DOCUMENT RESUME

ED 421 489 TM 028 448

AUTHOR Slack, Jill Berlin; St. John, Edward P.

TITLE A Model for Measuring Math Achievement Test Performance: A Longitudinal Analysis of Non-Transient Learners Engaged in a Restructuring Effort.

PUB DATE 1998-04-00
NOTE 38p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Acceleration (Education); Achievement Tests; Disadvantaged Youth; Elementary Education; *Elementary School Students; Evaluation Methods; Longitudinal Studies; *Mathematics Achievement; *Mathematics Tests; *Measurement Techniques; Models; Performance Factors; School Restructuring
IDENTIFIERS California Achievement Tests

ABSTRACT

This study investigated the mathematics achievement test performance of 62 non-transient elementary school learners in accelerated schools using a longitudinal design. Both the California Achievement Test (CAT) and the Louisiana Educational Assessment Program (LEAP) test were included in this investigation. In particular, this study sought to determine whether accelerated schools with distinct contextual features experienced significantly different test performances. A logistic regression was used to explore the relationship of several variables to the schools' performances. The variables were related to individual background, school environment, and curriculum and instruction factors. The researchers developed two logistic regression models to fit the uniqueness of the CAT and LEAP tests. Each model used a sequential analysis to examine the association of specific factors to test score improvement. The most consistent, significant finding across both models revealed that higher ability students were less likely to improve than lower ability students. This finding is consistent with the Accelerated Schools philosophy that "disadvantaged" students stand the most to gain from innovative teaching approaches. Additional findings showed the significant impact of age, gender, school environment, and curriculum and instruction on improvement. In particular, observations related to the latter factor revealed that students who were provided with math manipulatives/technology for longer periods were more likely to improve their standardized math scores than those who were provided with such instruction for shorter periods. (Contains 7 tables and 13 references.) (Author/SLD)

A Model for Measuring Math Achievement Test Performance: A Longitudinal Analysis of Non-Transient Learners Engaged in a Restructuring Effort

Presented at the Annual Meeting of the American Educational Research Association

San Diego, CA

April 16, 1998

by

Jill Berlin Slack
Southwest Educational Development Laboratory
Southeast Comprehensive Assistance Center

and

Edward P. St. John
University of Indiana

ABSTRACT

This study investigated the math achievement test performance of non-transient learners in accelerated schools using a longitudinal design. Both the California Achievement Test (CAT) and the Louisiana Educational Assessment Program (LEAP) test were included in this investigation.

In particular, this study sought to determine whether accelerated schools with distinct contextual features experienced significantly different test performances. A logistic regression was used to explore the relationship of several variables to the schools' performances. The variables were related to individual background, school environment, and curriculum and instruction factors.

The researchers developed two logistic regression models to fit the uniqueness of the CAT and LEAP tests. Each model used a sequential analysis to examine the association of specific factors to test score improvement.

The most consistent, significant finding across both models revealed that higher ability students were less likely to improve than lower ability students.

This finding is consistent with the Accelerated Schools philosophy that "disadvantaged" students stand the most to gain from innovative teaching approaches.

Additional findings showed the significant impact of age, gender, school environment, and curriculum and instruction on improvement. In particular, observations related to the latter factor revealed that students who were provided with math manipulatives/ technology for longer periods were more likely to improve their standardized math scores than those who were provided with such instruction for shorter periods.

Introduction

Researchers and practitioners have conducted several studies in recent years showing the benefits of disaggregated test score analysis (Bamburg & Medina, 1993; GAO, 1991; McCarthy & Still, 1993; Murphy & Schiller, 1992). To date, however, these studies have primarily concentrated on individual background characteristics such as race and socioeconomic class. Relatively few educational studies have focused on disaggregated data of non-transient students. Further, little research exists examining the influence of specific school practices on test score improvement.

The purpose of the present study was to conduct a longitudinal analysis with mathematics standardized test data of Louisiana students whose schools were engaged in the Accelerated Schools Project for at least a 4-year period to determine whether differences existed between their test score performances. A 4-year period was selected because research shows that students are more likely to experience test-score gains a few years after their schools have initiated and sustained the restructuring process (Hopfenberg et al., 1993; Schmoker, 1996).

Since the study involved longitudinal analysis of data of students in elementary schools over a 4-year period, non-transient sixth graders were considered the most viable participants. Consequently, 62 regular education students from three accelerated schools were included in the study because they remained in the same school for at least a 4-year period.

Due to the alternate use of two standardized achievement tests in Louisiana, the mathematics percentile scores from the California Achievement Test (CAT), and the raw scores from the math section of the Louisiana Educational Assessment Program (LEAP) were collected for two separate years. Consequently, the study included two sets of standardized test scores for each non-transient student. The percentile scores for the CAT were converted into normal curve equivalency scores (using a conversation table) to chart the progress students made on the tests.

The testing data was analyzed to investigate whether the three schools with distinct contextual features, as evidenced through profiles, experienced significant differences in math test performances. School variables, including student background, school environment and instructional approaches, comprised the contextual features of each school. A logistic regression was used to determine the influence of these variables on schools' test performance.

Literature Review

Testing has long been a focal issue of American education, but has recently demanded new attention as states and the federal government have increased both their investment and their attention to measuring educational achievement. In particular, since the publication of *A Nation at Risk* in 1983, politicians, educators, and the public have demanded greater efficiency and effectiveness to reform education (Toch, 1991).

With pressure emanating everywhere from the national level to the local level to improve school performance, educators have increasingly turned to standardized tests to prove the quality of their restructuring and reform efforts (Hymes, 1991). Consequently, "it is in terms of standardized test scores alone that the nation judges its schools" (Toch 1991, p. 206).

A concern flowing from this type of accountability is that test scores, particularly the way they are analyzed, do not tell the whole story. Unless analyzed with other contextual factors, they provide a narrow account of school progress (Schmoker, 1996). They also have limited influence on instruction (Haladyna, Nolen, & Haas, 1991).

Toch (1991) stated that score reports are so "full of obscurities and rife with abuses that it leads to a badly distorted picture of academic achievement" (p. 214). For example, and usually, school achievement relative to student outcomes is primarily based on aggregated test data (Murphy & Schiller, 1992).

Recent research has placed more emphasis on disaggregated test data to find out how specific groups of students are doing. These studies have primarily focused on differences between races, ages (GAO, 1991), and socioeconomic classes (McCarthy & Still, 1993); and have produced mixed results. For example, the latter study reported exceptional gains by low-socioeconomic students who attended schools engaged in a restructuring or reform effort, whereas the former study revealed no differences among races or ages of students who were enrolled in schools that participated in educational reform.

Bamburg and Medina (1993) also studied disaggregated data, but with an emphasis on teacher inquiry. They found that disaggregation of standardized test data increased the congruence between what teachers espoused and practiced; focused the instructional conversation; and established data-driven priorities.

Standardized tests, if thoughtfully analyzed have many possibilities for helping schools reveal both progress and areas that need improvement (Schmoker, 1996). The traditional orientation, which involves characterizing schools' academic achievement based on aggregated data does not consider many dynamics of the educational process, including school variation and school change.

Although methods to disaggregate data have been developed to provide a more accurate means of measuring achievement, they primarily focus on characteristics such as race, gender, age, and socioeconomic class. The research on disaggregated test data is scant outside these characteristics.

This study extends the research on longitudinal analyses of disaggregated data by suggesting that non-transience and certain school variables, including school environment, and curriculum and instructional practices, influence standardized test performance. The process of using these variables to study the test performance of non-transient students and the reasons for their growth or lack of progress will help schools, especially those involved in a restructuring or reform effort, understand the dynamics of their situation better.

Methodology

The purpose of this study was the longitudinal investigation of mathematical achievement test data of non-transient students and their schools in order to determine whether schools with distinct contextual features experience different test score performances. Two standardized achievement tests had to be used in this study because of the alternate use of the tests in the state in which the schools involved in this study reside. They included the California Achievement Test (CAT), a norm-referenced test, and the Louisiana Educational Assessment Program (LEAP), a criterion-referenced test.

Logistic regression was used to compare the test performances of the schools involved in this study. Only the non-transient students' test scores were considered in this investigation. Other school variables, such as individual background data and teachers' instructional approaches were also considered.

To ease this investigation, profiles of the participating schools were developed using information from school fact sheets and records; and teacher surveys of instructional approaches used with participants.

Three accelerated schools volunteered to take part in this study. All three are located in the Southeastern region of the United States and serve elementary students with similar socioeconomic status (SES) backgrounds. The schools are identified as X, Y, and Z.

Sixty-two, sixth grade, regular education students from these three schools were selected as participants for the study because they were classified as "non-transient," that is, they remained in the same school for at least a 4-year period. Fourteen out of 27 sixth graders in School X, 37 out of 59 sixth graders in School Y, and 11 out of 72 sixth graders in School Z qualified as non-transient students. Table 1 provides specific details related to each school's participants.

Participants' first and fourth grade CAT results, and third and fifth grade LEAP results were used in this investigation. The CAT was not administered to most of the participants in the second grade and, therefore, could not be used for the purposes of this study. The first grade CAT and third grade LEAP scores served as the baseline data.

Data were also collected from the second through fifth grade teachers on the instructional approaches they used on a daily basis with the participants. Cumulative folders were used to trace the number of years participants received a particular instructional approach.

The dependent variables for this study were the non-transient students' CAT and LEAP test scores in math. The independent variables focused on individual background, school environment (the school), and the number of year participants received each of five instructional approaches. The individual background

Table 1

Specific Details Relative to Each School's Participants

	School		
Participants	X	Y	Z
Number	14	37	11
% Black	100	100	82
% White			09
% Hispanic			09
% Female	50	62	82
% Retained	71	22	00
% Title 1	28	27	09

variables included age, gender, retention information, and base test scores. Initially, race was considered. However, since 96% of the participants were Black, this variable was omitted.

The instructional approaches included thematic units/teaching, multicultural education, tutoring, math manipulatives/technology, and Title I. The number of years participants were engaged in a particular approach ranged between zero and four. Notably, Title I was a pull out program for the first three of the four years included in this study. In the last year, Title I became a schoolwide program (i.e., students received Title I services in their regular classrooms). Nonetheless, the Title I program provided participants with math instruction for all four years.

The dependent (outcome) variables were dichotomous with "0" representing no test score growth and "1" representing at least one point of test score growth between the first and fourth grade CAT, and third and fifth grade LEAP. One point was selected as the threshold for improvement to account for the students with high base scores and less room for improvement.

Most of the independent variables were also dichotomous either directly or through design sets. For example, the number of years participants received Title I services represented a continuous variable (directly) grouped into three levels of dichotomous variables (a design set): one year, two years, and three years. The participant either fell into a specific category (which was coded as a "1") or did not (which was coded as a "0"). Table 2 provides a listing of the specific coding for each variable.

Because the dependent variables in this analysis involved non-continuous outcome measures, ordinary least squares regressions (OLS) were not considered appropriate statistical procedures (Aldrich & Nelson, 1984). Instead, logistic regression was employed to examine the relationship of specific variables to outcome measures (i.e., test score improvement).

Two logistic regression models were developed by the researchers--one for each dependent variable: CAT Math and LEAP Math. Each model used a sequential analysis, the stepping in of factors, to examine the relationship of specific factors to math test score improvement. The first step included only individual background variables. The second step added the school

Table 2

Variable Coding

Individual Variables		
Factor/Variable	**Coding**	**Comment**
Individual Background		
Age	Continuous	Indicates age
Gender	0 = Male 1 = Female	Compares females to males
Retention "once"	0 = No 1 = Yes	Compares students who were retained once with students who were never retained
Retention "twice"	0 = No 1 = Yes	Compares students who were retained twice with students who were never retained
Base Scores	Continuous	Indicates Grade 1 CAT and Grade 3 LEAP scores
School Environment		
School X	0 = No 1 = Yes	Compares School X with School Y
School Z	0 = No 1 = Yes	Compares School Z with School Y

(table continues)

Instructional Approach		
Thematic Units/Teaching (3 years)	0 = No 1 = Yes	Compares 3 years with 1 year
Thematic Units/Teaching (4 years)	0 = No 1 = Yes	Compares 4 years with 1 year
Multicultural Education (2 years)	0 = No 1 = Yes	Compares 2 years with 4 years
Multicultural Education (3 years)	0 = No 1 = Yes	Compares 3 years with 4 years
Tutoring (3 years)	0 = No 1 = Yes	Compares 3 years with 0 years
Tutoring (4 years)	0 = No 1 = Yes	Compares 4 years with 0 years
Math Manipulatives/ Technology (3 years)	0 = No 1 = Yes	Compares 3 years with 4 years
Title I (1 year)	0 = No 1 = Yes	Compares 1 year with 0 years
Title I (2 years)	0 = No 1 = Yes	Compares 2 years with 0 years
Title I (3 years)	0 = No 1 = Yes	Compares 3 years with 0 years

environment variables. The third step added the instructional approach variables.

The logic of each factor (individual background, school, and instructional approaches) was examined using two types of statistics. First, the changes in the delta Ps (probability measures) and the significance levels of each independent variable were reviewed. Second, the changes in the model statistics, particularly the log likelihood function (-2 Log L), pseudo R^2, and percent correctly predicted were compared across versions. The delta Ps were calculated using a formula recommended by Peterson (1984):

$$F(P) = \exp(L_1)/[1+\exp(L_1)]-\exp(L_0)/[1+\exp(L_0)]$$

where $L_0 = \ln p/(1-p)$ (p = baseline probability, ln = natural logarithm) and $L_1 = L_0 + \text{Beta}$.

The delta P statistics were used in two ways in this investigation. First, for the dichotomous variables, the delta P provided a measure of probability on which a specific independent variable was likely to change the dependent variable. Second, for continuous variables, the delta P was interpreted as meaning that a change in unit measure altered the probability of the outcome by a specific percentage.

Besides the delta Ps, a pseudo R^2 and other model statistics such as the log likelihood function (-2 Log L) and the percent correctly predicted were applied to determine whether the model improved as variables were stepped in. An increase in the pseudo R^2 from one series over the last was interpreted to mean a reduction in unexplained error. Further, an increase in the percent correctly predicted and a decrease in the -2 Log L from one step to the next was interpreted as meaning an improvement in the overall predictability and fit of the model, respectively.

Findings

This longitudinal study investigated the achievement test performance of non-transient students in order to determine whether differences existed in the test performances of three accelerated schools with distinct contextual features. Logistical regression was used to explore the relationship of several variables to the schools' achievement test performances. The variables were related to individual background, the school environment, and the instructional approaches used at the school. Table 3 presents descriptive statistics for the variables by

Table 3

Participant Demographics

School (Environment)	X	Y	Z
Individual Background			
Age			
10	21%	38%	45%
11	36%	62%	55%
12	36%	0%	0%
13	7%	0%	0%
Gender			
Male	50%	38%	18%
Female	50%	62%	82%
Retention "Once"	50%	22%	0%
Retention "Twice"	21%	0%	0%
Base Test Scores			
Cat Math Grade 1	43.0	36.0	62.0
LEAP Math Grade 3	358.0	361.0	364.0
Curriculum & Instruction Innovations (Teaching Approaches)			
Thematic Units/Teaching			
1 year	100%	0%	0%
3 year	0%	8%	91%
4 years	0%	92%	9%
Multicultural Education			
2 years	0%	0%	82%
3 years	100%	8%	18%
4 years	0%	92%	0%

(table continues)

School (Environment)			
Curriculum & Instruction Innovations (Teaching Approaches)			
Tutoring			
0 Years	0%	0%	100%
3 years	100%	5%	0%
4 years	0%	95%	0%
Math Manipulatives/ Technology			
3 years	100%	0%	91%
4 years	0%	100%	9%
Title I			
0 years	72%	76%	91%
1 year	0%	16%	0%
2 years	14%	8%	9%
3 years	14%	0%	0%

school. The statistics reveal the distinctiveness of the schools, especially with the teaching approaches.

Two logistic regression models were developed by the researchers to enhance the findings of the study. The dependent variables (i.e., participants' CAT and LEAP test performances) for both models were coded as dichotomous outcomes, with "0" = no improvement and "1" = at least one point of improvement. One point was selected as the threshold to allow for the participants with high base scores and smaller margins for improvement.

Table 4 presents the population with test score improvement broken down by the amount of improvement. As the table reveals, of the participants who experienced improvement, less than 5% fell into the "+1" point category and at least 80% fell into the "≥5" category for both tests. Moreover, most of those with lower base scores (representing the bottom half) fell into the "≥5" category, and most of those with higher base scores (representing the top half) fell into the "+1" to "+4" categories. This suggests that a one point threshold for improvement rather than two or more point threshold may have strengthened the design of the

Table 4

Population with Test Score Improvement Broken Down by Amount of Improvement

	CAT	LEAP
Point Category	Math	Math
	%	%
+1	4.0	4.0
+2	0.0	4.0
+3	4.0	4.0
+4	12.0	0.0
≥5	80.0	88.0

logistic regression model, allowing for a better comparison between the growth of the participants with high base scores and those with low base scores.

Like the dependent variables, most of the independent variables were also coded as dichotomous variables, with 1 = yes, and 0 = no. Of the individual background variables, females were compared with males, and students who were retained one or two years were compared with students who were never retained. Age and base test scores were continuous variables.

Due to the unbalanced number of participants representing the schools (14 in School X, 37 in School Y, and 11 in School Z), the logistic regression models were designed to compare Schools X and Z with School Y. Additionally, step three of the models (that included the instructional variables) was designed to compare each group of participants who received less than four years of multicultural education and math manipulatives/technology with the group of participants who received four years.

To counteract redundancies in the design matrix, the groups of participants who received three and four years of thematic units/teaching and tutoring were compared with the group of participants who received the fewest years. Finally, each group of participants who received Title I services was compared with those who did not receive such services.

Table 5 presents the results from the sequential analysis for the CAT Math Model. In step one, which examined the effect of the individual background factor on improvement, three variables had a significant, negative association with the dependent variable: Female, two years of retention, and CAT 1 Math scores.

Being a female decreased the probability of improvement by 28.3 percentage points ($p \leq .10$). Students who were retained two years were 43.7 percent less likely to improve than the students who were never retained ($p \leq .10$). Students with higher ability (higher base math scores) were 2.1 percent less likely to improve their fourth grade math scores than students with lower ability (lower base math scores) ($P \leq .01$).

In step two, when the school factor was added, being a female and/or a student who was retained for two years were no longer significant. However, age became significant; older students were 35.8 percent more likely to improve per year of age ($P \leq .10$). In

Table 5

Logistic Analysis of Test Score Improvement - CAT Math Model

Factor/Variable	Step 1 Delta P	Step 2 Delta P	Step 3 Delta P
Individual Background			
Age	0.290	0.358*	0.450*
Female	-0.283*	-0.128	0.089
Retention			
1 year	-0.089	0.450	0.450
2 years	-0.437*	0.448	-0.286
CAT 1 Math Scores	-0.021***	-0.022***	-0.047**
School Environment			
School X		-0.450	-0.448
School Z		-0.152	0.450
Instructional Approaches			
Thematic Units/Teaching			
3 years			-0.450
4 years			0.450
Multicultural Education			
2 years			0.450
3 years			0.450
Tutoring			
3 years			-0.450
4 years			0.450
Math Manipulatives/Tech.			
3 years			-0.450*
Title I			
1 year			0.450
2 years			-0.450
3 years			0.450
-2 Log L	45.240	35.434	17.581
Pseudo R^2	0.300	0.452	0.728
% Correctly Predicted	76.60	80.85	91.49
Model Chi-Square	19.383	29.189	47.042
Goodness of Fit	43.427	40.982	15.447

* = 0.10 level of significance
** = 0.05 level of significance
***= 0.01 level of significance

addition, the CAT 1 Math variable remained significant and negative at the 0.01 level ($p \leq .01$), with virtually no change in size. The changes in step two suggest an interaction between gender or two years of retention, and the school environment factor. Further, the changes demonstrate a relationship between the student's age and his or her school.

When the instructional factor was added in step three, age and base test scores remained significant. Age resulted in a 45.0 percent increase in the probability of improvement for each year of age ($p \leq .10$), a change from step two of the model. For the students who had higher math ability as measured by the first grade CAT, the result was a 4.7 decrease in probability of improvement ($p \leq .05$), also a change from step two.

Only one variable related to instruction was significant in step three. This involved students who had three years of math manipulatives/technology. These students were 45.0 percent less likely to improve than the students who had four years of math manipulatives/technology ($p \leq .10$).

Table 6 presents the results of the sequential analysis for the LEAP Math Model. In the first step,

Table 6

Logistic Analysis of Test Score Improvement - LEAP Math Model

Factor/Variable	Step 1 Delta P	Step 2 Delta P	Step 3 Delta P
Individual Background			
Age	0.087	0.217	0.292
Female	-0.055	-0.061	-0.004
Retention			
1 year	-0.174	-0.099	0.247
2 years	-0.097	0.200	0.550
LEAP 3 Math Scores	-0.022***	-0.029***	-0.031***
School Environment			
School X		-0.500***	-0.465
School Z		-0.493***	0.285
Instructional Approaches			
Thematic Units/Teaching			
3 years			0.550
4 years			-0.549
Multicultural Education			
2 years			-0.471
3 years			0.550
Tutoring			
3 years			-0.550
4 years			0.550
Math Manipulatives/Tech.			
3 years			-0.544
Title I			
1 year			0.048
2 years			-0.469*
3 years			-0.549
-2 Log L	68.347	52.487	46.896
Pseudo R^2	0.143	0.342	0.412
% Correctly Predicted	65.52	75.86	79.31
Model Chi-Square	11.436	27.296	32.887
Goodness of Fit	56.847	60.106	50.449

* = 0.10 level of significance
***= 0.01 level of significance

only the LEAP 3 Math scores were found to be significant. Students who had higher ability, as measured by the third grade LEAP, were 2.2 percent less likely to improve their fifth grade math scores than students with lower ability ($p \leq .01$).

In the second step, the LEAP 3 Math variable remained significant and negative. Students with higher base scores were 3 percent less likely to improve than students with lower base scores ($p \leq .01$). No other individual background variable was significant.

However, both school variables were significant at the 0.01 level ($p \leq .01$). The students in School X were 50 percent less likely to improve, and the students in School Z were 49.3 less likely to improve, than the students in School Y.

In step three, the LEAP base scores continued to be significant. An interesting change, however, was that the school environment variables were no longer significant. This step suggests an interaction between the schools and the instructional approaches. Further, this step demonstrates that students who received two years if Title I services were 47 percent less likely to improve than students who never received Title I services ($p \leq .10$).

With the inclusion of each factor in both models, the -2 Log L decreased. Further, the pseudo R^2 increased and the percent correctly predicted improved. These statistics indicate that the stepping in of each factor (individual background, school environment, and instructional innovations) improves the fit and overall predictability of the models.

Conclusion

Over the past decade, achievement testing has assumed an increasingly prominent position in educational reform (Schmoker 1996). Specifically, since a *Nation at Risk*, educators have used test results to plan and prove the quality of their restructuring efforts (Hymes, 1991; Toch, 1991).

At present, gains or declines in score patterns for aggregated units are mostly used as a standard by which the public judges the academic success of their schools (Murphy & Schiller, 1992). Although disaggregated test score analysis is becoming more prevalent, studies thus far primarily focus on individual characteristics such as race and socioeconomic class. This study extends the research on longitudinal analysis of disaggregated data by investigating whether non-transience and specific school variables such as instructional approaches impact test performance.

Several interesting findings resulted from this study. Perhaps the most striking finding across both models was that higher ability students were invariably less likely to improve than lower ability students (that is, those students who scored at or below the 30^{th} percentile on the first grade CAT, and at or below 350--the bottom third--on the third grade LEAP).

Table 7 provides further evidence of the benefits of non-transient students engaged in a reform effort (i.e., Accelerated Schools) by presenting the percent of low ability students experiencing improvement. As the table reveals, 70 percent of the low ability students improved their CAT math scores by at least 10 percentile points. Similarly, 70% percent of the low ability students improved their LEAP math scores by at least 10 points. This finding is consistent with the philosophy that academically disadvantaged students stand the most to gain from restructuring efforts focused on student learning (Levin, 1987; Newmann and Wehlage, 1995).

Another important finding focuses on the math

Table 7

Percent of Low Ability Students Experiencing Improvement

Points Improved	CAT Math	LEAP Math
>=5	85.7	80.0
>=10	70.5	70.0

Note. Low ability refers to those students who scored at or below the 30th percentile on the first grade CAT and at or below 350 on the third grade LEAP.

manipulatives/technology approach. A longer engagement in this approach was found to increase the probability of improvement significantly on the CAT.

Additionally, the Title I approach and the school environment factors were found to have a significant impact on the LEAP. These findings do not suggest that the other instructional approaches in this study were not effective at improving test scores. Rather they demonstrate that a one to two year differential in the amount of exposure to the instructional approaches may have a significant impact on improvement.

The findings associated with the school environment factor in the LEAP Math Model contain powerful messages for educators. First, schools have a significant impact on criterion-referenced test performance. That is, what goes on in the daily operations of the school influences test scores. And second, schools with distinct features may experience significantly different test performances, as found in this study. However, if the instructional approaches are comparably innovative and similarly implemented, the differences in their test performances become insignificant. The fact that the school environment variables were significant in step two of the LEAP Model but were not significant in step three gives support to this position and provides further evidence that what goes on in the overall occurrences of the school impacts test performance.

This study fills a void in the current literature concerning the effects of school restructuring on test performance by focusing on the individual background of non-transient students, their school environment, and their engagement with certain instructional approaches. Further, it suggests a need for reexamining district and state policies that judge, rank, and compare schools based on limited aggregated test data.

The models proposed in this study are both practical and workable, and offer policymakers, school officials, and faculty members insight on innovative ways of measuring school progress.

They also allow schools to determine whether, and under what conditions, school environment and instructional variables have an impact on test performance. This innovative orientation of measuring school progress can provide a crucial missing link for helping schools, especially those engaged in a restructuring effort, learn more about themselves.

References

Aldrich, J. H., & Nelson, F. D. (1984). Linear probability, logit and probit models. Beverly Hills, CA: Sage Publications.

Bamburg, J., & Medina E. (1993). Analyzing student achievement: Using standardized tests as the first step. In J. Bamburg (Ed.), Assessment: How do we know what they know? Dubuque, IA: Kendall-Hunt.

General Accounting Office. (1991) The effect of education reform on student achievement: GAO Report. Urban Education, 26, 160-175.

Haladyna, T. M., Nolen, S. B., & Haas, N. (1991). Raising standardized achievement test scores and The origins of test score pollution. Educational Researcher, 20(5), 2-7.

Hopfenberg, W. S., Levin, H. M. Chase, C., Christensen, G., Moore, M., Soler, P., Brunner,. I., Keller, B., & Rodriguez, G. (1993). Accelerated schools resource guide. San Francisco: Jossey-Bass.

Hymes, D. L. (1991). The changing face of testing and assessment: Problems and solutions. Arlington, VA: American Association of School Administrators.

Levin, H. (1987). Accelerated schools for disadvantaged students. Educational Leadership 44(6), 19-21.

McCarthy, J., & Still, S. (1993). Hollibrook accelerated elementary school. In Joseph Murphy & P. Hallinger (Eds.), Restructuring schooling (pp. 63-83). Newbury Park, CA: Corwin Press.

Murphy, J., & Schiller, J. (1992). Tools for diagnosis, prescription, and accountability. In J. Murphy & J. Schiller (Eds.), Transforming America's schools (pp. 187-224). LaSalle, IL: Open Court.

Newmann, F. M., & Wehlage, G. G. (1995). Successful school restructuring: A report card to the public and educators. Madison, WI: Center on Organization and Restructuring of Schools.

Peterson, T. (1984). A comment on presenting results of logit and probit models. American Sociological Review, 50,(1), 130-131.

Schmoker, M. (1996). Results: The key to continuous school improvement. Alexandria, VA: Association for Supervision and Curriculum Development.

Toch, T. (1991) The coins of realm: The paradox of

Standardized testing. In T. Toch (Ed.), In the name of excellence (pp. 205-232). New York: Oxford University Press.

ERIC® Clearinghouse on Assessment and Evaluation

University of Maryland
1129 Shriver Laboratory
College Park, MD 20742-5701

Tel: (800) 464-3742
(301) 405-7449
FAX: (301) 405-8134
ericae@ericae.net
http://ericae.net

March 20, 1998

Dear AERA Presenter,

Congratulations on being a presenter at AERA[1]. The ERIC Clearinghouse on Assessment and Evaluation invites you to contribute to the ERIC database by providing us with a printed copy of your presentation.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are gathering all the papers from the AERA Conference. We will route your paper to the appropriate clearinghouse. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality. You can track our processing of your paper at http://ericae.net.

Please sign the Reproduction Release Form on the back of this letter and include it with **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can drop off the copies of your paper and Reproduction Release Form at the **ERIC booth (424)** or mail to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to:

AERA 1998/ERIC Acquisitions
University of Maryland
1129 Shriver Laboratory
College Park, MD 20742

This year ERIC/AE is making a **Searchable Conference Program** available on the AERA web page (http://aera.net). Check it out!

Sincerely,

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

---

[1]If you are an AERA chair or discussant, please save this form for future use.

CUA

The Catholic University of America