

## **Validation of Competing Statistical Formulations of School Effectiveness**

Eugene Kennedy, Ph.D.  
Louisiana State University

Maryann Durland, Ph.D.  
Kentucky Institute of Educational Research

April 1999

---

Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada. Communications regarding this manuscript should be directed to Dr. Eugene Kennedy, 111 Peabody Hall, Louisiana State University, LA 70803

## **Abstract**

This study examined the relationship between school effectiveness indexes generated with three competing statistical models and measures of school climate and effective instructional practices. The objective of the study was to determine which of the three models examined yielded the strongest correlations with the climate and instructional quality indicators examined. The three statistical models considered were the (a) student-level regression model, (b) school-level aggregate regression model, and (c) multilevel model. Indexes of effectiveness were generated with each technique and then correlated with measures of school climate perceptions of teachers, administrators, students, and parents. These indexes were also correlated with measures of time on task and the Louisiana components of effective teaching. The results were interpreted cautiously due to the size of the sample. However, the tendency in the data was for the three procedures to yield similar correlations both in terms of direction and magnitude.

## Introduction

Over the past two decades there has been considerable attention focused on the identification of schools that appear to be unusually effective or ineffective at fostering desired student outcomes such as achievement, attendance, etc. (Raudenbush & Bryk, 1986). One of the persistent dilemmas facing researchers in this area has been the question of how to estimate the magnitude of a school effect. A variety of approaches have been considered in the past. These include (a) student-level regression models, (b) school aggregate regression models, and, more recently (c) multilevel models.

Multilevel models attempt to incorporate both student- and school-level variables in the modeling process and, in theory, more realistically mirror the processes operating in schools. Because of this, many researchers have enthusiastically embraced these models and considerable research has been directed toward their development. Specifically, these models have been expanded and modified to ever increasing levels of complexity in an effort to more realistically address the organization of modern schools. These developments include longitudinal models which permit study of cohort and panel designs (Raudenbush, 1989), cross-classified models which allow for the possibility that lower level units may be nested within *several* higher level units (Goldstein, 1994), and multivariate and structural equation models which permit modeling of multiple outcomes and allow for the presence of measurement errors (Muthen, 1994).

The application of sophisticated multilevel models to the identification of effective and ineffective schools has been discussed at length in the school effects literature. While most investigators would agree that these models offer the possibility of more realistic modeling of school processes, their sophistication may be an impediment to use in practical settings.

Specifically, it seems unlikely that even the most simple versions of these models will be easily comprehended by lay persons. Add to this the shrinkage phenomenon associated parameter estimates, and these models also have questionable ‘fairness’ appeal for school accountability purposes. Given the politically sensitive nature of high-stakes accountability systems, this is a serious limitation.

The ease with which a technique can be communicated should not be the sole criteria by which a statistical model is selected for accountability purposes. However, there is evidence that estimates of school effects based on multilevel models may not yield school rankings which differ significantly from more easily understood regression based procedures (see Fitz-Gibbon, 1996). In the present effort, we compare student-level, school aggregate and multilevel procedures for estimating school effects in terms of their relationships with ‘other’ indicators of school effectiveness. Specifically, we compare school effectiveness indexes based on regression and multilevel, models with school level summaries of (a) time on task, (b) school climate, and (c) instructional practices. It is our belief that this strategy permits a more meaningful selection of one statistical procedure over another—the logic being that the most defensible estimate of a school effect is the one that yields the strongest relationship with know indicators of school effectiveness.

## **Background**

### Statistical Models of School Effects

Raudenbush and Willms (1995) present a structural model of student achievement which serves as a useful platform for a discussion of the complexities of defining and estimating school effects. Accordingly,

$$\text{Student Achievement} = \text{Mean Achievement} + \text{Student Background} + \\ \text{School Context} + \text{School Process} + \text{Random Error}$$

According to this model, student achievement is a function of student background, (e.g., social class, previous achievement, etc.), school processes (e.g., organization and management practices, school climate, etc.), school context (e.g., size of student body, percent minority, etc.) and a component due to random chance. In this formulation, the unique impact of a school on students is conceptualized as the impact beyond the influence of student background factors. Raudenbush and Willms distinguish two types of school effects. Type A effects reflect the overall impact, both context and process influences, of attendance at a given school and Type B effects reflect the impact of a school on student performance that is attributable to school practices. Type A effects are likely of interest to parents that simply want the greatest boost to their child's performance irrespective of whether it stems from school context or process.

In this paper we focus on the estimation of Type A effects. According to Raudenbush and Willms, to achieve this it is only necessary to control for student background factors when estimating the unique impact of a school.

### Methodological Issues

Almost 30 years ago Dyer, Linn, & Patton (1969) recommended that conclusions

regarding school effects on student learning be based on the 'discrepancy' between actual mean achievement and the achievement expected given previous achievement and the social background of the students. As implied, the proposed strategy was simply to aggregate all student-level data to the school-level and, using multiple regression, calculate a standardized discrepancy score between actual student achievement and predicted student achievement. Criticisms of this approach have included (1) parameter estimation is generally inefficient, (2) problems of collinearity are common, (3) school effectiveness indices based on this model are generally unstable, and (4) the focus on means may mask important differences among subgroups of students with regard to school effects.

As an alternative to the school aggregate model, a number of researchers have studied student-level residual models. In this instance, students from all schools are pooled together and, without regard for school membership, the criterion of interest is regressed on student background and other student variables considered relevant. The residuals from this total sample regression are then averaged by school and taken as an index of school effectiveness. However, it can be shown that this strategy will yield biased estimates of school effects if there is heterogeneity of slopes among schools.

Finally, in response to the inherent multilevel nature of schools (students nested within classes; classes nested within schools, etc.) many researchers argue in favor of so-called hierarchical or multilevel statistical models as suitable for estimation of school effects. These models attempt to explicitly incorporate the layered nature of contemporary schooling into the analysis process and they permit the researcher to explicitly model complex within-school processes as consequences of between-school characteristics (see Mason, Wong, & Entwisle,

1983; Raudenbush & Bryk, 1986; Goldstein, 1986).

Following the notation of Raudenbush and Bryk (1986), let  $Y_{ij}$  be the outcome (e.g., achievement score) of student  $i$  ( $i=1, \dots, n_j$ ) in school  $j$  ( $j=1, \dots, J$ ). The within school regression model for predicting  $Y$  from, say, social class ( $X$ ), is given by,

$$Y_{ij} = \beta_{j0} + \beta_{j1}X_{ij} + e_{ij} \quad (1)$$

where  $\beta_{j0}$  is the intercept of the regression,  $\beta_{j1}$  is the slope and  $e_{ij}$  is an error term. The model in Equation 1 allows for the possibility that the regression parameters  $\beta_{j0}$  and  $\beta_{j1}$  vary across schools. If this variability is manifest, it could be due to differences in school policies, practices, or organization. A second model, the between model, might attempt to predict this variability on the basis of these or other school level attributes. If  $P$  represents an attribute of interest, then,

$$\beta_{jk} = \Phi_{0k} + \Phi_{1k}P_{jk} + a_{jk} \quad (2)$$

where  $k=0, 1$  for the model in Equation 2. If Equation 2 is substituted into Equation 1 the result is,

$$Y_{ij} = \Phi_{00} + \Phi_{10}P_{j0} + \Phi_{01}X_{ij} + \Phi_{11}P_{j1}X_{ij} + (e_{ij} + a_{j0} + a_{j1}X_{ij}) \quad (3)$$

This is a general model for the multilevel data problem and many researchers argue that it represents a significant advance in the effort to study schools and their effects on students.

Multilevel models have been criticized because of their complexity and because when used to estimate school effects, they make it difficult for small schools to appear effective (de Leeuw & Kreft, 1995; Morris, 1995). Also, there is evidence that, when the focus is on mean outcomes, the results from multilevel analyses and the school aggregate model do not differ substantially (Kennedy, Teddlie, & Stringfield, 1991).

## Methods

Data for the current study are taken from a statewide study of school effectiveness in Louisiana (For a detailed description of the measures used in this study, see Kemper, Kochan, Jarvis, and Durland, 1998). The focus of the current analyses is on public school students in the state during the 1997-98 school year. The design of this study was as follows:

- A. For each fourth grader in the state, the following data were obtained: (a) the standard score on the Core Battery of the Iowa Test of Basic Skills, (b) the student's third grade scale score on a criterion-referenced test administered statewide in 1997, and (c) an indicator of student participation in the state's free/reduced school lunch program.
- B. For all schools in the state, school effects were estimated with all three statistical techniques. In each instance, the dependent measure was fourth grade ITBS and the independent measures were the third grade criterion-referenced test score and an indicator of participation in the free/reduced lunch program. For the student level model, the studentized residuals from this regression were averaged by school for the school-level index. For the school aggregate model these indicators were averaged to the school and the regression executed. The studentized residuals from this regression constituted the effectiveness index for this model. For the multilevel model, both student-level and school-level variables were estimated simultaneously. The residuals for the centered intercept constituted the index of

effectiveness.

- C. Approximately 27 schools included in the analysis described above also participated in a school accountability study sponsored by the Louisiana Department of Education. In each instance, school averages were computed for each of the following:
  - A. Four school climate scales– School Safety, Academic Norms, Student Expectations, and Quality of Instruction.
  - B. Three indicators of time on task based on the Virgilio scale--Percent of Interactive Time on Task, Percent of Non-interactive Time on Task, and Percent of Total Time on Task.
  - C. Seven measures of effective instructional classroom practices :(1) Teacher Maintains an Environment Conducive to Learning, (2) Maximize Time for Instruction, (3) Provide Productive Learner Opportunities, (4) Effective Delivery of Instruction, (5) Appropriate Content, (6) Opportunities for Student Involvement, and (7) Assessment of Student Progress.

Simple correlations were computed between these averages and the indexes of school effectiveness generated with the three statistical procedures. Due to the size of the sample and the large number of statistics generated, strict statistical significance testing is not warranted. However, the data do indicate trends which are suggestive of future avenues of research.

## Results

Table 1 presents the regression results from the student-level and school aggregate models. In both instances, the models are reasonably effective at explaining variation in the criterion variable.

**Table 1. Student-Level and School-Level Regressions**

Predictors	Student-Level Model (Standardized Betas)	School-Level Model (Standardized Betas)
LEAP Composite (Student Level Data)	0.647	
Student F/R Lunch (Student Level Data)	0.198	
Average LEAP Composite (School Level Data)		0.57
Percent F/R Lunch (School Level Data)		0.35
Adj. R-Sq.	0.54	0.67

Table 2 presents the results of the mixed or multilevel model. Three models were fit. The first, the ‘unconditional model’ indicates that a considerable portion of the variation in student achievement on the fourth grade ITBS occurs between schools ( $Rho=0.23$ ). Model II shows that by considering two within-school predictors, performance on a criterion-referenced test (LEAPCOMP) and participation in the state’s free/reduced lunch program, the residual drops by 54 percent. Finally, Model III shows that by adding two school measures to the model PCTLUNCH (percentage of students participating in the free/reduced lunch program) and MLEAPCOMP (average performance on the criterion-referenced test), there is a substantial drop in the residual variation of the intercept of school means.

**Table 2. Mixed Model Results**

	Model I	Model II	Model III
Fixed Effects Estimates			
Intercept	494.1	493.0	168.2
LEAPCOMP	na	0.67	0.67
LUNCH	na	0.21	0.21
PCTLUNCH	na	na	0.85
MLEAPCOMP	na	na	0.59
Variance Estimates (Random)			
Intercept	2307	2501.7	643.9
Residual			
Value	7638	3484.7	3486.2
Percent Reduced		(54.4%)	(no change)
LEAPCOMP	na	0.04	0.04
LUNCH	na	0.02	0.02
Rho	0.23		
Akaike's Info. Criterion	-177544	-155343	-154962
Schwarz's Bayesian Criterion	-177552	-155372	154991

*Note. LEAPCOMP is a student level composite reflecting performance on a criterion-referenced test. LUNCH is an indicator of student participation in the state's free/reduced lunch program. PCTLUNCH is a school level measure of the percent of students receiving free/reduced price school lunches. MLEAPCOMP is a school level measure of average performance on the state's criterion-referenced test.*

Table 3 shows the simple correlations between the measures of school climate considered in the effectiveness indexes for the three models. Due to the size of the sample few of the values achieve statistical significance. However, the data do indicate that the three models yield similar patterns of relationships. The same pattern is apparent in Table 4 in which the focus is on time on task and instructional effectiveness. While more of these coefficients achieve statistical significance, the patterns across techniques are very similar.

**Table 3. Correlations Between School Effectiveness Indexes and School Climate (n=27)**

	Survey Group	Student-Level Model	School-Level Model	Mixed Model
School Safety	Principal	0.07	0.14	0.02
	Teachers	0.21	0.23	0.23
	Students	-0.06	-0.06	-0.06
	Parents	0.15	0.22	0.28
Student Expectations	Principal	0.15	0.09	0.14
	Teacher	0.20	0.18	0.19
	Students	-0.29	-0.24	-0.28
	Parents	0.04	0.14	0.14
Academic Norms	Principal	0.37	0.41*	0.36
	Teachers	-0.04	0.04	0.07
	Students	-0.18	-0.12	-0.11
	Parents	-0.10	-0.00	0.05
Instruction	Principal	0.23	0.31	0.27
	Teachers	0.06	0.11	0.16
	Students	-0.06	0.03	-0.02
	Parents	-0.07	0.02	0.09

**Table 4. Correlations Between School Effectiveness Indexes & Instructional Quality (n=21)**

	Student-Level Model	School-Level Model	Mixed Model
<i>Time on Task</i>			
Percent Interactive Time on Task	0.15*	0.18*	0.26*
Percent Non-Interactive Time on Task	-0.07*	-0.11*	-0.15*
Percent of Total Time on Task	0.15*	0.15*	0.23*
<i>Components of Effective Teaching</i>			
Environment Conducive to Learning	0.56	0.57	0.58
Maximize Time for Instruction	0.45	0.39*	0.45
Provide Productive Learner Opportunities	0.52	0.50	0.51
Effective Delivery of Instruction	0.52	0.50	0.54
Appropriate Content	0.41*	0.38	0.49
Opportunities for Student Involvement	0.45	0.46	0.47
Assessment of Student Progress	0.48	0.50	0.54

\* Value is not statistically significant at the 0.05 probability level.

### Conclusions

The purpose of this study was to examine the extent to which three competing statistical models for estimating school effects yielded similar correlations with recognized indicators of school effectiveness. These ‘other’ indicators included measures of school climate, time on task, and effective instructional practices. Due to the limited size of the sample, the results were interpreted with some caution. However, the patterns observed in the tables indicated that the three procedures examined yielded similar results with respect to the degree and direction of correlations.

## References

- de Leeuw, J. & Kreft, I. G. G. (1995). Questioning multilevel models. Journal of Educational and Behavioral Statistics, 20(2), 171-189.
- Dyer, H.S., Linn, R.L., and Patton, M.J. (1969). A comparison of four methods of obtaining discrepancy measures based on observed and predicted school system means on achievement tests. America Educational Research Journal, 6, 591-606.
- Fitz-Gibbon, C.T. (1996). Monitoring school effectiveness: Simplicity and complexity. In J. Gray, D. Reynolds, C. Fitz-Gibbon, & D. Jesson (Eds.) Merging traditions: The future of research on school effectiveness and school improvement (pp. 74-92). London: Cassell.
- Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. Biometrika, 73(1), 43-56.
- Goldstein, H. (1994). Multilevel cross-classified models. Sociological Methods & Research, 22(3), 364-375.
- Mason, W. M., Wong, G. M., & Entwistle, B. (1984). Contextual analysis through the multilevel I linear model. In S. Leinhardt (Ed.), Sociological methodology (pp. 72-103). San Francisco: Jossey-Bass.
- Morris, C.N. (1995). Hierarchical models for educational data: An overview. Journal of Educational and Behavioral Statistics, 20(2), 190-200.
- Kemper, L, Kochan, S., Jarvis, R. & Durland, M. (1998, April). Gathering and analyzing intensive school-level process data: A review of SEAP-II. Paper presented at the annual meeting of the American Educational Research Association, SanDiego, CA.
- Kennedy, E., Teddlie, C., & Stringfield, S. (1991, January). A multilevel analysis of phase II of the Louisiana school effectiveness study. Paper presented at the International Congress for School Effectiveness and School Improvement, Cardiff, Wales, UK.
- Muthen, B.O. (1994). Multilevel covariance structure analysis. Sociological Methods & Research, 22(3), 376-398.
- Raudenbush, S.W., & Bryk, A.S. (1986). A hierarchical model for studying school effects. Sociology of Education, 59, 1-17.
- Raudenbush, S. W. (1989). The analysis of longitudinal, multilevel data. International Journal of Educational Research, 13, 685-825.

Raudenbush, S. W., & Willms, J.D. (1995). The estimation of school effects. Journal of Educational and Behavioral Statistics, 20(4), 307-336.