DO NCTM REFORM PRACTICES MATTER? PREDICTING PRECOLLEGE STUDENT MATHEMATICS ACHIEVEMENT USING NAEP 2000 DATA

Mack C. Shelley, II, *Iowa State University*
Sarah T. Lubienski, *Iowa State University/University of Illinois, Champaign-Urbana*
Mack C. Shelley, II, 323 Snedecor Hall, Iowa State University, Ames, Iowa 50011-1210

KEY WORDS: Hierarchical linear models, imputation, numeracy

Concerns about equity in educational outcomes have taken on new urgency, amid concerns about recently increasing achievement gaps between white and black students (Campbell et al., 2000; Lee, 2002; Lubienski & Bowen, 2000). This research focuses on students' achievement and learning experiences in mathematics, central to policy objectives of developing a technologically literate work force and citizenry (Chubin & Pearson, 2001; National Science Foundation, 2000).

Family socioeconomic (SES) differences account for much of race-related achievement gaps (Jencks & Phillips, 1998), along with teacher expectations, school structure, student motivation, and student resistance (Banks, 1988; Ferguson, 1998; Steele & Aronson, 1998). In-depth examination of how achievement gaps vary by mathematical strand and disparities in students' access to effective mathematics instruction may illuminate underlying causes of achievement gaps.

Equity and mathematics achievement are related (Mitchell et al., 1999; Silver & Kenney, 2000). Race-based differences in students' beliefs and classroom experiences are related to mathematics achievement gaps, but such differences might be due to SES more than race (Strutchens & Silver, 2000). Although overall student performance varies only a few points by strand, significant and persisting variations exist across strands (Lubienski, 2001).

As student achievement has increased recently, some reform-based practices, such as unrestricted classroom use of calculators, have expanded dramatically (Braswell et al., 2001). Given Lubienski's finding that white, high-SES students were more likely than their less privileged counterparts to have unrestricted calculator access and to believe mathematics transcends memorization, race and SES may be confounding variables. NAEP data can inform the current debate, particularly because such confounding variables may be controlled. Our analysis endeavors to determine which selected reform-oriented instructional practices correlate positively with mathematics achievement, controlling for race, SES, and school context.

Estimating the effects of SES, race, and gender on mathematics achievement is complicated by NAEP's multiple-stage cluster sampling design, unequal selection probabilities associated with stratification and over-sampling of certain subpopulations, and measurement error associated with the matrix sampling scheme (Johnson, 1992; Johnson & Rust, 1992). Using item response theory, proficiency scores are estimated for each student based on their survey responses and other characteristics. Five plausible values are drawn randomly from the conditional distribution of proficiency scores for each student; results of separate analyses for the five plausible values then are synthesized (Mislevy et al., 1992; Rubin, 1987).

This study attempts to identify variables related to reform-based instruction that correlate with achievement, focused on race, SES, and gender differences. For example, students with greater access to calculators in the mathematics classroom scored higher on the NAEP mathematics assessment. We attempt to determine whether this correlation simply means that more white, high-SES students use calculators, or if calculator use does correlate with achievement after controlling for student characteristics including race, gender, and SES, and school context.

Data Analysis—HLM

Two-level hierarchical linear models (HLM) (Raudenbush & Bryk, 2002; Raudenbush et al., 2001) are used to estimate parameters for models with students (at level 1) nested within schools (level 2), to assess the contribution of reform-based instructional practices to explaining variation in mathematics achievement for students controlling for race, ethnicity, SES, gender, how often a language other than English is spoken at home, disability status, and whether the student is taking appropriate mathematics courses.

HLM resolves multilevel data problems of aggregation bias and imprecision (Bryk & Raudenbush, 1987, 1992; Goldstein, 1987; Raudenbush & Bryk, 2002), and is used widely for policy analysis and evaluating school effectiveness (Arnold, 1992; Bryk & Raudenbush, 1989; Gamoran et al., 1997; Lee & Bryk, 1989; Newman et al., 2001; Phillips, 1997; Raudenbush & Bryk, 1986).

The HLM2 procedure in HLM 5.04 software produces correct estimates for the "plausible values" configuration of the NAEP data structure, using the multiple, model-based imputation strategy for data missing at random (Little & Rubin, 1987; Rubin, 1987; Schafer, 1997). Parameter estimates are averaged from separate analyses of the five plausible values.

Data and Variables

The data source is a restricted-use CD ROM, provided by the National Center for Education Statistics, the containing 2000 data.

Level 1 and Level 2 data sets were merged using SPSS, based on the scrambled primary sampling unit and school code (SCRPSU).

The dependent variable (MEANACH) measures mathematics achievement of students in the 8th grade. It is calculated as the mean of student scores across five strands (number/operations, geometry, measurement, data analysis, and algebra/functions).

The Level 1 student weight (WEIGHT) is the across sample adjusted student weight. The Level 2 school weight (NEWSCHWT) is the product of five separate weights scaled down to the sample size.

The Level 1 base predictor variables, which appear in each equation, are the following student characteristics:

BLACK, coded black = 1, else = 0.

HISPANIC, coded 1 = Hispanic, else = 0.

STUD_ SES, student socioeconomic status, measured as the standardized Z-score value of the natural logarithm of the scores for the following eight variables that were combined into a single variable: 4 types of reading material in students' homes (newspapers, magazines, books, and encyclopedia, combined by NAEP into a single variable, with categories of 0-2 items, 3 items, or 4 items), computer and internet access at home, extent to which studies are discussed at home, school lunch and Title 1 eligibility, and education level of mother and father.

DSEX, student gender, measured as 1 = male, 0 = female.

ENGLISH, measuring how often a language other than English is spoken in the home.

DISABIL1, measuring whether the student has a non-orthopedic disability (coded 1) or not (coded 0).

HSMATH, coded 1 if the student is taking first-year algebra, geometry, second-year algebra, or an integrated math course, and 0 otherwise.

The Level 2 school characteristic predictor variables are:

ZSCH_SES, school SES. A school administrator at each sampled school completed a survey requesting information about the school, including the percentage of students qualifying for Title 1 funds and free/reduced lunch. For the purposes of this study, these two variables were averaged to produce a single school-SES variable, after which a natural logarithm was calculated and transformed into a Z-score. This variable speaks to the stratification of schools by student-level SES, in that remarkably few high-SES students of any race were in low-SES schools, and conversely, very few low-SES students of any race attended high-SES schools. School SES is measured so that larger values indicate a school with a greater array of problems, unlike student SES, for which larger values indicate a stronger family resource base.

SCH_RACE, school race, measured as a Z-score transformation of the square root of 100 times the percentage of students in the school who are identified as White or Asian.

PRIVAT_1, coded 1 = private school, 0 = else.

The major thrust of this analysis was to assess the impact of reform practices for teaching mathematics to precollege students. For each model, one of these following reform predictor variables was added to the set of predictors:

TCALC01 is coded 0 if a teacher reported calculator use below average and 1 if at or above average.

TMANIP01 is coded 1 if the teacher reported that students work with objects like rulers daily or weekly, and 0 if less frequent.

TMULTYR1 is coded 1 if multiple choice assessments are used never or just 1-2 times per year, and 0 if more frequent.

TMULTWK1 is coded 1 if a teacher reported the use of multiple-choice assessments once or twice a week, and 0 if monthly or yearly.

TEAREA01 is coded 1 if there is heavy teacher emphasis on reasoning, and 0 otherwise

TTRAD01 is coded 0 if there is heavy teacher-reported emphasis on facts and concepts, and skills for routine problem solving, and 1 otherwise.

TCOLPS is a Z-score of the square root of five composite variables measuring student interaction in doing their work.

NCTMVERY is coded 1 if the teacher is very knowledgeable about NCTM and 0 otherwise.

NCTMNOT is coded 1 if the teacher has little or no knowledge about NCTM, and 0 otherwise.

TEAWRITE is an index of teacher-reported writing by students in four activities.

TEACURR is coded such that a higher value means that teacher reports more emphasis on geometry, measurement, and data.

YEARSTAU is the number of years a teacher has taught at the elementary or secondary level.

STUBELIE is coded such that a higher value indicates a teacher who has students go beyond memorization and shows them more than one way to solve a problem.

STWRITPS is coded such that higher values indicate that a teacher has students write more often about solving a math problem.

STMANIP is coded such that higher values indicate more frequent use of measuring instruments and geometric solids.

STCALC is the result of a factor analysis, coded such that a higher score means that students more frequently use calculators for classwork, homework, tests, and math in school.

STHITECH is derived from a factor analysis and is coded such that a higher score means more frequent use of graphing calculators and symbol manipulators.

STCOLLAB is a factor coded such that higher scores imply that students more frequently solve problems with partners or groups, and talk with class about math problems.

Summary of Findings

Table 1 reports the average ordinary least squares (OLS) Level-1 regression coefficients, measuring the estimated change in 8^{th} grade MEANACH for one-unit increases in each reform initiative. These results are extracted from the separate equations, and represent the partial effect of each reform initiative, net of the base predictors in each equation. Several of these estimated effects are negative, possibly reflecting their use in low-achieving classes. For the most part, only modest effects are suggested.

Table 2 summarizes the effect of Level-2 variables on student math achievement, mediated through each of the reform initiatives. Robust standard errors are used, which makes it difficult for the coefficients to attain statistical significance. There are significant effects in three instances for school race and in three instances for private vs. public control. School SES is never significant.

Significance of the Research

The findings from this study can illuminate existing mathematics achievement disparities, raise questions about current policies and reforms, suggest potentially beneficial policy directions, and pinpoint areas needing further study. Implementation disparities may explain existing achievement gaps and should inform future instruction reform efforts (Newmann et al., 2001). We attempt to identify which reform-oriented instructional practices correlate with achievement for disadvantaged students, and which practices have not been implemented widely.

Given that the National Council of Teachers of Mathematics reforms are intended to help "all students" gain mathematical ability, researchers must address whether and how reform-based instructional practices affect students previously marginalized in mathematics. Mathematics is a gatekeeper to jobs in which workers are rewarded equitably. For example, women earn less than 70 cents for each dollar earned by males with comparable education, but this difference is negated when women take at least eight credits of college mathematics (American Association of University Women, 1992).

When talents remain untapped, society is deprived of workers with strong mathematical backgrounds, and students are not prepared to be critical thinking, mathematically literate citizens. By looking in depth at how achievement gaps vary by mathematical strand, and by examining disparities in students' access to effective mathematics instructional practices, the proposed research can shed new light on underlying causes of achievement gaps.

This study can help policymakers and educators create equitable classrooms, which ultimately will enhance life chances for minority, low-SES, and female students entering higher education and employment. The scope and representativeness of NAEP data can lend important evidence to inform policy debates.

References

AAUW. (1992). *How schools shortchange girls.* Wellesley, MA: Wellesley College Center for Research on Women.

Arnold, C.L. (1992). An introduction to hierarchical linear models. *Measurement and Evaluation in Counseling and Development, 25,* 58-90.

Banks, J.A. (1988). Ethnicity, class, cognitive, and motivational styles: Research and teaching implications. *Journal of Negro Education, 57*(4), 452-466.

Braswell, J.S., Lutkis, A.D., Grigg, W.S., Santapau, S.L., Tay-Lim, B., & Johnson, M. (2001). *The Nation's report card: Mathematics 2000.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Bryk, A.S., & Raudenbush, S.W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101,* 147-158.

Bryk, A.S., & Raudenbush, S.W. (1989). Toward a more appropriate conceptualization of research on school effects: A three level hierarchical linear model. In R.D. Bock (Ed.), *Multilevel analysis of educational data.* San Diego: Academic Press.

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

Campbell, J.R., Hombo, C.M., & Mazzeo, J. (2000). *NAEP 1999 trends in academic progress: Three decades of student performance, NCES 2000.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Chubin, D.E., & Pearson, W., Jr. (Eds.). (2001). *Scientists and engineers for the new millennium: Renewing the human resources.* Washington, DC: Commission on Professionals in Science and Technology.

Ferguson, R.F. (1998). Teachers' perceptions and expectations and the black-white test score gap. In C. Jencks & M. Phillips (Eds.), *The black-white test score gap* (pp. 273-317). Washington, DC: Brookings Institution Press.

Gamoran, A., Porter, A.C., Smithson, J., & White, P.A. (1997). Upgrading high school mathematics instruction: Improving learning opportunities for low-achieving, low-income youth. *Educational Evaluation and Policy Analysis, 19*(4), 325-338.

Goldstein, H. (1987). *Multilevel models in educational and social research.* New York: Oxford University Press.

Jencks, C., & Phillips, M. (Eds.). (1998). *The black-white test score gap.* Washington, DC: Brookings Institution Press.

Johnson, E.G. (1992). The design of the Nation Assessment of Educational Progress. *Journal of Educational Measurement, 29*(2), 95-110.

Johnson, E.G., & Rust, K.F. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics, 17*(2) 175-190.

Lee, V.E., & Bryk, A.S., (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education, 62,* 172-192.

Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity? *Educational Researcher, 31*(1), 3-12.

Little, R., & Rubin, D. (1987). *Statistical analysis with missing data.* New York: Wiley.

Lubienski, S.T., & Bowen, A. (2000). Who's counting? A survey of mathematics education research 1982-1998. *Journal for Research in Mathematics Education, 31*(5), 626-633.

Lubienski, S.T. (2001). *Are the NCTM Standards reaching all students? An examination of race, class, and instructional practices.* Paper presented at the American Educational Research Association, Seattle.

Mislevy, R.J., Johnson, E.G., & Muraki, E. (1992). Scaling procedures in NAEP. *Journal of Educational Statistics, 17*(92), 131-154.

Mitchell, J.H., Hawkins, E.F., Jakwerth, P., Stancavage, F.B., & Dossey, J.A. (1999). *Student work and teacher practices in mathematics.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

National Science Foundation, Committee on Equal Opportunities in Science and Engineering. (2000). *2000 biennial report to the United States Congress.* Arlington, VA: National Science Foundation.

Newmann, F.M., Smith, B., Allensworth, E., & Bryk, A.S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis, 23*(4), 297-321.

Phillips, M. (1987). What makes schools effective? A comparison of the relationships of communitarian climate and academic climate to mathematics achievement and attendance during middle school. *American Educational Research Journal, 34*(4), 634-662.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S., Bryk, A., Cheong, Y.F., & Congdon, R. (2001). *HLM5: Hierarchical linear and nonlinear modeling.* Lincolnwood, IL: Scientific Software International, Inc.

Raudenbush, S.W., & Bryk, A.S. (1986). A hierarchical model for studying school effects. *Sociology of Education, 59,* 1-17.

Rubin, D. (1987). *Multiple imputation for nonresponse in surveys.* New York: Wiley.

Schafer, J. (1997). *Analysis of incomplete multivariate data.* London: Chapman & Hall.

Silver, E.A., & Kenney, P.A. (Eds.). (2000). *Results from the seventh mathematics assessment of the national assessment of educational progress.* Reston, VA: National Council of Teachers of Mathematics.

Steele, C.M., & Aronson, J. (1998). Stereotype threat and the test performance of academically successful African Americans. In C. Jencks & M. Phillips (Eds.), *The black-white test score gap* (pp. 401-428). Washington, DC: Brookings Institution Press.

Strutchens, M.E., & Silver, E.A. (2000). NAEP findings regarding race/ethnicity: Students' performance, school experiences, and attitudes and beliefs. In E.A. Silver & P.A. Kenney (Eds.), *Results from the seventh mathematics assessment of the national assessment of educational progress.* Reston, VA: National Council of Teachers of Mathematics.

Table 1. Average OLS Level-1 Coefficients, 8th Grade, Estimated Change in MEANACH Attributable to Each Reform Initiative

Model	Teaching Practice	Average Coefficient
A	TCALC01	-29.88193
B	TMANIP01	-0.22052
C	TMULTYR1	15.84238
D	TMULTWK1	14.91274
E	TEAREA01	-94.45345
F	TTRAD01	-48.79070
G	TCOLPS	33.84481
H	NCTMVERY	4.33572
I	NCTMNOT	4.46546
J	TEAWRITE	5.58608
K	TEACURR	-21.31074
L	YEARSTAU	-1.37816
M	STUBELIE	3.35121
N	STWRITPS	0.15263
O	STMANIP	0.86472
P	STCALC	1.89126
Q	STHITECH	-2.28991
R	STCOLLAB	1.20780

Table 2. Least-squares Estimates of Fixed Effects (Robust Standard Errors) [*t*-statistic (*p*-value)] for Teaching Practices, 8th Grade (*df* = 13,282)

Teaching Practice	INTRCPT2	ZSCH_SES	SCH_RACE	PRIVAT_1
TCALC01	3.42 (1.96) *[1.796 (0.072)]*	0.00 (2.09) *[0.000 (1.000)]*	1.46 (2.61) *[0.561 (0.574)]*	1.85 (3.73) *[0.497 (0.618)]*
TMANIP01	-0.37 (1.80) *[-0.204 (0.839)]*	2.553972 (2.480236) *[1.030 0.304]*	-1.031617 (3.021644) *[-0.341 (0.733)]*	-4.433882 (3.726416) *[-1.190 (0.234)]*
TMULTYR1	2.769902 (2.090664) *[1.325 (0.185)]*	-0.434898 (2.08148) *[-0.209 (0.835)]*	-1.680663 2.950426 *[-0.570 (0.569)]*	-3.264888 3.022001 *[-1.080 (0.280)]*
TMULTWK1	-5.140987 (2.497190) *[-2.059 (0.039)]**	0.781761 (2.141813) *[0.365 (0.715)]*	3.554517 (3.250473) *[1.094 (0.275)]*	-5.027781 (5.903911) *[-0.852 (0.395)]*
TEAREA01	5.905223 (1.875022) *[3.149 (0.002)]**	-1.577709 (1.423093) *[-1.109 (0.268)]*	-4.779174 (2.365326) *[-2.021 (0.043)]**	4.844805 (3.185279) *[1.521 (0.128)]*
TTRAD01	-1.976338 (2.009581) *[-0.983 (0.326)]*	2.645099 (1.599338) *[1.654 (0.098)]*	2.138768 (2.671769) *[0.801 (0.424)]*	-4.595489 (3.465315) *[-1.326 (0.185)]*
TCOLPS	1.657155 (0.746848) *[2.219 (0.026)]**	0.126189 (0.844829) *[0.149 (0.882)]*	-0.844636 (1.090246) *[-0.775 (0.439)]*	-0.997366 (1.497242) *[-0.666 (0.505)]*
NCTMVERY	2.275967 (2.210386) *[1.030 (0.304)]*	0.638181 (1.574153) *[0.405 (0.685)]*	-1.925310 (2.134290) *[-0.902 (0.367)]*	4.045476 (4.640944) *[0.872 (0.384)]*
NCTMNOT	-4.285652 (2.094394) *[-2.046 (0.040)]**	1.227569 (2.604001) *[0.471 (0.637)]*	3.431540 (3.558834) *[0.964 (0.335)]*	0.385070 (3.876009) *[0.099 (0.921)]*
TEAWRITE	1.154953 (0.702072) *[1.645 (0.100)]*	0.936508 (0.810039) *[1.156 (0.248)]*	-1.318521 (1.002243) *[-1.316 (0.189)]*	0.246481 (1.472908) *[0.167 (0.867)]*
TEACURR	2.213212 (0.703222) *[3.147 (0.002)]**	-0.555412 (0.538696) *[-1.031 (0.303)]*	-0.298468 (0.699788) *[-0.427 (0.669)]*	3.097678 (1.374100) *[2.254 (0.024)]**
YEARSTAU	0.848948 (0.569301) *[1.491 (0.136)]*	0.166507 (0.630229) *[0.264 (0.792)]*	-2.278513 (0.920622) *[-2.475 (0.014)]**	-1.266522 (0.986661) *[-1.284 (0.200)]*
STUBELIE	5.283404 (0.780390) *[6.770 (0.000)]**	-0.154626 (0.371400) *[-0.416 (0.677)]*	-1.024818 (0.415417) *[-2.467 (0.014)]**	7.552690 (1.664391) *[4.538 (0.000)]**
STWRITPS	0.241740 (0.874660) *[0.276 (0.782)]*	-0.089210 (0.386283) *[-0.231 (0.818)]*	-0.044630 (0.559071) *[-0.080 (0.937)]*	1.249406 (1.877948) *[0.665 (0.506)]*
STMANIP	1.938456 (0.698544) *[2.775 (0.006)]**	-0.080828 (0.453455) *[-0.178 (0.859)]*	0.124463 (0.552352) *[0.225 (0.822)]*	2.565531 (1.543354) *[1.662 (0.096)]*
STCALC	1.462947 (0.630484) *[2.320 (0.020)]**	-0.249363 (0.724513) *[-0.344 (0.730)]*	0.343888 (0.711454) *[0.483 (0.628)]*	0.246081 (1.477011) *[0.167 (0.868)]*
STHITECH	-2.251794 (0.903398) *[-2.493 (0.013)]**	0.217576 (0.575763) *[0.378 (0.705)]*	0.004465 (0.524909) *[0.009 (0.993)]*	-4.393773 (2.097650) *[-2.095 (0.036)]**
STCOLLAB	1.980130 (0.988518) *[2.003 (0.045)]**	0.062520 (0.363575) *[0.172 (0.864)]*	0.005757 (0.363237) *[0.016 (0.988)]*	3.169639 (2.236632) *[1.417 (0.156)]*

* $p < .05$