DOCUMENT RESUME

ED 430 029 TM 029 773

AUTHOR	Yen, Shu-Jing; Schafer, William D.; Rahman, Taslima
TITLE	School Effect Indices: Stability of One- and Two-Level Formulations.
PUB DATE	1999-04-00
NOTE	20p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Elementary Education; Elementary School Students; *Least Squares Statistics; *Measurement Techniques; Predictor Variables; *Reliability
IDENTIFIERS	*Hierarchical Linear Modeling; Maryland School Performance Assessment Program; *School Effects

ABSTRACT

This study evaluated the comparative stability and agreement of three approaches to calculating school effects given both student-level and school-level data. The approaches were hierarchical linear modeling (HLM), ordinary least squares (OLS), and weighted least squares (WLS). Analyses were conducted using data from the 1998 Maryland School Performance Assessment Program for 23,461 third graders and 21,226 fifth graders. A two-level model was used for computing HLM school effects with four student-level predictors used to predict achievement in level one, and the school size used in level two to predict the level-1 intercept. The dependent achievement measure in OLS and WLS analyses was the average student score across the six content areas of the assessment. In OLS five variables were used as predictors, at the school level only, and in WLS the same strategy was used except that the sampling variance of the dependent variable was estimated for each school and used as the weighting variable. For OLS and WLS studentized residuals were used as the school effects measure. Results of the analyses indicate that, from a practical perspective, and all other considerations being equal, the HLM approach should be used for school effects measures on the basis of stability. The use of either of the school-level models appears to be viable in the event that only school-level data are available. Reasons for the greater stability of the HLM approach are discussed. (Contains 5 tables and 10 references.) (SLD)

School Effect Indices:
Stability of One- and Two-Level Formulations

Shu-Jing Yen
Maryland State Department of Education

William D. Schafer
University of Maryland, College Park
Maryland State Department of Education

Taslima Rahman
Maryland State Department of Education

Paper presented at American Educational Research Association, 1999 Annual Meeting, Montreal, Canada.

School Effects Indices: Stability of One- and Two-Level Formulations

Introduction

School effect indices (SEIs) are generally defined as differences between the school's actual mean performance and the school's expected mean performance based on the achievement of other schools with similar levels of student and school characteristics. At least three methods of formulating SEI have been proposed: studentized residuals computed by ordinary least squares (OLS) estimates (Mandeville & Anderson, 1987), studentized residuals computed by weighted least squares (WLS) estimates (Schafer, 1996), and HLM (Bryk & Raudenbush, 1992). The first two methods are one-level approaches in contrast to the two-level approach using Hierarchical Linear Modeling (HLM).

This study was designed to evaluate the comparative stability and agreement of these three approaches to calculating school effects given both student-level and school-level data. The stability of results produced by the three methods were judged in terms of their consistency across three different forms of similar tests administered to randomly equivalent groups within schools within grade level and across grade levels.

Theoretical Framework

It is recognized that characteristics of students and characteristics of schools may undermine the fairness of judging all schools on the same basis. Therefore, users of measures of school effectiveness have sought to take into account individual student characteristics such as prior achievement, ethnicity, and socioeconomic status (SES) as well as characteristics at the school level such as percentage of minority students, mean SES, and mobility. Mandeville and Anderson (1987) investigated the stability of school effects of South Carolina elementary schools where SEIs were defined as "studentized" residuals from the regression of students' achievement test scores onto earlier test performance and a measure of SES. They found the SEIs to be moderately stable across subject areas, but the SEIs reflecting the performance of students at different grade levels correlated only weakly, all less than 0.2. However, Schafer (1996) found cross-grade studentized residuals to be markedly larger than those of Mandeville (1988), when a similar method was used to measure Maryland school effects. Using residuals based on weighted least-squares regressions, the inter-grade correlations of SEIs ranged from .33 to .55.

Hierarchical linear modeling (Bryk & Raudenbush, 1992) is another prominent method of measuring student achievement by allowing for the investigation and possible control of various school-level factors that may otherwise confound such growth. Often educational effectiveness researchers (e.g., Phillips & Adcock, 1997; Webster & Mendro, 1997) employ two-level HLMs which control student variables at the first level, and school factors at the second level.

These procedures all provide indices that can be used to assess school effects. However, no direct comparison of them has been performed. The purpose of the present study is to evaluate the stability of these three indices, both across grades and across different samples of students within schools.

Method

Analyses were conducted using third and fifth grade data of the 1998 Maryland School Performance Assessment Program (MSPAP) that examines elementary schools in grades three, five, and eight in the areas of reading, writing, language usage, mathematics, science, and social studies.

MSPAP (Yen & Ferrara, 1996) is comprised of three test forms per grade. Each form is referred to as a cluster; forms are non-parallel test forms because content sub-areas are spiraled through them. Students are randomly assigned to testing groups to ensure that the students assigned to take each test form are equivalent in ability. During scaling and cluster equating process, three test forms are scaled onto a common scale (Maryland State Department of Education, 1998). With no loss of generality, testing clusters are also referred to as test forms in this paper.

Students' mean scaled scores across the six content areas were used as the outcome variable. For student level analyses, student characteristics including race, gender, English as a second language (ESL) status, special education status, and free and reduced lunch eligibility were used as explanatory variables. For school level analyses, variables representing school characteristics of the same set of variables plus school size were used in parallel ways in all three forms of SEI calculations.

Schools and Data

The complete student records of all Maryland public elementary schools with students in both third and fifth grade were obtained. Student-level and school-level data files were then created, one for each grade and cluster, using a rigorous selection process in three phases.

Phase 1 involved editing the student records and producing student-level predictors. Second semester students and students with incomplete test data were excluded. Four dummy variables were created: FEM (female =1, male =0); WHT (white =1, African American =0); BUY (receiving free/reduced price meal =0, paying full price meal =1); and REG (in Special education/ESL =0, in regular program =1).

Phase II involved aggregating across student records to produce school aggregates. In order to compare the stability of SEIs across clusters, only schools with all three clusters were included. Schools with less then 10 students were excluded. These data were eliminated because the means may be too unstable to be analyzed. To avoid potential estimation problem, schools with little or no variation with respect to the four predictors (FEM, WHT, BUY, and REG) were excluded. The proportion of female (FEM%), whites (WHT%), students paying full price meal (BUY%), and in regular programs (REG%) for each grade and cluster was first calculated. Schools were excluded if any of the clusters contain proportions that were considered as extreme, outside the range of 1% to 99%. This selection process resulted in the total of 286 schools in grade 3 and 267 schools in grade 5 out of the 886 elementary schools with grades 3 and 5. The total number of students included was 23,461 for grade 3 and 21,226 for grade 5. At the end of this phase, the student records underwent a second run of editing and school level files were then generated from the student level files.

Phase III involved computing student mean MSPAP score by taking the average of six MSPAP content area scale scores for each student. Students who received certain accommodations during testing did not receive a test score. Since excluding these students would exclude many special education, and ESL students; the lowest possible scale score for the content area was substituted for these students (Maryland State Department of Education, 1998). Missing test scores of non-accommodated students due to absences were replaced by the statewide mean.

Table I presents the distribution the five school-level predictors (FEM%, WHT%, BUY%, REG%, and SIZE-cluster size) and criterion variable (ACHV) by cluster and by grade. The correlation between the criterion variable and each of the five school-level predictors are presented in the last column (labeled 'correlation'). The inter-correlations among the five school-level predictors across clusters are reported in Table 2 for grades 3 and 5.

Table 1. Distribution of School-Level Variables

Grade	Cluster/Form	N[1]	Variable	Minimum	Maximum	Mean	Std Dev	Correlation[2]
3	A	286	ACHV	465.08	561.30	514.81	17.89	
			WHT%	2.50	98.20	63.58	26.34	0.48
			FEM%	23.50	80.00	48.03	10.10	0.04
			BUY%	4.80	97.50	66.81	21.38	0.53
			REG%	40.00	97.60	83.46	10.75	0.15
			SIZE	8	93	26.99	12.05	-0.01
	B	286	ACHV	454.45	573.77	516.55	19.29	
			WHT%	2.50	97.90	63.69	26.74	0.53
			FEM%	6.30	86.70	49.51	10.91	0.07
			BUY%	4.20	97.90	65.25	22.06	0.56
			REG%	47.60	97.90	84.12	9.41	0.19
			SIZE	8	79	26.05	11.80	-0.04
	C	286	ACHV	453.81	586.59	515.62	18.44	
			WHT%	2.00	97.90	62.41	25.94	0.50
			FEM%	13.00	80.00	50.00	10.13	0.03
			BUY%	2.70	97.40	65.62	21.72	0.61
			REG%	33.30	97.60	84.25	9.29	0.17
			SIZE	10	92	28.99	12.18	0.05
5	A	267	ACHV	454.87	562.27	517.10	19.69	
			WHT%	2.10	97.80	63.54	24.99	0.51
			FEM%	13.30	88.90	48.57	10.42	0.13
			BUY%	4.30	97.90	67.27	20.09	0.69
			REG%	30.80	97.60	82.49	11.67	0.36
			SIZE	7	90	29.28	14.28	-0.01
	B	267	ACHV	446.89	561.29	515.98	21.08	
			WHT%	2.90	98.00	65.28	24.54	0.50
			FEM%	16.70	83.30	47.94	11.29	0.09
			BUY%	4.50	97.70	68.77	21.13	0.65
			REG%	22.20	98.60	80.38	12.36	0.46
			SIZE	6	111	25.14	12.95	-0.02
	C	267	ACHV	458.20	572.96	517.66	19.39	
			WHT%	2.80	98.00	64.60	24.70	0.53
			FEM%	20.00	85.70	49.75	10.68	0.11
			BUY%	4.30	97.60	68.09	20.76	0.67
			REG%	33.30	98.10	82.97	11.05	0.26
			SIZE	7	114	25.07	12.01	0.11

Note: [1]Total Number of Schools. Note: [2]correlation between ACHV and predictors.
ACHV: Mean MSPAP score
WHT%: Percentage of white students
FEM%: Percentage of female students
BUY%: Percentage of students paying full lunch price
REG%: Percentage of students in regular program
SIZE: cluster size

Table 2. Correlations Among School-Level Predictors

Grade 3 (n=284)					
	FEM%	REG%	WHT%	BUY%	SIZE
FEM%	1.00	0.08	-0.05	-0.04	0.00
REG%	0.08	1.00	0.01	0.20	0.19
WHT%	-0.05	0.01	1.00	0.63	0.16
BUY%	-0.04	0.20	0.63	1.00	0.22
SIZE	0.00	0.19	0.16	0.22	1.00

Grade 5 (n=267)					
	FEM%	REG%	WHT%	BUY%	SIZE
FEM%	1.00	0.20	-0.03	0.05	0.03
REG%	0.20	1.00	0.08	0.26	0.26
WHT%	-0.03	0.08	1.00	0.57	0.12
BUY%	0.05	0.26	0.57	1.00	0.17
SIZE	0.03	0.26	0.12	0.17	1.00

FEM%: Percentage of female students
REG%: Percentage of students in regular program
WHT%: Percentage of white students
BUY%: Percentage of students paying full price meal
SIZE: School size

One-Level SEI methods

For the one-level SEI methods (Mandeville, 1987; Schafer, 1996), school-level mean MSPAP scores (ACHV) were regressed on five school-level explanatory variables: FEM%, WHT%, BUY%, REG%, and SIZE. The regression was done twice for each cluster for each grade level. One regression was based on the ordinary least square (OLS) estimation method while the other used weighted least square (WLS) where the weights were reciprocals of the criterion sampling variance (the square of the standard error of the mean) for each school. Since school means are more reliable for larger schools than smaller ones, WLS allows larger schools to have more contribution to the estimates than smaller schools. The residuals from the regression analyses were divided by their estimated standard errors to produce studentized residuals that served as the SEIs. These methods resulted in two residuals for each school; one based on OLS estimation and one on WLS.

Two-Level HLM method

For the HLM approach, we considered a Level-1 model where student mean MSPAP score (ACHV) was regressed on four dummy-coded variables (FEM, WHT, BUY, and REG):

$$Y_{ij} = \beta_{0j} + \sum_{q}^{4} \beta_{qj}(X_{qij} - X_{q..}) + r_{ij} \text{.....................(1)}$$

Where

Y_{ij} = MSPAP score for student i in school j,

B_{0j} = expected MSPAP score of student i whose X_{qij} is equal to the grand mean, $X_{q..}$,

B_{qj} = expected change in MSPAP score for a unit change in X_q, i.e., the expected differences between X_q=1 and X_q=0 in school j, and

r_{ij} = residual for student i in school j.

At Level-2, B_{0j} was regressed on school size (W_{1j}) and B_{qj} was constrained to be the same fixed value, γ_{q0}, for all schools:

$$\beta_{oj} = \gamma_{00} + \gamma_{01}(W_{1j} - W_{1.}) + \mu_{0j} \text{.........................} (2)$$

$$\beta_{qj} = \gamma_{q0} \text{..} (3)$$

Where

γ_{00} = expected MSPAP mean for schools whose W_{1j}=$W_{.}$,

γ_{01} = the relationship between the expected school mean B_{0j} and its school size,

μ_{0j} = unique effect of school j on the average MSPAP score after controlling for school size, and

γ_{q0} = fixed value of the slope B_{qj} across all schools.

Essentially, this is a random-intercept-model where the Level-1 intercept is assumed to vary across the Level-2 units (schools) but the within school slopes are constrained to be the same across schools (Phillips & Adcock, 1996). The unique effect of each school (μ_{0j}) after controlling for the explanatory variables were used as the SEI (Phillips & Adcock, 1996; Bryk & Raudenbush, 1992).

Results

The correlation coefficients of SEIs by grade and cluster are presented in Table 3. The three sub matrices along the main diagonal of the table indicate the consistency of each method across clusters or test forms. Since students are randomly assigned to forms within schools, these correlations indicate the consistency of SEI across three forms of similar tests administered to randomly equivalent groups within schools. For grade 3, the correlations of SEI between pairs of clusters for HLM (.61, .60, and .61) were slightly higher than those of other two methods (.57, .54, and .55 for WLS; .59, .54, and .55 for OLS).

The sub matrices on the off-diagonal of the table indicate the agreement among three methods for a given form or between pairs of forms. To examine the agreement among methods for a given form, the correlations among methods for the same form are compared. For grade 3, form A, for instance, the correlation between HLM and WLS (.93) was slightly lower than those between HLM and OLS (.95) and OLS and WLS (.97). Similar results were found when examining the agreement of three methods for form B and C. Lastly, the agreement among methods between any pairs of forms can be compared. The correlation between HLM and WLS is very similar to those between HLM and OLS, and between OLS and WLS, for all pairs of forms in grade 3. Parallel information for grade 5 is presented in the second part of Table 3.

Table 3. Intercorrelations among SEIs by Grade and Cluster

Grade 3 (n=284)

Method	Cluster/Form	**HLM** A	**HLM** B	**HLM** C	**WLS** A	**WLS** B	**WLS** C	**OLS** A	**OLS** B	**OLS** C
HLM	A	**1.00**	0.61	0.60	0.93	0.57	0.54	0.95	0.57	0.55
	B	0.61	**1.00**	0.61	0.56	0.94	0.52	0.57	0.95	0.54
	C	0.60	0.61	**1.00**	0.54	0.57	0.91	0.54	0.56	0.93
WLS	A	0.93	0.56	0.54	**1.00**	0.57	0.54	0.97	0.58	0.55
	B	0.57	0.94	0.57	0.57	**1.00**	0.55	0.57	0.97	0.55
	C	0.54	0.52	0.91	0.54	0.55	**1.00**	0.53	0.53	0.96
OLS	A	0.95	0.57	0.54	0.97	0.57	0.53	**1.00**	0.59	0.54
	B	0.57	0.95	0.56	0.58	0.97	0.53	0.59	**1.00**	0.55
	C	0.55	0.54	0.93	0.55	0.56	0.96	0.54	0.55	**1.00**

Grade 5 (n=267)

Method	Cluster/Form	**HLM** A	**HLM** B	**HLM** C	**WLS** A	**WLS** B	**WLS** C	**OLS** A	**OLS** B	**OLS** C
HLM	A	**1.00**	0.61	0.62	0.80	0.49	0.47	0.83	0.47	0.49
	B	0.61	**1.00**	0.59	0.46	0.86	0.48	0.46	0.88	0.51
	C	0.62	0.59	**1.00**	0.44	0.45	0.88	0.47	0.44	0.90
WLS	A	0.80	0.46	0.44	**1.00**	0.51	0.44	0.96	0.49	0.45
	B	0.49	0.86	0.45	0.51	**1.00**	0.49	0.50	0.97	0.49
	C	0.47	0.48	0.88	0.44	0.49	**1.00**	0.45	0.46	0.96
OLS	A	0.83	0.46	0.47	0.96	0.50	0.45	**1.00**	0.48	0.47
	B	0.47	0.88	0.44	0.49	0.97	0.46	0.48	**1.00**	0.47
	C	0.49	0.51	0.90	0.45	0.49	0.96	0.47	0.47	**1.00**

HLM: Hierachical Linear Model

WLS: Weighted Least Square

OLS: Ordinary Least Square

Note: Students are randomly assigned to forms within schools

To compare cross-grade consistency of SEIs, the average SEI across forms was computed at each grade for each method for schools where both grades exists. There are 184 schools included in this part of the analysis. The correlation coefficients of SEIs between grade 3 and 5 are presented in Table 4 for each method. The between-grade correlation was .64 for HLM, .58 for WLS, and .57 for OLS.

Table 4. Correlations of SEI between Grades by Methd

		Grade 3		
Method		HLM	WLS	OLS
	HLM	**0.64**	0.53	0.53
Grade 5	WLS	0.57	**0.58**	0.56
	OLS	0.58	0.57	**0.57**

Note: N=184

In order to investigate the predictability of school achievement with the studied models, the squared correlations between each of the three school effects indices and the raw school achievement means were evaluated for all three forms at each grade level (see Table 5). The result was interpreted as the extent to which the model provides measures that are sensitive to absolute achievement as opposed to achievement when variables in the predictor set are controlled. For grade 3, the squared correlations for the HLM indices are the highest (in the .80's), followed by OLS indices (.60's) and WLS indices (.50's). Similarly, the squared correlations for the HLM indices are the highest (in the .80's), followed by OLS indices (.40's) and WLS indices (.30's) for grade 5.

Table 5. Squared Correlations between SEI and School Means

Method	HLM			WLS			OLS		
	A	B	C	A	B	C	A	B	C
Grade 3	0.79	0.78	0.82	0.58	0.53	0.52	0.64	0.59	0.59
Grade 5	0.81	0.81	0.78	0.36	0.37	0.42	0.41	0.45	0.49

Note: N=184.

Summary and Discussion

This study evaluated the comparative stability and agreement of three approaches to calculating school effects given both student-level and school-level data. The approaches were hierarchical linear modeling (HLM), ordinary least squares (OLS) and weighted least squares (WLS). A two-level model was used for computing HLM school effects where in Level-1, four student-level predictors were used to predict student achievement; in Level-2, the school size was used to predict the Level-1 intercept. The four Level-1 predictors were regular vs. special program; white vs. non-white, female vs. male, and buy at full price vs. free or reduced price meals; the school-level predictor of intercept (centered model) was test group size. The school effects measure was the school-level error term in the intercept prediction equation.

The dependent achievement measure in OLS and WLS analyses was the average student score across the six content areas in the assessment. In OLS, the same five variables were used as predictors, but at the school level, only; school means were used for variables that were level-one predictors in the HLM method. The same strategy was used for WLS except that the sampling variance of the dependent variable was estimated for each school and used as the weighting variable. For OLS and WLS, studentized residuals were used as the school effects measure.

Stability was studied by replicating the school effects measures across randomly equivalent subgroups in schools where the subgroups were assessed on independent measures, called forms or clusters, of the same achievement construct. The average intercorrelation among forms at third grade was .61 for HLM, .56 for OLS, and .55 for WLS. The fifth grade intercorrelations were .61 for HLM, .48 for WLS and .47 for OLS. The stability of the hierarchical method (HLM) was greater than those for the school-level methods (OLS and WLS), with no clear difference between the latter two.

Another aspect of stability was consistency of school effects between grades three and five. For this analysis, the form differences were ignored and a single school effect was found at each grade level. The between-grade stability for HLM was .64, for WLS it was .58, and for OLS it was .57. This replicates the same pattern that was evident in the among-form stability results, with HLM most stable.

Agreement was studied by comparing the intercorrelations among the methods. When forms were the same, the agreement between OLS and WLS averaged .97, between HLM and OLS averaged .91, and between HLM and WLS averaged .89. Each of these is the average of six correlation coefficients, three at each grade level. The agreement between the two school-level methods was greater than that between either school-level method and HLM.

When forms were different, the agreement correlations were smaller. The average intercorrelation was .52 between OLS and WLS, and between HLM and OLS, and .51 between HLM and WLS. Each of these is the average of 12 correlation coefficients, six at each grade level. When different test forms are used for different students, there does not appear to be much difference among the correlations across pairs of methods.

Between-grade agreement was also studied. Again, there does not appear to be much difference among the methods. The two correlations between OLS and WLS averaged .57, the average of the two correlations between HLM and OLS was .56, and the two correlations between HLM and WLS averaged .55. These correlations are more consistent with Schafer's (1996) cross-grade findings than they are with Mandeville's (1988), where little stability was found. It is possible that MSPAP, a locally developed statewide performance assessment, is more an assessment of a school's overall educational program than is the test used by Mandeville, which may have focused on grade-specific curriculum objectives.

A simple school-level multiple regression equation usually generates a squared multiple correlation (R^2) between the criterion and the predictors. In this situation, the R^2 represents the proportion of school mean achievement variance that may be explained by the five predictors. The quantity ($1 - R^2$) represents the squared correlation between the residuals and mean school achievement. However, in all the procedures studied here, residuals from a simple regression equation were never used. In both the OLS and WLS cases, the residuals were studentized before use. In order to investigate the predictability of school achievement with the actual studied models, the squared correlations between the three school effects indices and the raw school achievement means were evaluated for all three forms at each grade level. The result was interpreted as the extent to which the model provides measures that are sensitive to absolute achievement as opposed to achievement when variables in the predictor set are controlled.

The six squared correlations for the HLM indices averaged .80; for OLS indices, the average squared correlation was .53; and the average squared correlation for WLS indices was .46. Because they are not derived from multiple regression equations, none of these R^2 values may be interpreted as representing a partitioning of between-school variance into portions due to and independent of the predictor variables. Nevertheless, it may be reasonable to consider a smaller R^2 to represent a greater ability to remove from between-school variance that portion explainable by the predictors. The WLS method appears best able to remove predictor effects, followed by the OLS method, and then the HLM method. However, the potential for the residuals to be correlated with the predictors in each of the three methods compromises that conclusion.

From a practical perspective, the results of this study suggest that, all other considerations equal, the HLM approach should be used for school effects measures on the basis of stability. Nevertheless, the use of either of the school-level models appears to be viable in the event that only school-level data are available. Both were almost as stable and did not differ much from the HLM approach, with agreement correlations in the middle .80's within forms and between-form agreement correlations virtually as high as those between the two school-level models. The high agreement between the two school-level models, with the average within-form intercorrelation of .97, suggests that there is little difference between them based on the criteria in this study.

The greater stability of the HLM approach might be the result of greater precision in the estimation of homogeneous regression coefficients in the school level models as opposed to the estimation of regression coefficients in the between-school models. Except for the complicating presence of the test group size as a level-2 predictor and the use of maximum likelihood estimation, the HLM residuals are analogous to deviations of adjusted school means from the grand mean in an analysis of covariance model. Adjustments to the school means are made using the within-group equations, which are estimated using data from all students. They may be more stable than equations generated directly from school means. Especially for small schools, HLM method should produce more stable measure of school effects. But that stability comes at the expense of an assumption that the regressions are homogeneous. If that assumption is not valid, then the coefficients do not estimate existing parameters. As in any analysis of covariance context, the presence of interaction between a covariate and a grouping variable threatens interpretation of adjusted means, particularly when students are not randomly assigned to groups, as in typical school effects research.

References

Bryk, A. S. & Raudenbush, S. W.(1992). The Hierachical linear model: Application and data analysis methods. Newberry Park, CA: Sage.

Mandeville, G. K. (1988). School effectiveness indices revisited: Cross-year stability. Journal of Educational Measurement, 25, 349-356.

Mandeville, G. K. & Anderson, L. W. (1987). The stability of school effectiveness indices across grade levels and subject areas. Journal of Educational Measurement, 24, 203-216.

Maryland State Department of Education (1998). MSPAP technical report, 1998. Maryland State Department of Education, Baltimore.

Phillips, W. G. & Adcock, E. P. (1997). Measuring school effects with hierachical linear modeling: Data handling and modeling issues. Paper presented at American Educational Research Association Annual Meeting, Chicago.

Raudenbush, S. W., & Willms, J. D. (1995) The estimation of school effects, Journal of Educational and Behavioral Statistics, 20, 4, 307-335.

Schafer, W. D. (1996). Study of higher-success and lower-success elementary schools: Phase two. Paper presented at the American Educational Research Association Annual Meeting, New York.

Webster, W. J., Mendro, R. L., (1997). The Dallas value-added accountability system, in Millman, J., editor, Grading teachers, grading schools: Is student achievement a valid evaluation measure? Corwin Press, Thousand Oaks, CA.

Webster, W. J., Mendro, R. L., Bembry, K. L., & Orsak, T. H. (1995). Alternative methodologies for identifying effective schools. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Yen, W. M. & Ferrara, S. (1997). The Maryland School Performance Assessment Program: Performance assessment with psychometric quality measurement, Educational and Psychological Measurement, 57, 60-84.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM029773

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: School Effect Indices: Stability of One- and Two-Level Formulations

Author(s): Shu Jing Yen William D. Schafer Taslima Rahman

Corporate Source: Maryland State Department of Education

Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
[X]	[ ]	[ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here,→ please

Signature: William D. Schafer	Printed Name/Position/Title: William D. Schafer Director of Student Assessment	
Organization/Address: Maryland State Department of Education Baltimore, MD 21201-2595	Telephone: 410-767-0081	FAX: 410-333-0052
	E-Mail Address: bschafer@msde.state.md.us	Date: 4/26/99

(over)