554	4.602
ACAEMPH	7.426	7.456	7.365	7.426	7.426	7.426	7.426	7.456
CONFLICT	3.468	3.489	3.426	3.468	3.468	3.468	3.066	3.087
CLASSGRP	0.667	1[a]	0[a]	0.667	0.667	0.667	0.667	0[a]
SALARY	3.751	3.751	3.751	3.751	3.751	3.751	3.751	3.751
LUNCH	1.408	1.408	1.408	1.408	1.408	1.408	0[a]	1.408

Note. [a]Manipulated policy instruments; see text.

Simulation 2: Teachers With or Without an Undergraduate Degree in Science

This simulation examines the predicted effect of teacher educational background on the endogenous indicators of the model. Specifically we study whether a teacher having a bachelor's degree in a science discipline exerts any marginal effect on predictors of science achievement and on science achievement directly with the effects of all other indicators held constant. The results of these simulations are given in Simulations 2a and 2b of Table 4. The results show that having a bachelor's degree versus not having a bachelor's degree does not change the predicted value of science achievement or its determinants within this model.

Simulation 3: Increase in Time Spent in Science Instruction

This simulation examines the effect of increasing the total time the science class meets (lecture plus lab) by 2 hours. The results are shown in Simulation 3 of Table 4. Here it can be seen that an increase of 2 hours leads to small but noticeable changes in the predicted values of the indicators relative to the baseline results of Simulation 0. The changes appear to be most noticeable with respect to teacher-reported activities of emphasizing experiments and discussion and teacher-reported objectives of increasing skills and awareness. The predicted values are above baseline, as one might expect. Nevertheless, an increase of 2 hours has virtually no effect on science achievement as measured in NELS:88.

Simulation 4: Schools With Different Percentages of Students on Free or Reduced Lunch

This simulation studies the predicted values of indicators for the *i*th student in the *g*th school where there are no students on free or reduced lunch, holding all other effects constant. The results are presented in Simulation 4 of Table 4. It can be seen that schools with zero students on free or reduced lunch report slightly lower predicted conflict and slightly higher academic press in the school. However, this simulation shows very little effect on within-school indicators, especially science achievement.

Simulation 5: Simultaneous Change in Three Variables

This simulation studies the simultaneous effect of changes in three indicators: (a) heterogeneous class grouping, (b) teachers with bachelor's degrees in a science discipline, and (c) an increase of 2 hours in the total time a science class meets. All other effects are held constant at their estimated values. The results are shown in Simulation 5 of Table 4. Here it can be seen that the combined effect of these changes is predicted increases in many, but not all, of the within-school indicators. In particular, a small predicted increase in science achievement is observed. A predicted increase in teacher-reported emphasis on science experiments, science discussion, and science awareness is also observed under this scenario.

Summary and Conclusions

The purpose of this article was to demonstrate the utility of a model-based approach to education indicator validation. To illustrate the approach, a multilevel model of science achievement based on a specification arising from the indicator systems literature was applied to the first follow-up of NELS:88. Scales reflecting indicators that were suggested to be important for monitoring the health of science education within and between schools were incorporated into the model. The model was found to have adequate fit to the data when the levels of the within-school indicators were allowed to be predicted by all between-school indicators. It should be pointed out that the observed effects in this study were quite small overall, owing partly to the fact that the NELS:88 test has too few items to be sensitive to school or classroom instructional indicators (see also Carey & Shavelson, 1989). Another reason for the small observed changes lies in the fact that the effects of the between-school variables on the within-school intercepts were very small. Thus, despite the fact that the intraclass correlations suggested the appropriateness of a multilevel model, the actual cross-level effects appeared negligible.

A number of methodological issues must be considered when interpreting the results of this study. First, as noted above, the signs of many of the coefficients reported in Table 3 do not make theoretical sense. The most probable explanation for this finding has to do with specification errors in the cross-level part of the model. In this example, all cross-level effects were from the between-school measure of academic press to the intercepts of the within-school variables. A major advantage to the modeling methodology advocated in this article (compared to standard multilevel regression) is that it allows effects in the opposite direction to be specified and estimated. That is, one can examine the influence of average student-level grades, say, on the academic press of the school. In addition, simultaneous cross-level effects can also be specified. However, with respect to this model, inspection of the modification indexes and expected parameter change statistics did not reveal the necessity of estimating those effects. Second, the unequal within-school sample sizes may also be contributing to the finding of small and nonsignificant cross-level effects. The influence of unequal sample sizes on cross-level effect estimation in multilevel structural models remains an open area of research. Third, the present investigation does not provide confidence intervals around the predicted values. The incorporation of prediction intervals is essential for a proper assessment of the precision of the estimates, especially if the model is to be used for indicator validation. With regard to the model used here, simultaneous prediction intervals from the reduced form of a multilevel model would be required. This also remains an open area of research.

Some substantive concerns must also be considered. Problems with the multilevel estimates may be due to the fact that this model of academic press is not the correct explanatory model for variation in the within-school indicators. That is, while the within-school and between-school models showed adequate statistical fit to their respective data, there may be no necessary relationship between the two. Again, the purpose of this article was not to argue for this model per se, but rather to argue for a model-based approach to indicator validation. An important benefit that can accrue from this modeling exercise, however, is the insight it can provide for theory building. Assuming that the methodological problems listed above were addressed, and a reanalysis led to similar findings, one might be compelled to reexamine the role of the academic press of the school as it pertains to teacher behavior and student outcomes. In addition to valuable information for indicator development, such an investigation would also lead to the progressive elaboration of a theory of schooling.

These methodological and substantive concerns notwithstanding, a unique feature of this article was the explicit attempt at estimating empirical associations among indicators within and between levels of the educational system. Rather than simply collecting indicators guided by the organizational framework and using them for reporting descriptive statistics, this article successfully demonstrates that it is possible to employ multilevel structural modeling methods to capture the salient relationships among science education indicators based on an organizational model of schooling wherein such relationships are implied. The type of modeling activity advocated in this article, we argue, is consistent with de Neufville's (1978) notions of the theoretical validity of indicators. Specifically, the simulation approach advanced in this article could be considered a type of quasi-experimental validity test (de Neufville, 1978, p. 178) for the selection of indicators, wherein one can observe how indicators change in response to changes in other indicators embedded within an explicit model of the educational system. Our results, though quite tentative, demonstrate the potential utility of a model-based approach to indicator validation and suggest that indicator data can be collected and organized into a quantitative model that can be used, in conjunction with other modes of inquiry, to inform policy directed toward indicator development.

Notes

[1] This software program is also downloadable from Muthén's World Wide Web page. It can be accessed at http://www.gse.ucla.edu/facpage/muthen.html/.

[2] In each case a variety of alternative factor structures with oblique rotation were explored. The criterion for choosing the number of factors was based on the change in the chi-square goodness of fit when the number of factors changed, as well as substantive significance and interpretability of the factor loadings. Scales were formed by unit weighing variables that loaded appreciably on the factors. These scales define the indicators under investigation.

[3] The statistic χ^2 refers to the likelihood ratio chi-square. The *GFI*, *AGFI*, and *RMSEA* are alternative fit indexes and stand for goodness-of-fit index, adjusted goodness-of-fit index, and root mean square error of approximation, respectively (see Bollen, 1989; Bollen & Long, 1993, for details).

APPENDIX

Final Within-School Model Equations

$$\text{SCIACH} = \alpha_1 + \beta_{y_{12}} (\text{GRADES}) + \beta_{y_{15}} (\text{UNDERST}) + \epsilon_1, \tag{A1}$$

$$\text{GRADES} = \alpha_2 + \beta_{y_{23}} (\text{PERCEP}) + \beta_{y_{25}} (\text{UNDERST}) + \epsilon_2, \tag{A2}$$

$$\text{PERCEP} = \alpha_3 + \beta_{y_{34}} (\text{CHALL}) + \beta_{y_{35}} (\text{UNDERST}) + \epsilon_3, \tag{A3}$$

$$\text{CHALL} = \alpha_4 + \beta_{y_{45}} (\text{UNDERST}) + \beta_{y_{46}} (\text{EXPER}) + \beta_{y_{47}} (\text{DISCUSS}) + \epsilon_4, \tag{A4}$$

$$\text{UNDERST} = \alpha_5 + \beta_{y_{56}} (\text{EXPER}) + \beta_{y_{57}} (\text{DISCUSS}) + \epsilon_5, \tag{A5}$$

$$\text{EXPER} = \alpha_6 + \beta_{y_{67}} (\text{DISCUSS}) + \beta_{y_{68}} (\text{SKILLS}) + \beta_{y_{6.14}} (\text{TIME}) + \epsilon_6, \tag{A6}$$

$$\text{DISCUSS} = \alpha_7 + \beta_{y_{76}} (\text{EXPER}) + \beta_{y_{79}} (\text{AWARE}) + \beta_{y_{7.12}} (\text{EMPLOY}) + \epsilon_7, \tag{A7}$$

$$\text{SKILLS} = \alpha_8 + \beta_{y_{8.10}} (\text{BA}) + \beta_{y_{8.12}} (\text{EMPLOY}) + \beta_{y_{8.13}} (\text{SURVIVE}) + \beta_{y_{8.14}} (\text{TIME}) + \epsilon_8, \tag{A8}$$

$$\text{AWARE} = \alpha_9 + \beta_{y_{9.11}} (\text{TGOALS}) + \beta_{y_{9.12}} (\text{EMPLOY}) + \beta_{y_{9.14}} (\text{TIME}) + \epsilon_9, \tag{A9}$$

$$\left.\begin{array}{c}\text{BA}\\ \text{TGOALS}\\ \text{EMPLOY}\\ \text{SURVIVE}\\ \text{TIME}\end{array}\right\}\ \text{Exogenous}$$

Final Between-School Model Equations

$$\begin{aligned}\text{ACAPRESS} = {}& \tau_1 + \beta_{z_{12}} (\text{STAFFDEV}) + \beta_{z_{13}} (\text{ACAEMPH}) \\ & + \beta_{z_{14}} (\text{CONFLICT}) + \beta_{z_{17}} (\text{LUNCH}) + u_1,\end{aligned} \tag{A10}$$

$$\text{STAFFDEV} = \tau_2 + \beta_{z_{25}} (\text{CLASSGRP}) + \beta_{z_{26}} (\text{SALARY}) + \beta_{z_{27}} (\text{LUNCH}) + u_2, \tag{A11}$$

$$\begin{aligned}\text{ACAEMPH} = {}& \tau_3 + \beta_{z_{32}} (\text{STAFFDEV}) + \beta_{z_{13}} (\text{ACAEMPH}) + \beta_{z_{35}} (\text{CLASSGRP}) \\ & + \beta_{z_{36}} (\text{SALARY}) + \beta_{z_{37}} (\text{LUNCH}) + u_3,\end{aligned} \tag{A12}$$

$$\text{CONFLICT} = \tau_4 + \beta_{z_{45}} (\text{CLASSGRP}) + \beta_{z_{46}} (\text{SALARY}) + \beta_{z_{47}} (\text{LUNCH}) + u_4, \tag{A13}$$

$$\left.\begin{array}{c}\text{CLASSGRP}\\ \text{SALARY}\\ \text{LUNCH}\end{array}\right\}\ \text{Exogenous}$$

References

Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies (with discussion). *Journal of the Royal Statistical Society, Series A, 149*, 1–43.

Bidwell, C. E., & Kasarda, J. D. (1975). School district organization and student achievement. *American Sociological Review, 40*, 55–70.

Boardman, A. E., Davis, O. A., & Sanday, P. R. (1977). A simultaneous equations model of the educational process. *Journal of Public Economics, 7*, 23–49.

Bock, R. D. (1989). *Multilevel analysis of educational data*. San Diego, CA: Academic Press.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. Newbury Park, CA: Sage.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Carey, N., & Shavelson, R. J. (1989). Outcomes, achievement, participation, and attitudes. In R. J. Shavelson, L. M. McDonnell, & J. Oakes (Eds.), *Indicators for monitoring mathematics and science education: A sourcebook* (pp. 147–191). Santa Monica, CA: RAND Corporation.

Catterall, J. S. (1989). Resources and commitment. In R. J. Shavelson, L. M. McDonnell, & J. Oakes (Eds.), *Indicators for monitoring mathematics and science education: A sourcebook* (pp. 25–39). Santa Monica, CA: RAND Corporation.

Cohn, E., & Millman, S. D. (1975). *Input-output analysis in public education*. Cambridge, MA: Ballinger.

Darling-Hammond, L. (1992). Educational indicators and enlightened policy. *Educational Policy*, 6, 235–265.

de Neufville, J. I. (1978). Validating policy indicators. *Policy Sciences, 10*, 171–188.

Elliott, P. R., & Kaplan, D. (1995). *Beyond goodness-of-fit: A perspective on the practice and potential of structural equation modeling*. Manuscript submitted for publication.

Glasman, N. S., & Biniaminov, I. (1981). Input-output analysis of schools. *Review of Educational Research, 51*, 509–539.

Goldberger, A. S., & Duncan, O. D. (1973). *Structural equation models in the social sciences*. New York: Academic Press.

Jöreskog, K. G. (1977). Structural equation modeling in the social sciences: Specification, estimation, and testing. In P. R. Krishnaiah (Ed.), *Applications of statistics* (pp. 265–287). Amsterdam: North-Holland.

Jöreskog, K. G., & Sörbom, D. (1991). *LISREL 7: Analysis of linear structural relationships by the method of maximum likelihood*. Chicago: Scientific Software, Inc.

Kaplan, D. (1988). The impact of specification error on the estimation, testing, and improvement of structural equation models. *Multivariate Behavioral Research, 23*, 69–86.

Kaplan, D. (1989). A study of the sampling variability and z-values of parameter estimates from misspecified structural equation models. *Multivariate Behavioral Research, 24*, 41–57.

Kaplan, D. (1990a). Evaluation and modification of covariance structure models: A review and recommendation. *Multivariate Behavioral Research, 25*, 137–155.

Kaplan, D. (1990b). Rejoinder on evaluating and modifying covariance structure models. *Multivariate Behavioral Research, 25*, 197–204.

Kaplan, D. (1991). On the modification and predictive validity of covariance structure models. *Quality & Quantity, 25*, 307–314.

Kaplan, D., & Elliott, P. R. (1994). *A multilevel structural model of science achievement from an indicator system perspective: Implications for educational policy analysis.* Final report to the AERA Grants Program Committee.

Kaplan, D., & Elliott, P. R. (1997). A didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling, 4*, 1–24.

Levin, H. M. (1970). A new model of school effectiveness. In A. Mood (Ed.), *Do teachers make a difference?* Washington, DC: U.S. Department of Health, Education, and Welfare.

McDonnell, L. M. (1989). The policy context. In R. J. Shavelson, L. M. McDonnell, & J. Oakes (Eds.), *Indicators for monitoring mathematics and science education: A sourcebook* (pp. 241–269). Santa Monica, CA: RAND Corporation.

Michelson, S. (1970). The association of teacher resources with children's characteristics. In A. Mood (Ed.), *Do teachers make a difference?* Washington, DC: U.S. Department of Health, Education, and Welfare.

Murnane, R. J., & Raizen, S. A. (1988). *Improving indicators of the quality of science and mathematics education in Grades K–12*. Washington, DC: National Academy Press.

Muthén, B. (1989). Latent variable modeling in heterogenous populations. *Psychometrika, 54*, 557–585.

Muthén, B. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement, 28*, 338–354.

Muthén, B. (1994). Multilevel covariance structure analysis. *Sociolgical Methods & Research, 22*, 376–398.

Muthén, B., & Kaplan, D. (1985). A comparision of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38*, 171–189.

Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology, 45*, 19–30.

Muthén, B., Kaplan, D., & Hollis, M. (1987). Structural equation modeling for data that are not missing completely at random. *Psychometrika, 51*, 431–462.

Muthén, B., & Satorra, A. (1989). Multilevel aspects of varying parameters in structural models. In R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 87–99). San Diego, CA: Academic Press.

National Center for Education Statistics. (1988). *National education longitudinal study of 1988*. Washington, DC: Author.

National Educational Goals Panel. (1994). *National educational goals report: Building a nation of leaders: 1994*. Washington, DC: Author.

Nelson, G., & Muthén, B. (1991). *Analysis preparation steps for multilevel analysis using LISCOMP* (Tech. rep.). Los Angeles: University of California, Los Angeles.

Oakes, J. (1986). *Educational indicators: A guide for policymakers*. New Brunswick, NJ: Center for Policy Research in Education, Rutgers University.

Porter, A. C. (1991). Creating a system of school process indicators. *Educational Evaluation and Policy Analysis, 13*, 13–29.

Raudenbush, S. W., & Willms, J. D. (1991). *Schools, classrooms, and pupils: International studies of schooling from a multilevel perspective*. New York: Academic Press.

Saris, W. E., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. In C. C. Clogg (Ed.), *Sociological methodology 1987* (pp. 105–129). San Francisco: Jossey-Bass.

SAS Institute Inc. (1989). *SAS interactive matrix language*. Cary, NC: Author.

Shavelson, R. J., McDonnell, L. M., & Oakes, J. (1989). *Indicators for monitoring mathematics and science education: A sourcebook*. Santa Monica, CA: RAND Corporation.

Sörbom, D. (1989). Model modification. *Psychometrika, 54*, 371–384.

Willms, J. D. (1992). *Monitoring school performance: A guide for educators*. Washington, DC: The Falmer Press.

Authors

DAVID KAPLAN is Associate Professor of Educational Studies, Department of Educational Studies, University of Delaware, Newark, DE 19716; dkaplan@udel.edu. He specializes in educational statistics and psychometrics.

PAMELA R. ELLIOTT is Research Associate in the Center for Community Study and Assistant Professor of Psychology, University of Rochester, Rochester, NY 14627; pelliott@prodigal.psych.rochester.edu. She specializes in educational statistics and psychometrics.

Received October 16, 1995
Revision received February 26, 1996
Accepted May 8, 1996