Practical Applications of Hierarchical Linear Models to District Evaluations

Gary W. Phillips, Ph.D
National Center for Education Statistics, U.S. Department of Education
Eugene P. Adcock, Ph.D
Prince George's County Public Schools, Maryland

This paper provides a practical application of hierarchical linear models (HLM) in an evaluation of effective schools for a large school district (the Prince George's County School District in Maryland). The HLM model is used first to rank elementary schools on their effectiveness at improving student learning in reading and mathematics and is also used to evaluate which factors contribute to school effectiveness. Teacher training was found to be the largest factor at contributing to school effectiveness after controlling for school context variables (School poverty and percent minority). It was found that this approach not only provides a rigorous statistical procedure, but also was easy to communicate to education policy makers. Plans for future analysis are also include.

This paper presents a "value-added" study of the effectiveness of 119 Prince George's County elementary schools' reading and mathematics programs. As suggested by Bryk and Weisberg (1976), a "value-added" approach is based on the use of a growth model to estimate the amount of growth that would be expected for a group participating in an educational program or school if they did not participate but instead were in the "regular" or "comparison" program. The actual change of the participants is compared to the predicted change and the difference is the "value-added." This approach is particularly well suited for evaluating school effectiveness or program effectiveness in their natural setting.

Following recent school effect studies of McPherson, 1992, Sanders & Horn, 1994, and Raudenbush & Willms, 1995, our practical application of the value-added model focuses on the influences of school practices (vs school context) which provide instructional treatments that raise student academic achievement regardless of the level at which the students enter the educational venue. A value-added school Effectiveness Index (EI) for the county was obtained from a new analysis of the statewide 1994 Maryland School Performance Assessment Program (MSPAP) controlling for student family socio-economic status and school population's percent of student poverty (Adcock, 1995). Hierarchical linear modeling analysis results provided an EI value for ranking each school's performance as either "ineffective," "no value-added," or "effective." Additional analyses examined the impact of a variety of teacher, school, and student background variables on schools' instructional effectiveness.

Method

In order to achieve the goals of the evaluation it was decided that the best statistical methodology would be hierarchical linear modeling. Hierarchical linear model analyses are like statistical microscopes in that they allow researchers and policy makers to see relationships in the data unconfounded by other variables. For example, the study attempted to determine the effectiveness of schools at promoting student achievement with the effects of student SES and school poverty controlled. In addition HLM was used to assess which school variables contribute to school effectiveness. It should be noted that only extant data were used in the study. Plans are under way to expand the data base so that the influence of other variables (such as school resources and instructional practices) may be assessed.

Recent articles on HLM applications were helpful in conceptualizing and explaining the analysis to policy makers. For example, Raudenbush and Willms (1995) distinguished between *Type A* and *Type B* school effects. *Type A* effects are often the interest of parents and real estate agents, whereas *Type B* effects are of more interest to education policy makers and evaluators. In a *Type A* effect we consider a school effective when students do well "regardless of whether that school's effectiveness derives from the superb practice of its staff, from its favorable student composition, or from the beneficial influence of the social and economic context of the community in which the school is located. But it would clearly be unfair to reward school staff purely on the basis of their *Type A* effects, given that the staff is only partly responsible for those effects" (p.310). The *Type B* effect is the effect of school practice on student learning unconfounded by school context variables. HLM models are ideally suited to estimate *Type B* effects because they provide an index of school practice variables (curriculum content, instructional practice, and school resources) after factoring out the influence of school context variables (student demographics, community characteristics). "The *Type B* effect is the effect school officials consider when evaluating the performance of those who work in the schools. A school with an unfavorable context could produce a large *Type B* effect through the effort and talent of its staff. The school would rightly earn the respect of school evaluators even though parents shopping for a large *Type A* effect might not want to choose that school" (p. 310).

Past Practice

Before proceeding with the HLM model it is instructive to review an approach that many other evaluators have used in the past. In order to rank the schools based on an index of *Type B* school effectiveness that is unconfounded with student and school poverty education researchers have often used an ordinary least-squares regression (OLS) equation which includes a school poverty measure. One example of this would be the following single level equation

$$Y_{ij} = \beta_0 + \beta_1(X_{.j} - X_{..}) + r_{ij}. \qquad (1)$$

In equation 1 β_0 represents the predicted level of student achievement when school ($X_{.j}$) poverty equals the grand mean, β_1 represents the effects of poverty for schools, and r_{ij} is the error term. The model is essentially an OLS regression model with schools as the unit of analysis.

When this equation is used then the usual measure of *Type B* effectiveness is the difference between the actual mean performance of the school and the predicted performance based on school poverty, i.e., $Y_{.j} - Y_{..} - \beta(X_{.j} - X_{..})$

Although these OLS estimates have often been used, they have several statistical problems in comparison to HLM models. In the first place, they are unbiased but less efficient. The HLM estimates are both unbiased and more efficient. This is accomplished in HLM through the Bayesian procedure which uses not only the data available within a school for the regression equation but also uses all available data from other schools. The regression equation for each school is a weighted composite based on the information available in that school and the information available in the entire data set. The relative weights from these two sources depend on the precision of the parameter estimates. As the sample size of the school increases the weight of the school information dominates the parameter estimate. A by-product of the HLM solution to providing more stable estimates in smaller schools, is the added benefit that HLM more clearly partitions the variance within- and between schools, disentangles hypothesis testing for student versus school effects, and provides a general, yet flexible, way of modeling even with large numbers of student and school variables.

HLM Model

Instead of using the above single level model, the Prince George's County Effectiveness School Evaluation used a two level HLM model to assess *Type B* effects.

Level I

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - X_{..}) + r_{j}, \qquad (2)$$

where

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

β_{0j} = expected MSPAP score for a student whose value on X_{ij} is equal to the grand mean, $X_{..}$.
β_{0j} is an adjusted mean for school j such that $\beta_{0j} = \mu_{y\,j} - \beta_{1j}(X_{ij} - X_{..})$,

β_{1j} = expected change in MSPAP scores for a unit change in SES (i.e., the expected difference between SES = 1 and SES = 0) in school j, and
r_{ij} = residual for student I in school j.

Level II

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(W_{1j} - W_{1.}) + \mu_{0j}, \qquad [3]$$

where
γ_{00} = expected MSPAP mean for a PG County schools for students whose $W_{1j} = W_{1.}$,

γ_{01} = the relationship between the expected school mean achievement (β_{0j}) and percent poverty in the school (W_{1j}),

μ_{0j} = unique effect of school j on the average achievement after controlling for W_{1j}.

$$\beta_{1j} = \gamma_{10}, \qquad (4)$$

where

γ_{10} = the fixed value of the slope (β_{1j}) across all schools (pooled within-school regression coefficient).

The above HLM model is called a random-intercept model because the β_{0j} is assumed to vary randomly across the level II units (schools). However, in the model, the within-school slopes are assumed to be constant across schools.

An important by-product of the HLM model is that it can be used to derive an index of the *Type B* effectiveness of schools at raising academic achievement after controlling for relevant student and school level variables. Once the index of effectiveness is obtained then schools can be ranked according to this index. The current index only controls for student and school level poverty. However, as data become available, and policy makers decide which variables they would like to include, then the index can be refined in the future. The effectiveness index μ_{0j} used in the evaluation to date is derived by the following steps.

1. Substitute equations {3} and {4} into equation 2

$$Y_{ij} = \{\gamma_{00} + \gamma_{01}(W_{1j} - W_{1}.) + \mu_{0j}\} + \{\gamma_{10}\}(X_{ij} - X..) + r_{ij},$$

$$\mu_{0j} = Y_{ij} - [\gamma_{00} + \gamma_{01}(W_{1j} - W_{1}.) + \gamma_{10}(X_{ij} - X..)].$$

2. Average over i within j,

$$\mu_{0j} = Y._{j} - [\gamma_{00} + \gamma_{01}(W_{1j} - W_{1}.) + \gamma_{10}(X._{j} - X..)]. \qquad (5)$$

The effectiveness index in equation 5 is a measure of the schools level of academic achievement after controlling for student background effects, $\gamma_{10}(X._{j} - X..)$, and school context effects, $\gamma_{01}(W_{1j} - W_{1}.)$. It can be interpreted as the difference 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 poverty).

Results

The above index was calculated for all 119 elementary schools and is included in the full report. Schools that are more than one standard deviation above what is expected (based on their levels of poverty) are considered effective. Schools that are within one standard deviation are considered doing about as well as can be expected (no value-added). Schools that are one standard deviation below are considered ineffective and are not performing up to the levels of other schools with similar levels of poverty.

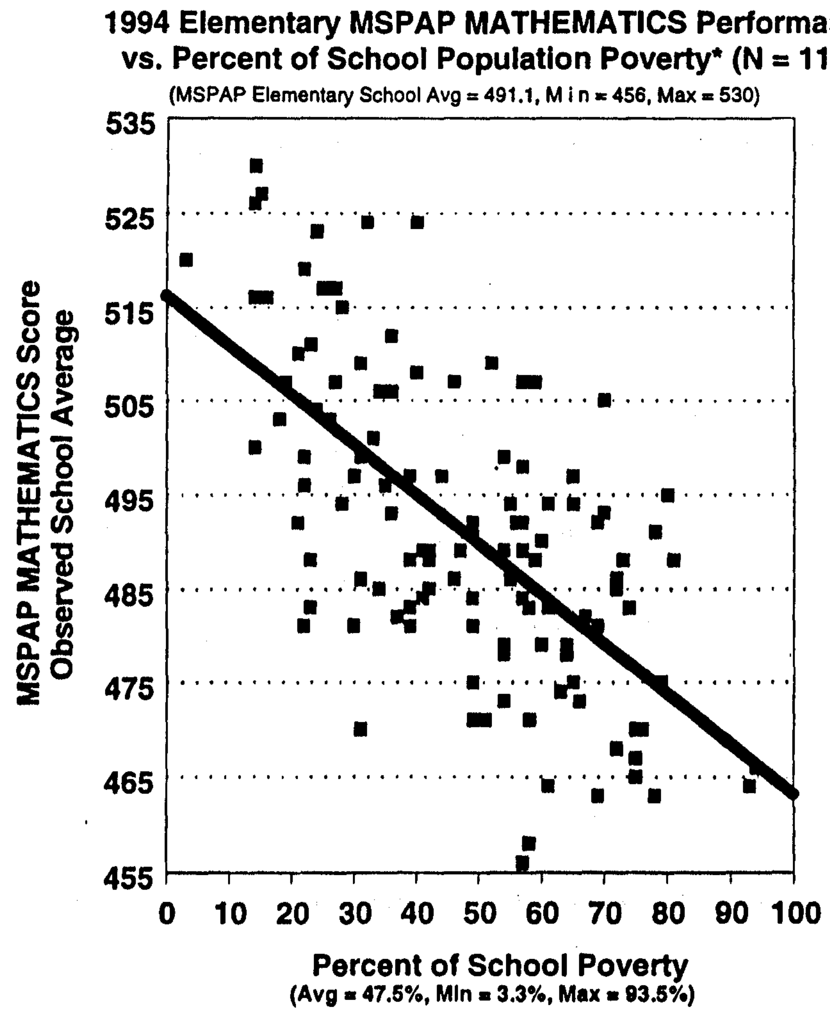
Figure 1 (mathematics) and figure 2 (reading) provide a graphic representation of the relationship between the schools actual observed average score and the schools level of poverty.

Figures 1 and 2 clearly show that there is a strong negative relationship (the correlation was -.70 for reading and -.64 for math) between the schools achievement and the population of poverty in the school. As the level of poverty goes down the school tends to achieve more. These graphs represent the type of *Type A* effects discussed above. Unfortunately, in figures 1 and 2 it is impossible to disentangle the effects of school practice from school context. An "evaluation" of schools would need to first control for school context variables. This is accomplished in figures 3 and 4 by controlling for student and school SES.

Figure 3 (mathematics) and figure 4 (reading) plot the effectiveness index against the schools level of poverty. These graphs are examples of *Type B* effects that school officials need in order to determine which schools have the most effective practice.

The data points above the upper boundary line in Figures 3 and 4 are those schools identified as effective while those below the lower boundary line are ineffective. It should be noted that at all levels of poverty, there are many schools that meet expectation (within one standard deviation), some that are effective (above one standard deviation) and some that are ineffective (below one standard deviation). It should be noted that the effectiveness index is not correlated with school poverty. This is why the effectiveness index as an accountability measure is an improvement over the mean MSPAP score. The average MSPAP score is highly correlated with school poverty and in fact 40% of the variance in school math performance can be attributed to school poverty as can 50% of the variance in reading. The important thing about figures 3 and 4 is that they provide a way of comparing schools with a more even playing field. It shows that schools with similar levels of poverty have differing levels of student achievement. Some schools are not achieving well even though they have low levels of poverty and some are doing very well in spite of very high levels of poverty.

Figure 1

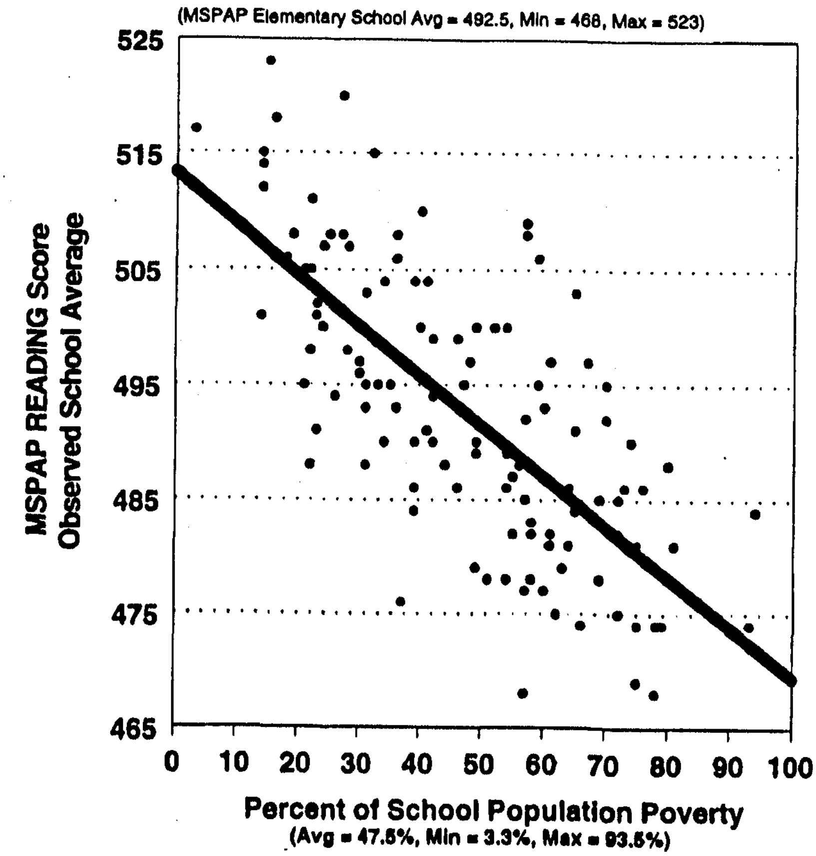


* Correlation between school poverty and MSPAP Math = -.64
* School poverty calculated from Elementary grade students receiving F/R Meals and administered the MSPAP.

Research, Evaluation & Accountability

Figure 2

1994 Elementary MSPAP READING Performance vs. Percent of School Population Poverty* (N = 119)

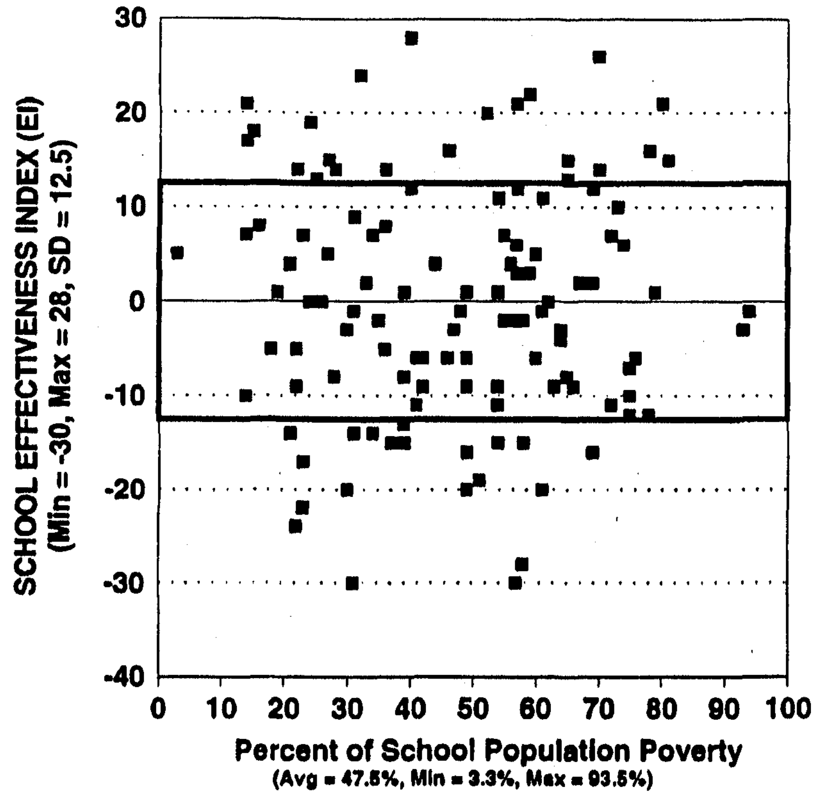


* Correlation between school poverty and MSPAP READING = -.70
* School poverty calculated from Elementary grade students receiving F/R Meals and administered the MSPAP.
Research, Evaluation & Accountability

Figure 3

MATHEMATICS EFFECTIVENESS INDEX (EI)
vs. School Population Poverty

1994 MSPAP Performance for 119 Schools

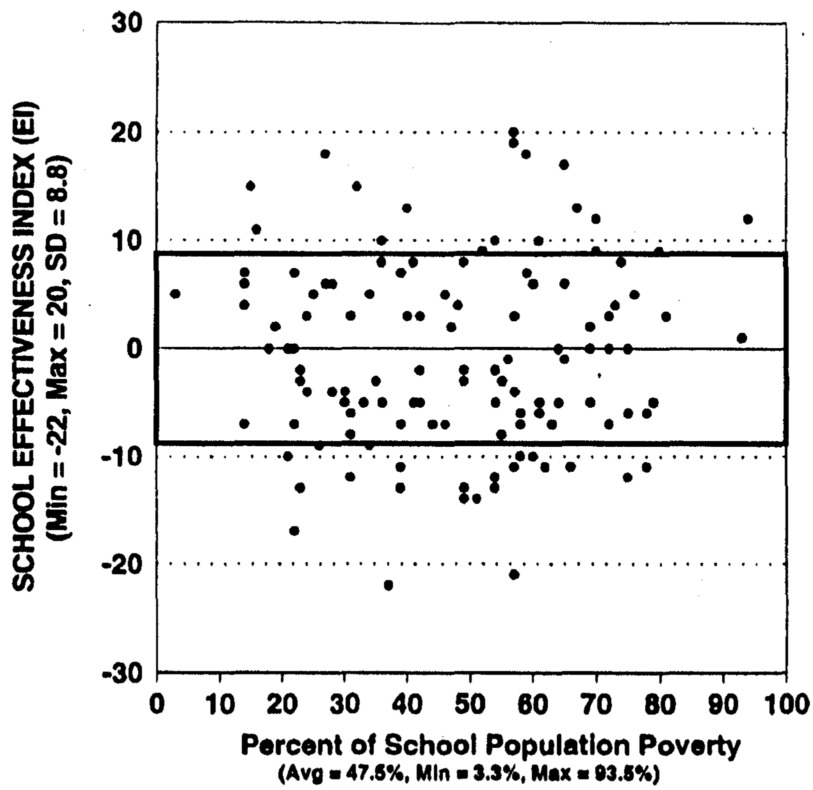


Note: Dots outside the middle zone indicate schools one or more standard deviations (+/- 12.5) from expected performance.Correlation is ZERO(r = 0.01, p = 0.94).
School poverty calculated from Elementary grade students receiving Free/Reduced Meals and administered the MSPAP.
Research, Evaluation & Accountability

Figure 4

READING EFFECTIVENESS INDEX(EI) vs. School Population Poverty

1994 MSPAP Performance for 119 Schools



Note: Dots outside the middle zone indicate schools one or more standard deviations (+/- 8.8) from expected performance. Correlation is ZERO (r = 0.02, p = 0.86).
School poverty calculated from Elementary grade students receiving Free/Reduced Meals and administered the MSPAP.
Research, Evaluation & Accountability

Additional Analyses

After ranking the schools based on the effectiveness index we also are interested in those variables that help to explain the rankings. The question we are attempting to answer is what factors help explain why some schools are more effective than others (i.e., a *Type B* effect based on school practice variables) after we have controlled for school context variables). This line of inquiry is only in its initial stages in the Prince George's County Evaluation. Additional data need to be collected that relate to additional school practices such as fiscal resources, teacher characteristics, instructional practices and curriculum offerings. However, as a first attempt at this analysis, extant school level data were used in which school poverty and percent minority are treated as school context variables and level of teacher training and Milliken status are treated as school practice variables. The HLM model that was fit to the data was as follows:

LEVEL I

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - X_{..}) + r_{ij}, \text{ where} \qquad (6)$$

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

β_{0j} = expected MSPAP score for a student whose value on X_{ij} is equal to the grand mean, $X_{..}$. β_{0j} is an adjusted mean for group j such that $\beta_{0j} = \mu_{y\,j} - \beta_{1j}(X_{ij} - X_{..})$.

β_{1j} = expected change in MSPAP scores for a unit change in SES (i.e., the expected difference between SES = 1 and SES = 0) in school j, and

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

LEVEL II

$\beta_{oj} = \gamma_{00} + \gamma_{01} (W_{1j} - W_{1 \cdot}) + \gamma_{02} (W_{2j} - W_{2 \cdot}) + \gamma_{03} (W_{3j} - W_{3}) + \gamma_{04} (W_{4j} - W_{4 \cdot}) + \mu_{0j}$, where (7)

γ_{00} = expected MSPAP mean for a non-Milliken school whose $W_{1j} = W1.$, $W_{2j} = W_{2}.$, $W_{3j} = W_{3}.$, $W_{4j} = 0$,

γ_{01} = the relationship between the expected school mean achievement (β_{0j}) and percent poverty of the school (W_{1j}) after controlling for other school level variables,

γ_{02} = the relationship between the expected school mean achievement (β_{0j}) and percent minority of the school (W_{2j}) after controlling for other school level variables,

γ_{03} = The relationship between expected school mean achievement (β_{0j}) and levels of teacher training in the school (W_{3j}), after controlling for other school level variables,

γ_{04} = difference between Milliken and non-Milliken expected school mean achievement (W_{4j}) after controlling for other school level variables, and

μ_{oj} = unique effect of school j on the average achievement after controlling for W_{1j}, W_{2j}, W_{3j} and W_{4j}.

$\beta_{1j} = \gamma_{10} + \gamma_{11} (W_{1j} - W_{1 \cdot}) + \gamma_{12} (W_{2j} - W_{2 \cdot}) + \gamma_{13} (W_{3j} - W_{3 \cdot}) + \gamma_{04} (W_{4j} - W_{4 \cdot})$, where (8)

γ_{10} = expected MSPAP slope for a non-Milliken school whose $W_{1j} = W1.$, $W_{2j} = W_{2}.$, $W_{3j} = W_{3}.$, $W_{4j} = 0$,

γ_{11} = the relationship between the expected school slope (β_{1j}) and percent poverty of the school (W_{1j}) after controlling for other school level variables,

γ_{12} = the relationship between the expected school slope (β_{1j}) and percent minority of the school (W_{2j}) after controlling for other school level variables,

γ_{13} = The relationship between expected school slope (β_{1j}) and levels of teacher training in the school (W_{3j}), after controlling for other school level variables,

γ_{14} = difference between Milliken and non-Milliken expected school slopes (W_{4j}) after controlling for other school level variables.

To distinguish this more elaborate model (equations 6, 7 and 8) from the one used to rank the schools (equations 2, 3 and 4) we will refer to the earlier model as HLM_1 and the current model as HLM_2.

The results in Tables 1 and 2 are the primary findings from the fuller HLM_2 model. In each case an HLM analysis was conducted that included all available variables. Variables that did not show a significant relationship were deleted in the final model. The results for Table 2 were as follows: across all 119 schools as the percent of poverty increased 10% the mean math MSPAP score dropped 2.1 points; as the percent minority increased 10% the mean MSPAP score decreased 1.5 points; and, as the level of teacher training increased one level (e.g., from the bachelors to the bachelors plus 30 credit hours) the average MSPAP score increased 7 points. The level of teacher training was by far the variable with the strongest influence on the achievement of schools. It is also important to note that whether the school was a Milliken school was also a variable in the initial model. However, there was no significant difference in Milliken versus non-Milliken schools (after controlling for student SES and school poverty) so the variable was dropped in the final model. The results were similar, but less dramatic, for reading in Table 3.

In addition to assessing the influence of the effect of the above variables on average school achievement, the HLM_2 was also used to assess a question of equity. The issue here is the extent to which the schools achievement is really due to the SES of the student population. An index of this is captured by the level I β_1 coefficients. The level I β_1 coefficient represents the within -school relationship of student achievement to student SES. A large β_1 indicates that there is a large relationship between achievement and SES within the school. A more desirable situation would be a small β_1 which indicates that the schools level of achievement is related to variables other than the SES of the student population. Tables 2 and 3 also contain these analyses. The results indicate: as the percent poverty increases 10% the β_1 decreases 1 point for math and 1.3 for reading; as the percent poverty increases 10% the β_1 drops 1 point for both math and reading; and as the average level of teacher training increases one level the β_1 increases by 4.7 points in reading. This last statistic is significant in that it means reading achievement is more related to the student's SES in the schools with the highest level of teacher training. This finding was not observed for math.

Table 1: Primary Findings for HLM_2 in Math

	Initial Model	Final Model
Model for Predicted School Means, β_0		
Intercept, _00	490.2 (1.2)	490.6 (1.1)
Percent Poverty, _01	-2.1 (0.6)*	-2.1 (0.6)*
Percent Minority, _02	-1.6 (0.6)*	-1.5 (0.6)*
Teacher Training,_03	7.0 (3.1)*	7.0 (3.2)*
Milliken Program, _04	2.0 (3.4)	
Model For SES Slope, β_1		
Intercept, _10	18.1 (0.9)*	18.2 (0.8)*
Percent Poverty, _11	-1.0 (0.5)**	-1.0 (0.5)*
Percent Minority, _12	-1.5 (0.65*	-1.7 (0.4)*
Teacher Training, _13	2.2 (2.3)	
Milliken Program, _14	0.6 (2.7)	

* There is at least a 95% chance that the true regression effect is not equal to zero.
** There is at least a 90% chance that the true regression effect is not equal to zero.

Table 2: Primary Findings for HLM_2 in Reading

	Initial Model	Final Model
Model for Predicted School Means, β_0		
Intercept, _00	492.5 (0.8)	492.2 (0.8)
Percent Poverty, _01	-1.8 (0.4)*	-1.8 (0.4)*
Percent Minority, _02	-1.0 (0.4)*	-1.1 (0.4)*
Teacher Training,_03	4.8 (2.1)*	4.9 (2.1)*
Milliken Program, _04	-1.4 (2.4)	
Model For SES Slope, β_1		
Intercept, _10	16.5 (0.9)*	16.4 (0.8)*
Percent Poverty, _11	-1.2 (0.5)*	-1.3 (0.5)*
Percent Minority, _12	-1.0 (0.5)**	-1.0 (0.5)*
Teacher Training, _13	4.6 (2.4)*	4.7 (2.4)*
Milliken Program, _14	-0.7 (2.8)	

* There is at least a 95% chance that the true regression effect is not equal to zero.
** There is at least a 90% chance that the true regression effect is not equal to zero.

Future Plans

The results described in this paper are based on an initial effort to evaluate the effectiveness of schools with a limited number of variables. Future plans include 1) increasing the grade levels to include both elementary and middle school, 2) including measures of science in addition to math and reading, 3) using both SES and percent minority as context variables, and 4) extending the number of school practice variables to include teacher training, teacher experience, fiscal resources available in the school, and educational effort from the student, teacher and the parents. In addition, more distant plans call for the use of a three-level HLM model in which change over time is modeled at the first level, student differences at the second level and school effects at the third level.

References

Adcock, E. P. (1995). *Value-Added Effective Schools Study for Elementary Schools: 1994 Maryland School Performance Assessment Program Results*, Research, Evaluation & Accountability, Prince George's County Public Schools, MD.

Bryk, A. S., & Weisberg, H. I. Value-added analysis: A dynamic approach to the estimation of treatment effects. *Journal of Educational Statistics,* 1976, 1, 127-155.

McPherson, A. F. (1992). *Measuring added value in schools* (NCE Briefing No. 1). London: National Commission of Education.

Raudenbush, S. W., & Willms, J.D. (1995) The estimation of School Effects, *Journal of Educational and Behavioral Statistics*, Winter 1995, Vol. 20, No. 4, pp. 307-335.

Sanders, W. L., & Horn, S. *The Tennessee Value-Added Assessment System (TVASS): mixed Model Methodology in Educational Assessment,* 1994.