

# EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal

Editor: Sherman Dorn

College of Education

University of South Florida

Volume 14 Number 14

June 1, 2006

ISSN 1068–2341

---

## Examining Instruction, Achievement, and Equity with NAEP Mathematics Data<sup>1</sup>

Sarah Theule Lubienski

University of Illinois at Urbana-Champaign

Citation: Lubienski, S. T. (2006). Examining instruction, achievement, and equity with NAEP mathematics data. *Education Policy Analysis Archives*, 14(14). Retrieved [date] from <http://epaa.asu.edu/epaa/v14n14/>.

### Abstract

The purpose of this article is two-fold. First, it reports on a study of the distribution of reform-oriented instructional practices among Black, White and Hispanic students, and the relationship between those practices and student achievement. The study identified many similarities in instruction across student groups, but there were some differences, such as Black and Hispanic students being assessed with multiple-choice tests significantly more often than were White students. Using hierarchical linear modeling, this study identified several significant positive—and no negative—relationships between reform-oriented practices and 4th-grade student achievement. Specifically, teacher emphasis on non-number mathematics strands, collaborative problem solving, and teacher knowledge of the NCTM *Standards* were positive predictors of achievement. An analysis of interaction effects indicated that the relationships between various instructional

---

<sup>1</sup> This project was funded through a National Assessment of Educational Progress Secondary Analysis Grant from the National Center for Education Statistics, Institute of Education Sciences. The author would like to thank Eric Camburn, Mack Shelley, Jay Verkuilen, Lateefah Id-Deen, Megan Brown and Chris Lubienski for their helpful advice during various stages of this work. Only the author is responsible for the analysis and interpretations presented in this report.



Readers are free to copy, display, and distribute this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available at <http://creativecommons.org/licenses/by-nc-nd/2.5/>. All other uses must be approved by the author(s) or **EPAA**. **EPAA** is published jointly by the Mary Lou Fulton College of Education at Arizona State University and the College of Education at the University of South Florida. Articles are indexed by H.W. Wilson & Co. Please contribute commentary at <http://epaa.info/wordpress/> and send errata notes to Sherman Dorn ([epaa-editor@shermamdorn.com](mailto:epaa-editor@shermamdorn.com)).

practices and achievement were roughly similar for White, Black and Hispanic students.

The second purpose of this article is to make comparisons with another study that used the same NAEP data, but drew very different conclusions about the potential for particular instructional practices to alleviate inequities. A study published in *EPAA* by Wenglinsky (2004) concluded that school personnel can eliminate race-related gaps within their schools by changing their instructional practices. Similarities and differences between these two studies are discussed to illuminate how a researcher's framing, methods, and interpretations can heavily influence a study's conclusions. Ultimately, this article argues that the primary conclusion of Wenglinsky's study is unwarranted.

Keywords: equity, hierarchical linear modeling; mathematics achievement; mathematics instruction; NAEP.

### **Examinando instrucción, logros, y equidad usando resultados de matemáticas de NAEP**

#### **Resumen**

El propósito de este artículo es doble. Primero, reportar los resultados de un estudio sobre la distribución de prácticas orientadas para producir reformas educativas entre estudiantes negros, blancos e hispánicos y el lazo entre esas prácticas y los logros de los estudiantes. Este estudio identificó muchas semejanzas en la instrucción de los grupos de estudiantes, pero también algunas diferencias. Por ejemplo los estudiantes negros e hispánicos son evaluados con pruebas de opción-múltiple considerablemente más a menudo que los estudiantes blancos. Usando modelos lineales jerárquicos, este estudio identificó varias relaciones significativas positivas y relaciones no-negativas entre prácticas orientadas para producir reformas educativas en los resultados educativos de estudiantes de 4to. grado. Específicamente, el énfasis de los profesores en la enseñanza de aspectos no-numéricos en matemáticas, la cooperación en la resolución de problemas, y el nivel de conocimiento de los profesores de los estándares de NCTM fueron predictores positivos del logro educativo. Un análisis de los efectos de las interacciones indicó que los lazos entre las diferentes prácticas y los logros educativos fueron bastante similares para los estudiantes blancos, negros e hispánicos.

El segundo propósito de este artículo fue hacer comparaciones con otro estudio que utilizó los mismos datos de NAEP, pero obtuvo conclusiones muy diferentes acerca del potencial de las prácticas orientadas para producir reformas educativas para aliviar desigualdades. Wenglinsky (2004), en un estudio publicado en *EPAA* concluyó que el personal de las escuelas puede eliminar las diferencias educativas relacionadas con aspectos raciales cambiando sus prácticas educacionales. Semejanzas y diferencias entre estos dos estudios se discuten para iluminar cómo el marco referencial, los métodos, y las interpretaciones de un investigador influyen sustantivamente en las conclusiones de un estudio. En última instancia, este

artículo discute que la conclusión primaria del estudio de Wenglinsky no esta plenamente justificada.

## Introduction

Identifying instructional practices that both boost achievement and promote equity has been of increasing concern among educators, researchers, and policy makers recently. This article reports the results of a study that focuses on the distribution of instructional practices advocated by the National Council of Teachers of Mathematics (NCTM), and the relationship between those practices and diverse students' achievement.

Recently a similar study was published in EPAA by Harold Wenglinsky (2004). Wenglinsky's study is comparable to this study in many important ways. Both studies utilized the 2000 National Assessment of Educational Progress (NAEP) mathematics data. Both studies used hierarchical linear modeling (HLM) to examine relationships between instructional practices and achievement, both overall and for particular subgroups. Additionally, both studies identified positive correlations between some reform-oriented instructional practices and overall student achievement. Yet, the studies began with different framings—this with an eye toward NCTM-endorsed practices, and Wenglinsky's with an eye toward the Bush Administration's No Child Left Behind (NCLB) act, which requires schools to closely monitor student achievement and reduce race-related achievement gaps. These differences in framing led, in part, to differences in the particular statistical models and methods employed. There was also a difference in the care with which findings were interpreted, including the extent to which causal attributions were made. Ultimately, different conclusions were reached. Specifically, Wenglinsky concluded that by changing instructional practices, "any gaps within a given school can be completely eliminated" (p. 17). In contrast, this study's conclusions are far less definitive and optimistic.

This article begins with a report of the current study, including its NCTM-based framing, its methods, and results. The article concludes with a comparison of its methods, results and interpretations with those of Wenglinsky, and ultimately raises questions about Wenglinsky's conclusions. Finally, issues pertaining to the analysis and reporting of NAEP and other large-scale data sets are highlighted.

## Background

Although NAEP mathematics scores have generally risen over the past 15 years (Braswell, Daane & Grigg, 2003; Kloosterman & Lester, 2004) there is some debate as to whether the achievement gains occurred because of, or in spite of, reforms promoted in the NCTM *Standards*. Although this cross-sectional study cannot offer a definitive resolution to this debate, it does offer a "bird's-eye" view of the distribution of some reform-oriented instructional methods, and their correlations with achievement for various student groups.

This study is situated at the intersection of work on reform-oriented instruction, mathematics achievement, and equity. Primary aspects of "reform-oriented mathematics instruction" are outlined first, followed by brief discussions of previous work regarding reform-oriented instruction and achievement, reform-oriented instruction and equity, and equity and mathematics achievement. This is followed by a more specific discussion of NAEP data, including a description of NAEP and findings of previous examinations of NAEP data regarding mathematics achievement, instruction, and equity.

## Reform-Oriented Mathematics Instruction

In 1989, NCTM published the *Curriculum and Evaluation Standards*, which, along with additional documents published subsequently (NCTM, 1991, 1995, 2000), called for mathematics instruction to be centered around students' reasoning, collaborative problem solving, and mathematical communication (both verbal and written). NCTM argued that a wider variety of tools (including manipulatives and calculators) and more meaningful forms of assessment should be employed. In addition, NCTM revised curricular goals for grades K–12 to include greater emphasis on measurement, geometry, data analysis, probability, algebra, as well as number concepts. Finally, NCTM called for “mathematical power for all students,” including those students previously under-represented in mathematics-based careers (NCTM, 1989, 1991, 1995, 2000).

### Reform-Oriented Mathematics Instruction and Student Achievement

The benefits of instruction aligned with the NCTM *Standards*—or “reform-oriented instruction” as it is termed here—have been the subject of much debate. Evidence from schools that have used new, reform-oriented curricula has generally been encouraging, with students outscoring control groups on a variety of measures and in a variety of contexts (e.g., Reys, Reys, Lapan, Holliday & Wasman, 2003; Riordan & Noyce, 2001; Schoenfeld, 2002; Senk & Thompson, 2003).

However, some critics of reform have pointed to less encouraging evidence, such as the fact that scores on NAEP's long-term-trend mathematics test remained flat during the 1990s, after a period of growth in previous decades (Loveless & Diperna, 2000). One need only make a brief visit to websites such as [Mathematicallycorrect.com](http://Mathematicallycorrect.com) or [NYCHold.com](http://NYCHold.com) to see that, despite the benefits of reform that some researchers report, much of the public is not convinced of the merits of reform-oriented instruction on a broad scale.

### Reform-Oriented Mathematics Instruction and Equity

Scholars have long argued that lower-SES and minority students have received more than their share of rote-based mathematics instruction (e.g., Anyon, 1981; Ladson-Billings, 1997; Means & Knapp, 1991). NCTM's vision of problem-centered instruction for all students challenges the status quo and is intended to correct past inequities (NCTM, 1989, 1991, 1995, 2000).

Now that the NCTM reforms are being implemented, scholars have begun to ask whether some students enter the mathematics classroom better positioned than others to learn in the ways envisioned in the *Standards* (Lubienski, 2000a, 2000b). Hickey, Moore, and Pellegrino (2001) found that reform-oriented instruction improved low- and high-SES students' problem solving skills, but the same instruction increased the SES-related gap in students' performance on the concepts and estimation portion of the Iowa Test of Basic Skills. However, still other studies have suggested that reform-minded practices are particularly beneficial for lower-SES and minority students (e.g., Boaler, 2002; Ladson Billings, 1997; Newmann & Wehlage, 1995; Schoenfeld, 2002; Silver, Smith, & Nelson, 1995; Stiff, 1990).

After analyzing national test score trends, Lee (2002) noted that Black-White gaps in achievement decreased during the 1970s–80s emphasis on “basic skills,” but increased during the 1990s, when emphasis shifted to higher-order thinking. Others have provided additional evidence of this trend (Campbell, Hombo, & Mazzeo, 2000; Jencks & Phillips, 1998). These studies raise the

question of whether these patterns are caused by reform-oriented instruction, differential access to such instruction, or other confounding variables.<sup>2</sup>

### *Equity and Achievement*

In recent years, many researchers have struggled to understand the underlying causes of persistent inequities in academic achievement, especially race-related achievement gaps.<sup>3</sup> Clearly, SES differences involving parent education, occupation, income, and educational resources in the home account for much of these gaps (Jencks & Phillips, 1998; Peng, Wright, & Hill, 1995). Several studies of race-related achievement gaps have also examined school-related factors, including the roles of teachers, curricula, school funding, student motivation, and student resistance (e.g., Banks, 1988; Cook & Ludwig, 1998; Ferguson, 1998a, 1998b, 1998c; Ogbu, 1995; Steele & Aronson, 1998). Such discussions have tended to focus on the overall academic performance and experiences of minority students, as opposed to an in-depth examination of achievement and instructional practices in a particular subject area, such as mathematics. This trend was noted by Lee (2002), who concluded his general analysis of patterns in achievement data by urging subject matter specialists to further examine inequities in their areas of expertise.

This study does not attempt to enter into debates about the many factors outside of schools that contribute to achievement inequities, but instead focuses on instructional variables over which educators have control. This study focuses specifically on students' achievement and learning experiences in mathematics, which is a particularly important subject to consider in relation to equity because it is a key gatekeeper for entry into high status occupations. Researchers in mathematics education have given some attention to race-related gaps in mathematics achievement, but have rarely examined race and SES simultaneously (Lubienski & Bowen, 2000; Tate, 1997). By exploring the relationship between particular instructional practices and achievement utilizing hierarchical linear models that include both race and SES, this study examines the extent to which race-related achievement gaps that persist after controlling for SES may be related to differences in students' access to particular mathematics instructional practices, as measured by NAEP.

### *The National Assessment of Educational Progress*

NAEP is the only nationally representative, ongoing assessment of U.S. academic achievement. NAEP measures student performance at 4th, 8th, and 12th grades in mathematics and other subject areas. NAEP also provides survey information from students and their teachers regarding mathematical backgrounds, beliefs, and instructional practices.

Since 1990, the Main NAEP mathematics assessment has been guided by a framework based on NCTM's *Curriculum and Evaluation Standards for School Mathematics* (1989). Hence, the Main

---

<sup>2</sup> The publication of the 2003 NAEP results prompted even more discussion about these issues, as both NCTM and NCLB were credited by various parties for improvements in scores and decreases in achievement gaps between 2000 and 2003. However, the 2003 data were not yet available for secondary analysis at the time of this study.

<sup>3</sup> For the sake of simplifying the text, the term "race" is used very loosely to mean race or ethnicity when referring to the NAEP categories of Black, White and Hispanic students. Additionally, to be consistent with NAEP data, the terms "Black", "White," and "Hispanic" are used throughout this article.

NAEP assesses students' performance on both multiple-choice and constructed-response items over the five mathematics strands emphasized by NCTM: number/operations, geometry, measurement, data analysis, and algebra/functions. Additionally, some NAEP survey questions administered to students and teachers were designed to identify the extent to which students' classroom experiences are aligned with NCTM's vision for mathematics instruction.

### **Previous NAEP Findings on the Distribution of Reform-Oriented Mathematics Instruction**

Strutchens and Silver (2000) gave detailed attention to race-related disparities in 1996 NAEP data on mathematics achievement, students' beliefs about mathematics, and teachers' instructional practices and emphases. They found that Black and Hispanic students were at least as likely as White students to have access to manipulatives, "real-life" mathematics problems, and student collaboration in their mathematics classrooms. However, according to teacher reports, White eighth graders were more likely than Black or Hispanic students to receive some aspects of reform-oriented instruction, such as calculator access, fewer multiple-choice assessments, and a heavy emphasis on reasoning.

Students' mathematical attitudes and beliefs, although shaped by a variety of factors, are linked to the instruction students receive. Strutchens and Silver (2000) reported that Black and Hispanic students were more likely than White students to agree with the statements, "There is only one way to solve a math problem" and "Learning mathematics is mostly memorizing facts." However, they cautioned that the race-related differences they reported might be due more to SES than race.

More recently, Strutchens, Lubienski, McGraw and Westbrook (2004) examined the 2000 Main NAEP mathematics data and confirmed the above 1996 findings, with the exception of differential access to teacher-reported emphasis on reasoning, for which there were no longer race-related disparities in the 2000 data. Lubienski and Shelley (2003) extended this work and found that the race-related gaps persisted even after controlling for SES. Additionally, in their analysis of 1996 State NAEP data, Swanson and Stevenson (2002) identified SES-related differences in instruction, with more affluent schools tending to utilize more reform-oriented practices, as measured by a single composite of 16 variables.

Overall, previous analyses of NAEP data have indicated some potentially important ways in which White, higher-SES students are experiencing more of the fundamental instructional shifts called for by NCTM than less privileged students. These differences are reminiscent of those discussed by Means & Knapp (1991), Anyon (1981) and others who observed poor and minority students receiving more than their share of drill-based, computation-focused instruction.

### **Previous NAEP Findings on Reform-Oriented Mathematics Instruction, Achievement and Equity**

The official NAEP report for the 2000 main mathematics assessment highlighted several instruction-related variables that correlated with achievement. For example 8th graders with unrestricted access to calculators scored significantly higher than did their peers without such access. Similarly, the report stated that 4th, 8th, and 12th grade students who agreed with the statement, "Learning math is mostly memorizing facts," scored significantly lower than did students who disagreed with the statement (Braswell, Lutkis, Grigg, Santapau, Tay-Lim, & Johnson, 2001). However, given that White, high-SES students have been more likely than their less privileged counterparts to have unrestricted calculator access and to believe that mathematics is more than just fact memorization (Lubienski, 2002; Strutchens et al., 2004), race and SES are likely confounding variables in the correlations noted by Braswell et al. (2001).

Hence, the question remains whether reform-oriented instructional practices, as reported in NAEP teacher surveys, are positive predictors of achievement after controlling for confounding variables. If so, then the differences in instructional practices noted above might contribute to race- and SES-related achievement differences.

A prior study by Raudenbush, Fotiu, and Cheong (1998), utilizing 1992 State NAEP data, found that teacher-reported emphasis on reasoning in mathematics instruction correlated positively with achievement even after controlling for race and SES, and that White, high-SES students were more likely to have such a teacher. However, disparities in students' access to teachers emphasizing reasoning were no longer significant in the 2000 NAEP data.

Finally, as noted previously, at the same time this study was being conducted, Wenglinsky (2004) used HLM to analyze the 2000 NAEP Mathematics Data, examining whether particular instructional practices related to schools' overall achievement and the size of their race-related gaps. He found that teacher-reported time on task, use of routine exercises and a geometry emphasis correlated with higher achievement for students, in general, whereas frequent testing, emphasis on facts, and project work correlated negatively with achievement. He also concluded that an emphasis on measurement "was the most beneficial practice" (p. 16) for Black students, while an emphasis on data analysis appeared beneficial for Hispanic students. However, as will be discussed in more detail later, his conclusions require further consideration.

The 2000 Main NAEP data included larger samples than previous administrations, and also included dozens of teacher-reported variables relating to reform-oriented instruction (many of which were deleted in 2003). An in-depth analysis of instruction and achievement using the 2000 data can illuminate relationships among reform-oriented instructional measures, student achievement, and equity. Still, given that NAEP data are cross-sectional and not longitudinal, no NAEP-based study can definitively determine which instructional methods are most effective for particular groups of students. Still, the scope and representative nature of NAEP data can lend important evidence to inform current debates and to point toward areas in need of further research.

## **Research Questions**

In the context of the NCTM reform movement and concerns about its impact on mathematics achievement and equity, this study addresses three questions. First, the study examines the extent to which reform-oriented instructional practices are reaching all students, regardless of race. Second, the study investigates whether particular reform-oriented instructional practices correlate positively or negatively with mathematics achievement, after controlling for race, SES, and other potentially confounding variables. Finally, the study considers whether reform-oriented practices correlate similarly with achievement for diverse student groups, regardless of student race or SES.

Taken together, these questions probe whether inequities in access to reform-oriented instruction might contribute to achievement gaps, with a particular focus on Black-White and Hispanic-White gaps that persist even after controlling for student- and school-level SES. Identifying inequities in access to instructional methods that correlate positively or negatively with achievement can shed light on variables potentially underlying achievement gaps, enrich our understanding of students' experiences with learning mathematics, and suggest important areas for further study. While not assuming that instruction-related variables are the only, or even primary, cause of achievement gaps, it is important to give attention to the area that educators are best positioned to address.

## Method

Several methodological features of this study merit discussion including the NAEP samples used, special challenges of NAEP analyses, the specific variables utilized, and the analyses conducted.

### *The Samples*

The 2000 Main NAEP data used in this study were accessed from a restricted-use CD-ROM.<sup>4</sup> Data regarding the mathematics achievement of a nationally representative sample of 13,511 4th graders who were assessed in late winter/early spring, 2000, were included, as well as data from student background surveys and teacher reports of instructional practices. The unweighted sample of students was 64% White, 17% Hispanic, 13% Black, and 6% other groups. The analyses reported here were part of a larger study that gave attention to both 4th and 8th grades, and that examined instruction-related variables as reported by both students and teachers.<sup>5</sup> For the sake of space limitations and comparability with Wenglinsky's study, analyses of fourth grade achievement and teacher-reported data are the primary focus here. However, additional findings from the larger study are footnoted when particularly relevant.

### **Methodological Challenges of NAEP Data Analyses**

Several features complicate the analysis of NAEP data. To obtain a representative sample of students, schools are stratified based on urbanicity, minority population, size, and area income, and then schools within each stratum are selected at random. Finally, students are selected randomly within schools. Deliberate oversampling of certain strata, such as schools with high enrollments of minority students, results in more reliable estimates for the oversampled subgroups, and then student and school weights are used to adjust for both unequal probabilities of selection and nonresponse. To account for the clustered sampling, NAEP data also contains replicate weights for each student and school, which are used in calculating sampling errors using the jackknife repeated replication method. Teacher weights are not assigned, because NAEP selects samples of students and then surveys their teachers; teacher data are linked to student data and are interpreted at the student level. As a concrete example, NAEP analyses would not indicate that 80% of teachers reported allowing unrestricted calculator use, but that 80% of *students* had teachers who reported allowing unrestricted calculator use.

To reduce the test-taking burden on individual students, NAEP administers only a subset of items to each student. Hence, individual students' achievement is not measured reliably enough to be assigned a single "score." Instead, using Item Response Theory (IRT), NAEP estimates a distribution of plausible values for each student's proficiency, based on the student's responses to administered items and other student characteristics. When analyzing NAEP achievement data, separate analyses are conducted with the five plausible values assigned to each student. The five sets of results are then synthesized, following Rubin (1987) on the analysis of multiply-imputed data. For

---

<sup>4</sup> Researchers who apply for a license from the National Center for Education Statistics may obtain restricted-use data.

<sup>5</sup> A detailed (120-page) report of the methods and results of the full study was submitted to the National Center of Education Statistics (Lubienski, Camburn & Shelley, 2004). Interested readers may download the report from [www.ed.uiuc.edu/naep](http://www.ed.uiuc.edu/naep).



more detailed information regarding the structure of NAEP data, see Johnson (1992) and Johnson and Rust (1992).

### Demographic and Instruction-Related Variables

Several student- and school-level demographic variables were included in this analysis, along with teacher-reported variables pertaining to instructional practices:<sup>6</sup>

*Student race.* Binary “Black” and “Hispanic” variables were created from NAEP’s student race/ethnicity variable (taken from student self-reports, or school records when necessary).

*School race.* There was no school race variable in the 2000 NAEP mathematics data set. As a proxy, the percentage of White/Asian students in each school’s sample was calculated.<sup>7</sup>

*Student SES.* After consideration of the much-debated meanings of “socioeconomic status” and “social class” (e.g., Duberman, 1976; Weis, 1988), a comprehensive SES variable was created using factor analysis. Six variables were combined to produce a new student SES variable: types of reading material in students’ homes (newspapers, magazines, books, and encyclopedias), computer and Internet access at home, extent to which studies are discussed at home, and eligibility for school lunch and Title 1 (a federal program for disadvantaged students). Parent education levels were not reported for 4th graders in 2000 and were therefore not included. The final variable was standardized with a mean of 0 and standard deviation of 1.

*School SES.* At each sampled school, an administrator provided survey data regarding the percentage of students qualifying for Title 1 funds and free/reduced lunch. These two variables were ordinal with rough categories of percentages (e.g., 0–10, 11–25, etc.). The final school SES measure was a composite of the student-level SES variable aggregated to the school level and the percentage of students eligible for free/reduced lunch and Title 1.<sup>8</sup>

*Gender.* NAEP’s “Gender” variable (coded as “boy” = 1, else = 0) was included in the analyses because prior research suggests that gender correlates significantly with mathematics achievement (e.g., Fennema, Carpenter, Jacobs, Franke & Levi, 1998; Lubienski, McGraw & Strutchens, 2004).<sup>9</sup>

*Disability.* Given that students with disabilities tend to score lower than others on NAEP (Foegen, 2004) and that these students could be subject to different instructional practices than their peers, a binary student disability variable was used to control for whether students have a non-orthopedic disability (e.g., learning disability, visual impairment, behavioral disorder).

---

<sup>6</sup> Teacher background/certification variables were also considered, including undergraduate major and whether teachers held master’s degrees. These variables were ultimately omitted from the fourth-grade analyses due to a lack of significance. However, in the 8th-grade analyses, secondary mathematics certification was significant.

<sup>7</sup> White and Asian students were combined because these groups tend to have higher achievement than other groups. The resulting variable was skewed and somewhat bi-modal (revealing school segregation patterns). A natural logarithmic transformation was used to create a somewhat more normally distributed variable, which was then standardized with mean = 0 and standard deviation = 1. Although a more normally distributed variable was desirable for inclusion in the models, the tradeoff in using such a transformation is that the resulting variable is more difficult to interpret.

<sup>8</sup> The final composite was standardized with a mean of 0 and standard deviation of 1. For additional details about the creation of this and other demographic indicators, see Lubienski, Camburn & Shelley, 2004.

<sup>9</sup> NAEP’s teacher-reported data generally does not vary by gender because each teacher survey is linked to all students selected from his/her class. However, the larger study also included student-reported instruction-related data, which can vary by gender. For the sake of consistency, gender was included in all models.

*School sector.* Public/private school status has been found to correlate with achievement (e.g., Bryk, Lee & Holland, 1993; Lubienski & Lubienski, 2006) and might also relate to the instructional practices employed. The NAEP variable “sctype” was recoded, with Catholic and other private schools = 1 and public schools = 0.

*Teacher-reported instruction-related variables.* During initial explorations of teacher-reported variables that could conceivably be viewed as measuring some aspect of reform-oriented practices, the net was cast widely to include 31 such variables. Factor analysis was used to create composites of highly correlated instruction-related variables, thereby reducing the number of predictors included in the HLM analyses and decreasing the danger of “fishing” for correlations among multiple variables.<sup>10</sup>

Several variables did not seem to fit with the others and were excluded because upon further consideration it did not make conceptual or statistical sense to include them. These variables included the frequency with which students took tests, did problems from textbooks, and used computers in mathematics class. Although these variables could be construed to relate to reform-oriented instruction, closer inspection of the content of the questions combined with their lack of correlation with other reform measures suggested that these variables were not essential measures of reform-oriented instruction. Ultimately, 24 variables remained, with most clustering around six themes: calculator use, facts and skills, collaborative problem solving, non-number curricular emphasis, writing about mathematics, and manipulative use.<sup>11</sup>

Teacher emphasis on reasoning, use of multiple choice assessments, and teachers’ knowledge of the NCTM *Standards* tended to correlate loosely with the other variables, but did not associate strongly with any single factor or with each other. These variables were included among the final set of instruction-related measures, but were treated individually.

Six factor analyses were conducted—one with each of the 6 clusters of variables—to create a single, standardized factor (with a mean of 0 and standard deviation of 1) representing each theme. In each case, only one factor resulted with an Eigenvalue greater than 1, so that factor was used to represent the cluster. The loadings of each of the original variables on the final resulting factors are listed in Table 1, along with Cronbach’s alpha, an indicator of how closely the items correlate with one another.<sup>12</sup>

---

<sup>10</sup> The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) with all of the variables in a single factor analysis was roughly .8, indicating that factor analysis was appropriate.

<sup>11</sup> Variables clustered similarly in the 8th-grade factor analysis, providing some evidence that the NAEP survey items are capturing some meaningful differences in teacher instruction, despite the rather rough response categories used and the self-reported nature of the data.

<sup>12</sup> When designing a survey and developing item clusters, Cronbach’s Alphas between .7 and .9 are desirable. However, lower values were considered acceptable for the purposes of this study, in which the goal was not to design a new survey, but to create composites of existing survey items that capture various aspects of reform-oriented instruction. Conceptual connections among items (e.g., whether items refer to calculators or manipulatives) and a desire for consistency in the created composites across 4th and 8th grade were also considered in the development of the composites.

Table 1  
*Teacher-Reported Instruction-Related Factors—Questions with Loadings*

| Questions Grouped by Factors                                                                                     | Response Options                                                         | (Alpha) or Loading |
|------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------|--------------------|
| 1) Calculators                                                                                                   |                                                                          |                    |
| How often do the students in this class use a calculator?                                                        | Almost every day, 1–2 times a week, 1–2 times a month, Never/hardly ever | .72                |
| Do you permit students in this class to use calculators for tests?                                               | Yes, No                                                                  | .79                |
| Do you permit students in this class unrestricted use of calculators?                                            | Yes, No                                                                  | .71                |
| 2) Emphasis on Facts & Skills                                                                                    |                                                                          |                    |
| How much emphasis did you or will you give to learning mathematics facts and concepts?                           | Heavy emphasis, Moderate emphasis, Little/no emphasis                    | (.62)              |
| How much emphasis did you or will you give to learning skills and procedures needed to solve routine problems?   | Heavy emphasis, Moderate emphasis, Little/no emphasis                    | .83                |
| How much emphasis did you or will you give to number and operations?                                             | Heavy emphasis, Moderate emphasis, Little/no emphasis                    | .74                |
| How much emphasis did you or will you give to learning skills and procedures needed to solve routine problems?   | Heavy emphasis, Moderate emphasis, Little/no emphasis                    | .70                |
| 3) Collaborative Problem Solving                                                                                 |                                                                          |                    |
| How often do the students in this class discuss solutions to mathematics problems with other students?           | Almost every day, 1–2 times a week, 1–2 times a month, Never/hardly ever | (.76)              |
| How often do the students in this class work and discuss mathematics problems that reflect real life situations? | Almost every day, 1–2 times a week, 1–2 times a month, Never/hardly ever | .84                |
| How often do the students in this class talk to the class about their mathematics work?                          | Almost every day, 1–2 times a week, 1–2 times a month, Never/hardly ever | .74                |
| How often do the students in this class solve mathematics problems in small groups or with a partner?            | Almost every day, 1–2 times a week, 1–2 times a month, Never/hardly ever | .72                |
| How much emphasis did you or will you give to learning how to communicate ideas in mathematics effectively?      | Heavy emphasis, Moderate emphasis, Little/no emphasis                    | .65                |
| 4) Non-Number Curricular Emphasis                                                                                |                                                                          |                    |
| How much emphasis did you or will you give to learning how to communicate ideas in mathematics effectively?      | Heavy emphasis, Moderate emphasis, Little/no emphasis                    | .62                |
| How much emphasis did you or will you give to geometry?                                                          | Heavy emphasis, Moderate emphasis, Little/no emphasis                    | (.68)              |
| How much emphasis did you or will you give measurement?                                                          | Heavy emphasis, Moderate emphasis, Little/no emphasis                    | .78                |
| How much emphasis did you or will you give measurement?                                                          | Heavy emphasis, Moderate emphasis, Little/no emphasis                    | .65                |

| Questions Grouped by Factors                                                                                                                                                                         | Response Options                                                               | (Alpha) or Loading |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------|--------------------|
| How much emphasis did you or will you give data analysis, statistics, and probability?                                                                                                               | Heavy emphasis, Moderate emphasis, Little/no emphasis                          | .76                |
| How much emphasis did you or will you give Algebra and functions (informal introduction of concepts)?                                                                                                | Heavy emphasis, Moderate emphasis, Little/no emphasis                          | .69                |
| 5) Writing About Mathematics<br>How often do you use short (e.g., a phrase or sentence) or long (e.g., several sentences or paragraphs) written responses to assess student progress in mathematics? | 1-2 times a week, 1-2 times a month, 1-2 times a year, Never/hardly ever       | .76<br>(.69)       |
| How often do you use individual or group projects or presentations to assess student progress in mathematics?                                                                                        | 1-2 times a week, 1-2 times a month, 1-2 times a year, Never/hardly ever       | .70                |
| How often do the students in this class write a few sentences about how to solve a mathematics problem?                                                                                              | Almost every day, 1-2 times a week, 1-2 times a month, Never/hardly ever       | .76                |
| How often do the students in this class write reports or do mathematics projects?                                                                                                                    | 1-2 times a week, 1-2 times a month, 1-2 times a year, Never/hardly ever       | .67                |
| 6) Manipulatives<br>How often do the students in this class work with objects like rulers?                                                                                                           | Almost every day, 1-2 times a week, 1-2 times a month, Never/hardly ever       | (.66)<br>N/A*      |
| How often do the students in this class work with counting blocks or geometric shapes?                                                                                                               | Almost every day, 1-2 times a week, 1-2 times a month, Never/hardly ever       | N/A*               |
| 7) Reasoning<br>How much emphasis did you or will you give to developing reasoning and analytic ability to solve unique problems?                                                                    | Heavy emphasis, Moderate emphasis, Little/no emphasis                          | N/A*               |
| 8) Multiple-Choice Assessment<br>How often do you use multiple-choice tests to assess student progress in mathematics?                                                                               | 1-2 times a week, 1-2 times a month, 1-2 times a year, Never/hardly ever       | N/A*               |
| 9) Knowledge of NCTM Standards<br>How knowledgeable are you about the NCTM Curriculum and Evaluation Standards for School Mathematics?                                                               | Very knowledgeable, knowledgeable, Somewhat knowledgeable, Little/no knowledge | N/A*               |

\*Loadings not relevant because variable was not part of a composite of three or more variables.

Given that the goal was to use these variables as predictors of achievement in HLM models, it was preferable for them to be either continuous, normal variables or binary. The single, isolated variables (teacher emphasis on reasoning, multiple-choice assessment use, knowledge of the NCTM *Standards*) were ordinal, not continuous. A few of the variables created through a combination of factors were heavily skewed or bimodal. These issues were addressed by creating binary variables as follows:

*Calculators* (which was a bimodal composite) was recoded so that above average calculator use = 1, else = 0.

*Facts and skills* was recoded so that heavy emphasis on facts and concepts, skills for routine problems, and number/operations = 0, otherwise 1.

*Reasoning* was recoded so that heavy emphasis = 1, moderate or light emphasis = 0

*Multiple choice assessment use* was originally an ordinal variable with 4 categories. There seemed to be substantial differences between teachers using multiple choice assessments weekly, monthly, and annually/never, so two binary variables were created: weekly = 1 and less than weekly = 0, and once or twice annually or never = 1, otherwise 0. This effectively separates the weekly, monthly, and annually/never groups.

*Knowledge of NCTM Standards* originally had four categories: Very knowledgeable, knowledgeable, somewhat knowledgeable, and little/no knowledge. Two binary variables were created: 1 = very knowledgeable about the *Standards*, otherwise = 0; and 1 = little or no knowledge about the *Standards*, otherwise = 0. This set distinguishes between the two extremes and combines the two middle categories.

The remaining continuous variables, collaborative problem solving, non-number curricular emphasis, writing about mathematics, and manipulatives were standardized with a mean of 0 and standard deviation of 1. With the exception of “weekly multiple-choice assessment use” and “little or no knowledge of the NCTM *Standards*” each variable was coded so that a higher number indicated a greater alignment with the NCTM *Standards*.

### *Data Analysis*

The initial, descriptive phase of data analysis addressed the first research question: Are reform-oriented instructional practices reaching all students, regardless of race? HLM models were then developed to answer the second and third research questions: Which reform-oriented instructional practices correlate positively or negatively with mathematics achievement after controlling for confounding variables? Do those correlations vary by student race and/or SES?

*Phase 1—Descriptive analyses of instruction by race.* Means of the newly created instruction-related variables were compared for White, Black and Hispanic students to examine whether differences emerged for the instructional composites created for this study. These comparisons were made using AM Statistical Software, designed by the American Institutes for Research to handle the special weighting and jackknifing needs of complex data sets such as NAEP. Two-tailed T-tests were used to determine if means significantly differed between White and Black students and between White and Hispanic students. When interpreting results, issues of multiple comparisons were considered using Bonferroni corrections.

*Phase 2—HLM analyses of instruction and achievement.* Because of the nested nature of the data (students and teachers within schools), two-level HLMs were used to examine whether particular reform-based practices positively or negatively predicted achievement while controlling for potentially confounding variables at both the student and school level. HLM statistical software was designed specifically to accommodate multi-level datasets, including those with plausible values

(Raudenbush & Bryk, 2002). As HLM computes statistics related to NAEP achievement, each parameter is estimated for each of the five plausible values, and the five estimates are then averaged.<sup>13</sup>

In the HLM models, students (level 1) were nested within schools (level 2).<sup>14</sup> The level of classroom or teacher was not included as a separate level because NAEP uses random samples of students and not teachers, and there were no teacher codes in the data to allow for analysis at the teacher level. Given these constraints, and given the study's primary focus on student-level disparities in instruction and achievement (as opposed to school-level issues), teacher-level data were treated as level-1 data. In this way, the instructional practices linked with students were those they had experienced during that school year. It is important to note that, given the lack of a prior achievement measure in NAEP, this study does not examine the *change* in students' achievement during the school year. Therefore, it is very possible that relationships between instruction and achievement that appear weak or insignificant in this study could be found to be stronger in longitudinal studies.

Because of concerns about collinearity among the 9 teacher-reported instructional practice variables, separate HLM analyses were conducted with each of the variables to determine the relationship of each with student achievement. The student- and school-level demographic variables described above were also included as predictors in the models. Given the focus on general relationships between NCTM practices and achievement, as opposed to variation in their slopes by school, slopes were fixed in the HLM models, and continuous predictors were centered around their overall means. Binary predictors were not centered. The changes in coefficients for Black and Hispanic students that occurred after adding each instructional variable to the model were examined in an attempt to gauge the possible impact of each instructional practice on the race-related achievement gaps that persisted after controlling for SES. This change in coefficients was examined separately with HLM models for each of the 9 instruction-related variables, and interaction effects were included in the final models to examine whether the coefficient for each instruction-related variable differed by student race and SES. Finally, a larger HLM model was created to examine the change in coefficients and variance when the 9 instruction-related variables were included simultaneously, yet this model was interpreted cautiously because of collinearity among the 9 instruction-related predictors.

## Results

To help the reader interpret the results discussed here, some information about NAEP scores is necessary. NAEP uses a 500-point scale on which 4th graders scored an average of 228 in 2000. The fourth-grade Hispanic-White gap was 24 points, and the Black-White gap was 31 points. The standard deviation for the 2000 fourth-grade scale scores was 31 points. Hence, a difference of 3 points can be considered an effect size of roughly 0.1. Note that the size of the Black-White

---

<sup>13</sup> A more detailed explanation of the data analysis methods that the HLM program uses is available in Bryk and Raudenbush (1992) and Raudenbush and Bryk (2002).

<sup>14</sup> Due to missing survey data, HLM samples were reduced to include 9,999 students across 611 schools. The demographics for these reduced samples were very similar to the demographics of the entire data set, lessening concerns about the results of the analyses being skewed due to missing data.

fourth-grade gap was a full standard deviation (very large effect size of 1), and the Hispanic-White gap had an effect size of 0.8.<sup>15</sup>

Table 2

*Means of Grade 4 Teacher-Reported Instruction-Related Factors by Student Race*

| Factor                                                      | White |      | Black  |      | Hispanic |      |
|-------------------------------------------------------------|-------|------|--------|------|----------|------|
|                                                             | Mean  | S.E. | Mean   | S.E. | Mean     | S.E. |
| Greater than Average Calculator Use = 1 (binary)            | .35   | .03  | .37    | .04  | .27*     | .03  |
| Not Heavy Emphasis on Facts and Skills = 1 (binary)         | .22   | .02  | .23    | .02  | .22      | .02  |
| Collaborative Problem Solving                               | -.02  | .05  | .06    | .06  | -.03     | .06  |
| Non-number Curricular Emphasis                              | -.06  | .06  | .16*   | .09  | .11*     | .07  |
| Writing About Math                                          | -.04  | .06  | .12    | .07  | .06      | .06  |
| Manipulatives                                               | -.05  | .06  | .12*   | .06  | .09      | .07  |
| Heavy Emphasis on Reasoning = 1 (binary)                    | .61   | .02  | .65    | .03  | .58      | .03  |
| Weekly Multiple Choice Assessment Use = 1 (binary)          | .09   | .01  | .23*** | .03  | .22***   | .02  |
| Yearly or Never Multiple Choice Assessment Use = 1 (binary) | .44   | .03  | .30*** | .03  | .32***   | .03  |
| Very Knowledgeable About NCTM Standards = 1 (binary)        | .07   | .01  | .04    | .01  | .06      | .01  |
| Little/No Knowledge About NCTM Standards = 1 (binary)       | .35   | .03  | .36    | .02  | .41      | .03  |

Note: Means for Black and Hispanic students were compared with means of White students in the significance tests.

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$

*Means by Race of Instruction-Related Variables*

The means and standard errors for the teacher-reported instructional composites for White, Black and Hispanic fourth graders are presented in Table 2.<sup>16</sup> There were no significant race-related differences in teacher emphasis on reasoning and facts/skills, teacher knowledge of the NCTM *Standards*, collaborative problem solving, and writing about mathematics. Black and Hispanic

<sup>15</sup> Although there have been various methods proposed for calculating effect sizes (Thompson, 2002), this discussion refers to Cohen's (1988)  $d$ , which is computed by dividing the difference between the means of two samples by the standard deviation of the combined population sample.

<sup>16</sup> In the full study, 28 t-tests were used to compare White-Black and White-Hispanic means for 14 teacher- and student-reported instruction-related variables. Hence, it can be considered appropriate to hold  $p < .05/28 = .002$  as the standard for determining statistical significance, using the Bonferroni correction for multiple comparisons. However, others might argue for a different clustering of variables for the Bonferroni correction (e.g., dividing .05 by 2, because two comparisons were made for each variable). Hence, the standard .05, .01, and .001 significance levels are reported in the tables, leaving readers to interpret results as they see fit.

students were at least as likely as White students to have access to manipulatives and a non-number curricular emphasis. For example, whereas White students were 0.05 standard deviations below the mean for use of manipulatives, Black students were 0.12 standard deviations above the mean. Hence, Black students actually appeared to be getting slightly more access to some reform-oriented practices than were their White peers (with means for Hispanic students generally in between those for Black and White students). However, consistent with previous findings (Strutchens, et al., 2004), Black and Hispanic students were significantly more likely to be assessed with multiple choice tests than were White students. For example, 44% of White students were assessed with multiple choice tests no more than once or twice a year, whereas this percentage was 30% for Black students and 32% for Hispanic students (because this is a binary variable, the means can be interpreted as percentages).<sup>17</sup>

### *HLM Analyses of Instruction-Related Factors and Achievement: The Example of Calculator Access*

HLM analyses were undertaken to examine the relationship between particular reform-oriented instructional practices and mathematics achievement, as measured by NAEP. Because of concerns of multi-collinearity, separate models were created for each of the 9 instructional factors. Due to space limitations, the full results of each of the HLM models are not presented here. Instead, full details of the models involving the teacher-reported calculator composite is presented as an example, and then the main results involving the remaining instruction-related factors are summarized.

Table 3 presents the set of models run with the 4th-grade calculator use composite. Recall that the calculator use variable was binary, with 1 = above average and 0 = at or below average. Models 1, 2, and 3 remained constant for all grade 4 HLM analyses regardless of the instruction-related variable in question. The base model (Model 1) shows that the mean achievement across all sampled schools was 230.4 points. It also indicates that roughly one third of the variance in achievement was between schools (intraclass correlation=.34), and two thirds of the variance was among students within schools. According to model 2, the mean for Black students was about 23 points lower than that of their non-Hispanic peers within the same school (White/Asian students were the primary comparison group, with a mean achievement of 235)<sup>18</sup>, whereas Hispanic students scored about 17 points lower. The addition of these two student-level race variables accounted for almost 40% of the variance between schools, but only about 5% of the variation within schools. In Model 3, we can see that student and school SES, gender, disability, and school sector all significantly predicted achievement. For example, an increase in one standard deviation in SES was associated with a 7.6 point increase in achievement at the student level and 6.1 points at the school level. Similarly, the coefficients reported in model 3 indicate a 3.8 point advantage for males and a 30.3 point disadvantage for students with disabilities. Additionally, the private school students in the

---

<sup>17</sup> More frequent multiple-choice testing for Black and Hispanic students was also found at grade 8. Black and Hispanic eighth graders were also less likely to be given access to calculators by their mathematics teachers.

<sup>18</sup> The American Indian/Alaskan Native subgroup comprises only about 2% of students. Because of their small sample size, these students were not denoted by a separate variable, but were included in the general default group of “non-Black”, “non-Hispanic” students.



sample performed a significant 4.8 points lower than the public school students sampled.<sup>19</sup> The coefficient for school race was insignificant (though it was close to significant, and it was significant when school SES was not included in the model). Model 3 also indicates that even after controlling for these other contextual variables, there are still highly significant race-related gaps within schools of 12.5 (Hispanic) and 17.1 (Black) points. Taken together, the demographic factors in Model 3 explained 70% of the variance between schools, and 16% of the variance within schools.

Table 3  
*HLM Models of NAEP Achievement by Teacher-Reported Calculator Use and Achievement, Grade 4*

| Variable                             | Model 1                   | Model 2                   | Model 3                   | Model 4                   | Model 5                   |
|--------------------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Level 1 (student & his/her teacher)  |                           |                           |                           |                           |                           |
| Intercept                            | 230.4                     | 235.0***                  | 236.6***                  | 236.6***                  | 236.3***                  |
| Black                                |                           | -22.8***                  | -17.1***                  | -17.1***                  | -16.8***                  |
| Hispanic                             |                           | -17.3***                  | -12.5***                  | -12.5***                  | -12.1***                  |
| Student SES                          |                           |                           | 7.6***                    | 7.6***                    | 7.1***                    |
| Boy                                  |                           |                           | 3.8***                    | 3.8***                    | 3.8***                    |
| Disability                           |                           |                           | -30.3***                  | -30.3***                  | -30.4***                  |
| Calculator Use                       |                           |                           |                           | -0.1                      | -0.6                      |
| Calculator x Black                   |                           |                           |                           |                           | -0.7                      |
| Calculator x Hispanic                |                           |                           |                           |                           | -1.4                      |
| Calculator x SES                     |                           |                           |                           |                           | 1.5                       |
| Level 2 (school)                     |                           |                           |                           |                           |                           |
| School SES                           |                           |                           | 6.1***                    | 6.1***                    | 6.1***                    |
| School Race/Ethnicity                |                           |                           | 1.0                       | 1.0                       | 1.0                       |
| Private School                       |                           |                           | -4.8***                   | -4.8***                   | -4.6**                    |
|                                      | <i>Variance Component</i> | <i>Variance Component</i> | <i>Variance Component</i> | <i>Variance Component</i> | <i>Variance Component</i> |
| Random Effects                       |                           |                           |                           |                           |                           |
| Intercept (variance between schools) | 275.4                     | 168.4                     | 83.6                      | 83.7                      | 83.5                      |
| Level-1 (variance within schools)    | 537.9                     | 510.4                     | 454.2                     | 454.2                     | 454.1                     |
| Intraclass Correlation               | .34                       | .25                       | .16                       | .16                       | .16                       |

N= 9999 students and 611 schools.

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$

In Model 4, we see that once all of these potentially confounding variables are controlled, students whose teachers reported giving a higher than average amount of calculator access to students scored an insignificant 0.1 point lower on the NAEP mathematics assessment than did students with teachers reporting less calculator access in their classrooms.<sup>20</sup> Finally, Model 5 controls

<sup>19</sup> Further investigation revealed that although overall achievement was substantially higher in private schools than in public schools, this relationship reversed after controlling for SES and other demographic variables. See Lubienski & Lubienski (2006) regarding a follow-up study.

<sup>20</sup> Interestingly, the *student*-reported calculator variable negatively predicted achievement at grade 4. Teacher-reported calculator use and achievement were more positively related at grade 8, perhaps because advanced classes utilize calculators more often. Again, see Lubienski, Camburn and Shelley (2004) for more information.

for all of these factors and also includes student-level interaction terms to examine whether the relationship between calculator access and achievement differs by student race. None of the interaction terms were significant. The addition of the calculator variables in Models 4 and 5 did not help explain additional variance between or within schools.

### *Summary of HLM Results*

Each of the other 9 teacher-reported instruction-related variables were treated in a likewise manner, with the final results condensed in Table 4 (interaction terms are discussed later).

*Relationship between instruction and achievement.* The HLM coefficient for each instruction-related measure is presented in Table 4, each taken from a model equivalent to Model 4 in Table 3. With the exception of weekly multiple choice assessment use and being “not very knowledgeable” about the NCTM *Standards*, for each variable a positive HLM coefficient indicates that a practice aligned with the NCTM *Standards* positively predicts achievement after controlling for student- and school-level demographics and other potentially confounding variables.

Table 4

*HLM Coefficients of Teacher-Reported Instruction-Related Factors when Predicting Student Mathematics Achievement, After Controlling for Demographic Variables*

| Variable                       | Coefficient                   |
|--------------------------------|-------------------------------|
| Calculators                    | -0.1                          |
| De-emphasize facts & skills    | 1.1                           |
| Collaborative problem solving  | 1.1*                          |
| Non-number curricular emphasis | 1.6**                         |
| Writing about mathematics      | 0.2                           |
| Manipulatives                  | 0.6                           |
| Reasoning                      | 1.0                           |
| Multiple choice assessment use | -0.9 (annual use)             |
|                                | -0.4 (weekly use)             |
| Knowledge of NCTM standards    | 4.2* (very knowledgeable)     |
|                                | 0.13 (not very knowledgeable) |

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$

For each of the three significant coefficients found, the direction of the relationship indicated that NCTM-based instruction and knowledge were positively related to achievement. Specifically, collaborative problem solving, teacher knowledge of the NCTM *Standards*, and having a non-number curricular emphasis were all significant, positive predictors of fourth-grade achievement. The results of the larger study were more striking, in that the five teacher-reported variables found to significantly predict achievement at grade 8 held the same pattern.<sup>21</sup>

*Reduction of race-related gaps with teacher-reported variables.* By comparing the coefficients for Black and Hispanic students before each instructional practice is included in the model (see Table 3, Model 3) and their corresponding coefficients after each practice is added (see Model 4),

<sup>21</sup> Collaborative problem solving and knowledge of the NCTM *Standards* were also significant predictors of achievement in grade 8. In addition, calculator use, an emphasis on reasoning, and a de-emphasis of facts and skills were also significantly, positively related to 8th-grade achievement.

one can examine the extent to which disparities in access to particular instructional practices might account for a portion of the achievement gaps. If an instructional practice correlates strongly with achievement, and if that practice is utilized much more with White students than with Black or Hispanic students, then we might see a substantial improvement in the slopes for Black and Hispanic students once we add the instructional variable to the model. In other words, after controlling for the fact that Black and Hispanic students have less access to such instructional practices, we would see the magnitude of the Black and Hispanic coefficients decrease. However, an examination of the change in race-related coefficients for each teacher-reported instructional variable added revealed that the change was near .1 or less. Even when adding all of the instruction-related variables together in the same model, the change in the slopes was .2 or less at both 4th and 8th grades, indicating a less than 1% change in the 17–30 point gaps.<sup>22</sup>

Hence, these results indicate that the disparities in reform-oriented instruction, as measured in these models by the teacher-reported NAEP data, do not help explain much of the race-related achievement gaps. Yet again, researchers might find a stronger relationship if using more sensitive measures and examining student experiences and growth over several years (Rowan, Correnti, & Miller, 2002). It is worth noting that in the full study, student-reported NCTM-aligned beliefs (math is not simply fact memorization and there are multiple ways to solve problems) were strong, positive predictors of achievement at both 4th and 8th grades. Such beliefs are formed over years of students' experiences learning mathematics. Race-related gaps slightly but significantly decreased when these and other student-reported factors were included in HLM models.

*Interaction effects.* Three interaction effects (Black, Hispanic, and SES) were examined for each of the teacher-reported variables. Of these, only one interaction was significant: Non-number curricular emphasis had a positive interaction with SES, indicating that a non-number curricular emphasis correlated more positively with achievement for higher-SES students than lower-SES students. Specifically, as shown previously in Table 4, a student of average SES having a teacher with non-number emphasis one standard deviation above the mean, scored an average of 1.6 points higher than a student whose teacher reported an average amount of emphasis on non-number topics. However, given that the “non-number X SES” coefficient was 1.2, if a student were 1 standard deviation above the mean in terms of SES, that non-number curricular emphasis advantage would actually be  $1.6 + 1.2 = 2.8$  points. If a student were 2 standard deviations below the mean SES, then the coefficient would actually be  $1.6 - 2.4 = -0.8$  points.<sup>23</sup>

---

<sup>22</sup> Only an additional 1% of the overall variance in achievement was explained when all of the instruction-related variables were added to the demographic model. However, multicollinearity among predictors, as well as the fact that these data are not longitudinal, necessitate caution in interpreting these results.

<sup>23</sup> There were three significant interactions among the *student*-reported variables. Two of these involved SES and might be viewed as following a pattern consistent with the teacher-reported non-number emphasis interaction. The coefficients for student-reported calculator use, and student-reported collaborative problem solving were greater for high-SES students than for low-SES students, suggesting that perhaps these aspects of instruction could further the advantages of high-SES students. However, again, given the cross-sectional (as opposed to longitudinal) nature of the data, and given the number of interactions tested, these findings should be viewed as merely suggestive of issues for further study.

## Discussion

### *Reform-Oriented Instruction, Achievement, And Equity*

This study's descriptive analyses showed relatively few race-related inequities in fourth-graders' access to instructional practices aligned with the NCTM *Standards*. Black and Hispanic students were actually more likely than White students to have a teacher report a strong non-number curricular emphasis and frequent manipulative use. However, consistent with previous findings (Strutchens, et al., 2004), Black and Hispanic students were significantly more likely to be regularly assessed with multiple choice tests than were White students.

This study's HLM analyses determined that several reform-oriented factors were significantly related to student achievement after controlling for SES, race, disability status, gender, and school sector. Specifically, teachers' non-number curricular emphasis, use of collaborative problem solving and knowledge of the NCTM *Standards* were significant, positive predictors of fourth grade mathematics achievement.

Although the primary focus of this article is on the grade 4 results, it is worth noting that at both fourth and eighth grades, in every case when a teacher-reported, reform-oriented instructional factor was significantly related to achievement, the relationship was positive. Additionally, student-reported beliefs aligned with the NCTM Standards were strong, positive predictors of achievement at grades 4 and 8, and such beliefs were more prevalent for White students than for Black and Hispanic students. Given these differences in beliefs, it is very possible that additional race-related instructional disparities exist that are not captured by the NAEP teacher survey items.

Despite the positive relationships between reform-oriented instruction and achievement identified in this study, the overall implications for ways to improve equity are less clear. The reductions in the "slopes" for Black and Hispanic students produced by adding the teacher-reported instructional variables to the models were very small. Additionally, some instructional practices that correlated positively with achievement, such as teacher-reported non-number curricular emphasis and collaborative problem solving, were actually more prevalent for Black and Hispanic students than for White students. Moreover, the few interaction effects that were significant in the full study suggested ways in which NCTM-based practices correlated more positively with achievement for high-SES students than for low-SES students. Instead of illuminating possible causes of achievement gaps, these facts seem to only further complicate the search for instruction-related causes.

Overall, the NCTM-based instructional practices examined in this study related positively to achievement when they related at all. The consistency of this pattern at grades 4 and 8 (as revealed in the full study) would seem to provide encouraging news for reformers. However, this and other results of the study must be interpreted with care, as is discussed in the next section.

### *Limitations*

Given the cross-sectional nature of NAEP data, we cannot be sure whether reform-oriented practices actually *caused* higher achievement, or whether higher-achieving students were more likely to receive reform-oriented instruction. Either case raises important questions about the reasons for the relationship and its ultimate effects on sustaining or furthering achievement disparities.

There are several additional cautions to be discussed. First, NAEP classroom practices data are based on teacher self-reports for that school year only. The accuracy of teachers' memory of

practices utilized throughout the year and perceived pressure to portray instruction in particular ways could have affected teachers' responses. Additionally, the three- or four- point scales used on many of the teacher survey items were rough and perhaps insensitive to important differences in teacher practices. Many important questions were not asked that might move beyond surface features of instruction (e.g., manipulative use) to probe at more fundamental instructional issues (e.g., the extent to which instruction builds upon and centers around student understanding), as well as to identify larger, structural inequities (e.g., school funding). Still, the fact that when put in a factor analysis, most teacher-reported instruction-related variables "clumped" with each other in sensible ways, and the fact that several significant relationships between student demographics and instructional factors were found, indicate that the NAEP mathematics teacher survey questions are, indeed, measuring some important aspects of variability in instruction.

Again, there is no measure of prior achievement in NAEP, and so it is students' overall achievement, and not growth in achievement, that serves as the outcome variable. This limitation, combined with the limits of the teacher-reported data noted above, suggest that this study may be overly conservative in determining the strength of impact that instructional measures can have on both student learning and on achievement gaps. If teacher practices were measured with more sensitive measures over time, and if the data allowed for examinations of student achievement gains, it is likely that we would see a greater instructional impact on achievement and race-related gaps than what is indicated here (Rowan, Correnti, & Miller, 2002). On the other hand, standard errors for instruction-related coefficients were perhaps smaller (less conservative) than they would have been if the clustering of students within classrooms was accounted for in the models (i.e., if teachers could have been treated at a "classroom level"). Hence, again, the results of this study should be viewed as merely suggestive of relationships that are important to explore further using in-depth, longitudinal methods that can take student, teacher/classroom and school levels into consideration.

### *Comparison across Two Studies*

We now return to the question of how this study compares with that of Wenglinsky (2004), who also used HLM with the 2000 NAEP mathematics data to examine relationships among instruction, achievement and equity. Some findings of the two studies were complementary. For example, this study found a positive relationship between teachers' non-number curricular emphasis and achievement. Wenglinsky also obtained positive coefficients for teachers' emphasis on geometry, measurement and algebra (although only the coefficient for geometry was significant). There were also consistencies in some factors that did not correlate with achievement in either study, including manipulative use and writing about mathematics, as well as teachers' college major and degree (which were subsequently deleted from models in this study due to their lack of significance).

However, vastly different conclusions were reached about the potential for particular instructional practices to close achievement gaps. This study identified only weak, insignificant interactions between particular instructional practices and Black and Hispanic student achievement. Although some significant relationships between instructional practices and overall achievement were found, these relationships did not vary significantly by student race. Additionally, all relationships between instruction and achievement found in this study are interpreted with great caution due to the cross-sectional nature of NAEP data.

In contrast, Wenglinsky concluded from his study "that a series of instructional practices, when used in concert, can substantially reduce both the Black-White and Latino-White achievement gaps" (p. 3). Specifically, Wenglinsky asserted that frequent test taking enlarges the Black-White gap,

and that an emphasis on measurement helps reduce the gap. (Although not the main point of the concerns raised here, it is worth noting that the actual coefficients Wenglinsky provided appear to indicate the opposite—an emphasis on measurement appeared to predict a larger Black-White gap, while frequent testing predicted a smaller gap.) He also concluded that an emphasis on data analysis is particularly beneficial for Hispanic students.

There are four fundamental differences between the two studies that underlie their divergent conclusions. First, Wenglinsky aggregated teacher-level data to the school level, whereas this study treated teacher-level data at the student-level. Again, there was no classroom or teacher identification code in the NAEP data, making it nearly impossible to treat “classroom” as a separate level in HLM. Confidentiality concerns might partially underlie NAEP’s exclusion of teacher codes, but another reason is that students—not teachers—are randomly selected within a school. Therefore, NAEP experts recommend connecting teachers with individual student data when making claims about teachers. This was the approach taken in this study, compatible with its primary focus on disparities among students’ classroom experiences and achievement. On the other hand, given Wenglinsky’s primary focus on NCLB and within-school practices and achievement gaps, his decision to aggregate instructional practices to the school level is certainly conceptually defensible, having the potential to create a stronger measure of the general instructional climate of the school. The interactions between student race and instruction were then treated as cross-level interactions in Wenglinsky’s study, which fit well with his focus on within-school gaps (whereas in this study, the interactions were treated at the student-level). Overall, the difference in treatment of the teacher data when designing the HLM models (analyzing it at the student versus the school level) may be one contributor to the differences in the studies’ findings.

However a second and more important difference between the studies is the number of variables included simultaneously in the models. The study reported here utilized factor analysis to reduce roughly 30 instruction-related variables to nine factors, which were then each examined in separate models. In contrast, in addition to several demographic measures, Wenglinsky included 20 variables pertaining to teaching practices and 3 variables pertaining to teacher background in his model to predict the main intercept, and he also used the 23 teacher-related variables to predict the within-school “slope” (or gap) for both Black and Hispanic students. Wenglinsky found that the slopes for Black and Hispanic students were significant before adding the teacher-related variables but no longer significant after adding those variables. Wenglinsky then concluded, “Thus, by including the 20 instructional practices, the second HLM can explain away the entire within-school racial gap” (p. 16).

Wenglinsky’s full model, then, involved the determination of over 70 coefficients. Forty-six of these were predicting the Black or Hispanic slope, the primary focus of his study. By chance alone we would expect roughly 4 or 5 of those 46 predictors to be statistically significant at the  $p < .1$  level (the base level of significance he used), with 2 or 3 of those significant at the  $p < .05$  level. In fact, his model identified only 3 significant predictors of the Black or Hispanic slopes (2 at the .05 level, and 1 at the .1 level). Hence, it is quite possible that these 3 variables are “false positives.” In fact, it is clear that Wenglinsky’s full model is problematic, as evidenced by inflated standard errors and the fact that the Hispanic slope went from being -8 (with a standard error of 1) in his base model, to a positive 27 points (with standard error of 22). The slope for Black students went from -16 (standard error of 1) to -9 (standard error of 26). It is worth noting that the Black-White gap of -9 was considered “eliminated” because the gap was not significant, yet the standard error had become so large that even the original Black-White gap would not be significant. Again, the huge reversal of the Hispanic slope suggests serious instability in the model, likely caused by the large number of predictors, many of which are collinear (as evidenced by the results of the factor analyses in the study reported here).

Hence, while the composites of instructional measures used in the study reported here can be more difficult to interpret than the individual variables included in Wenglinsky's models, his full model reveals the danger of including too many variables in an HLM, particularly at the school level, as Wenglinsky, himself notes in a footnote: "degrees of freedom are sharply reduced by including so many school-level independent variables" (p. 16).

Other, minor differences between the studies' methods could be discussed, including the exact calculations of school race or SES measures, or the fact that some variables, such as disability or time on math were included in one study but not the other. However, the two remaining differences between the studies that merit discussion lie not in their methods of analysis but in their interpretation of the results.

First, NAEP data are cross-sectional—not longitudinal—and therefore not intended for drawing causal conclusions regarding instruction and achievement. Wenglinsky himself notes this in his brief, isolated discussion of the limitations of his study, explaining: "This means that nothing is known about the causal direction of the results" (pp. 16–17). And yet causal language and conclusions are prevalent throughout the remainder of the article, beginning with the abstract in which he states that he uses HLM with NAEP data "to identify instructional practices that reduce the achievement gap. It finds that, even when taking student background into account, instructional practices can make a substantial difference."

Wenglinsky's optimistic conclusion that, according to his results, school administrators can "succeed at closing the racial achievement gap in their schools," (p. 17) is unwarranted. Again, a correlation between particular practices and achievement may not be causal, particularly given that another plausible explanation exists—i.e., that higher achieving students might tend to receive different instruction than lower-achieving students. And again, the large number of predictors in Wenglinsky's full model should also raise major concerns about drawing conclusions from the particular relationships identified.

Second, even if one *could* conclude from Wenglinsky's study that particular instructional practices reduced the within-school Black-White and Hispanic-White gaps to 0, one must interpret this with the understanding that SES was controlled for in the models, and therefore it is the race-related "leftover" gap (the part not related to SES) within schools that was reduced in the models. Hence, in practice, there would still be very large within-school gaps between Black and White students in most schools, as well as between Hispanic and White students, given the strong correlation between race and SES. Additionally, the focus of NCLB and Wenglinsky's study on within-school gaps ignores race- and SES-related gaps between schools, which dangerously places responsibility for gap reductions on school personnel and ignores larger societal inequities, such as persistent disparities in community resources that schools alone cannot overcome (Berliner, 2005).

### *Implications for Future Research*

#### **Uses and Abuses of Cross-Sectional Data**

One can understand why researchers utilizing NAEP and other cross-sectional data are tempted to overstate rather than understate the conclusions that can be drawn. Soft claims surrounded by a sea of caveats tend to be ignored by publishers, the popular media, and policy makers. This dynamic points toward the need for studies indicating no relationship between important variables to be reported along with those with more exciting conclusions. Critiques of NAEP's cross-sectional nature also raise questions about the usefulness (or lack thereof) of NAEP and other similar large-scale data sets (Christensen & Angel, 2005; Lubienski & Lubienski, 2006).

One might wonder, if causal claims cannot be made from studies of NAEP data, then what good is NAEP?

In the study reported here, NAEP data were useful for examining whether reform-oriented instructional practices were distributed equally across U.S. students, regardless of race. This is aligned with NAEP's strength—describing current achievement and instruction-related patterns on a large, nationally-representative scale. Analyses of NAEP data can also shed light on which instructional practices do and do not correlate with student achievement while controlling for many potential confounding variables. However, a measure of prior achievement is not available in NAEP, and the causal order of any relationships identified will be unclear. Because of their potential for widespread attention and influence on policy, large-scale studies, in particular, must be communicated with care, with proper cautions regarding the limitations of the results emphasized.

In the case of the study reported here, whether particular NCTM-based practices caused higher achievement, or whether high-achieving students were more likely to be taught with NCTM-based practices is unclear. Still, the relationships identified raise important questions for further research within classrooms. As Wenglinsky also notes, NAEP analyses cannot replace studies involving in-depth classroom observations.

### Further Research on Equity

After multiple reform efforts aimed at changing mathematics instruction and reducing inequities, much work remains. One finding that is clear in both Wenglinsky's study and this study is that there are large race- and SES-related achievement gaps, and even after controlling for SES using multiple demographic variables, the unexplained race-related gap within schools is disturbingly large.<sup>24</sup>

NAEP offers one avenue for examining disparities in achievement and classroom practices. The patterns identified in this study suggest directions for additional longitudinal and qualitative studies that examine causes of, and ways to address, the patterns identified here. Overall, researchers should continue to examine achievement disparities, considering instructional factors identified in this study, as well as other potential influences not considered here, such as differential access to various resources at both home and school.

## References

- Anyon, J. (1981). Social class and school knowledge. *Curriculum Inquiry*, 11, 3–42.
- Banks, J. A. (1988). Ethnicity, class, cognitive, and motivational styles: Research and teaching implications. *Journal of Negro Education*, 57(4), 452–466.
- Berliner, David C. (2005). Our impoverished view of educational reform. *Teachers College Record*, 108, 949–995.

---

<sup>24</sup> Rothstein (2004) rightly points out that the SES variables used in studies such as these are limited because they do not measure important historical and contextual factors, such as accumulated family wealth. It is likely that a better SES measure would account for a greater proportion of race-related achievement gaps. Still, the measure used in this study is more comprehensive than SES variables typically utilized in NAEP reports.



- Boaler, J. (2002). Learning from teaching: Exploring the relationship between reform curriculum and equity. *Journal for Research in Mathematics Education*, 33(4), 239–258.
- Braswell, J. S., Daane, M. C., & W. S. Grigg (2003). *The nation's report card: Mathematics highlights 2003*. Washington, DC: U.S. Department of Education. Institute for Education Sciences, National Center for Education Statistics.
- 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., Lee, V. E., & Holland, P. B. (1993). *Catholic schools and the common good*. Cambridge MA: Harvard University Press.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park: Sage Publications.
- Campbell, J. R., Hombo, C. M., & J. Mazzeo (2000). *NAEP 1999 trends in academic progress: Three decades of student performance, NCES 2000–469*, Washington, DC: U.S. Department of Education.
- Christensen, J. & Angel, L. (2005). What NAEP can't tell us about charter school effectiveness. *Education Week*, 25(14), 35, 37.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2<sup>nd</sup> ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cook, P. & Ludwig, J. (1998). The burden of acting White. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 375–400). Washington, DC: Brookings Institution Press.
- Cooper, B & Dunne, M. (2000). *Assessing children's mathematical knowledge: Social class, sex and problem-solving*. Buckingham: Open University Press, UK.
- Duberman, L. (1976). *Social inequality: Class and caste in America*. Philadelphia, J.B. Lippincott.
- Fennema, E., Carpenter, T. P., Jacobs, V. R., Franke, M. L., & Levi, L. (1998). A longitudinal study of gender differences in young children's mathematical thinking. *Educational Researcher*, 27(5), 6–11.
- Ferguson, R. F. (1998a). 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.
- Ferguson, R. F. (1998b). Comment on Cook & Ludwig's "The burden of 'acting White': Do Black adolescents disparage academic achievement?" In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 394–397). Washington, DC: Brookings Institution Press.

- Ferguson, R. F. (1998c). Can schools narrow the Black-White test score gap?. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 318–374). Washington, DC: Brookings Institution Press.
- Foegen, A. (2004). *Students with disabilities and the 2000 NAEP mathematics assessment: achievement levels and sample characteristics*. Paper presented at the annual meeting of the American Educational Research Association, San Diego.
- Hickey, D. T., Moore, A. L., & Pellegrino, J. W. (2001). The motivational and academic consequences of elementary mathematics environments: Do constructivist innovations and reforms make a difference? *American Educational Research Journal*, 38(3), 611–652.
- 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 National 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.
- Kloosterman, P., & Lester, F. K., Jr. (Eds.) (2004). *Results and interpretations of the 1990 through 2000 mathematics assessments of the National Assessment of Educational Progress*. Reston, VA: National Council of Teachers of Mathematics.
- Ladson-Billings, G. (1997). It doesn't add up: African American students' mathematics achievement. *Journal for Research in Mathematics Education*, 28(6), 697–708.
- Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity? *Educational Researcher*, 31(1), 3–12.
- Loveless, T., & Diperna, P. (2000). *How well are American students learning? Focus on math achievement* (Brown Center Report, Vol. 1, No. 1). Washington, DC: Brookings Institution.
- Lubienski, C., & Lubienski, S. T. (2006). *Charter, private, public schools and academic achievement: New evidence from NAEP mathematics data*. New York: National Center for the Study of Privatization in Education, Teachers College, Columbia University. Retrieved May 29, 2006, from <http://www.ncspe.org/readrel.php?set=pub&cat=126>.
- Lubienski, S. T. (2000a). A clash of class cultures? Students' experiences in a discussion-intensive seventh-grade mathematics classroom. *Elementary School Journal*, 100, 377–403.
- Lubienski, S. T. (2000b). Problem solving as a means toward “mathematics for all”: An exploratory look through a class lens. *Journal for Research in Mathematics Education*, 31(4), 454–482.

- Lubienski, S. T. (2002). A closer look at Black-White mathematics gaps: Intersections of race and SES in NAEP achievement and instructional practices data, *Journal of Negro Education*, 71(4), 269–287.
- 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., Camburn, E., & Shelley, M. C. (2004). *Reform-oriented mathematics instruction, achievement, and equity: Examinations of race and SES in 2000 Main NAEP Data*. Report submitted to the National Center for Education Statistics.
- Lubienski, S. T. & Lubienski, C. (2006). What NAEP can tell us about school achievement. *Education Week*, 25(26), 28, 30.
- Lubienski, S. T., McGraw, R., & Strutchens, M. (2004). NAEP findings regarding gender: Mathematics achievement, student affect, and learning practices. In P. Kloosterman., & F. K. Lester, Jr. (Eds.) *Results and interpretations of the 1990 through 2000 mathematics assessments of the National Assessment of Educational Progress* (pp. 305–336). Reston, VA: NCTM.
- Lubienski, S. T. & Shelley, M. C. (2003). *A closer look at U.S. mathematics instruction and achievement: Examinations of race and SES in a decade of NAEP data*. Paper presented at the American Educational Research Association, Chicago.
- Means, B. & Knapp, M. S. (1991, December). Cognitive approaches to teaching advanced skills to educationally disadvantaged students. *Phi Delta Kappan*, 73, 282–289.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation Standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics, (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.
- Newmann, F. M., & Wehlage, G. G. (1995). *Successful school restructuring: A report to the public and educators*. Washington, DC: Office of Educational Research and Improvement.
- Ogbu, J. U. (1995). Understanding cultural diversity and learning. In J. A. Banks & C. A. McGee Banks (Eds.), *Handbook of research on multicultural education* (pp. 582–593). New York: Macmillan.
- Peng, S. S., Wright, D. & S. T. Hill (1995). *Understanding racial-ethnic differences in secondary school science and mathematics achievement*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Raudenbush, S.W., Fotiu, R., & Cheong, Y.F. (1998). Inequality of access to educational opportunity: a national report card for eighth grade math. *Educational Evaluation and Policy Analysis*, 20, 253–268.
- Reys, R., Reys, B., Lapan, R., Holliday, G., and Wasman, D. (2003). Assessing the impact of Standards-based middle grades mathematics textbooks on student achievement. *Journal for Research in Mathematics Education*, 34(1), 74–95.
- Riordan, J. E., and Noyce, P. E. (2001). The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematical Education*, 32(4), 368–398.
- Rothstein, R. (2004). *Class and schools: Using social, economic, and educational reform to close the Black-White achievement gap*. Washington, DC: Economic Policy Institute.
- Rowan, B., Correnti, R. & Miller, R. J. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the *Prospects* study of Elementary Schools. *Teachers College Record*, 104, 1525–1567.
- Rubin, D. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley.
- Schoenfeld (2002). Making mathematics work for all children: Issues of standards, testing, and equity. *Educational Researcher*, 31(1), 13–25.
- Senk, S. & Thompson, D. (Eds.) (2003). *Standards-based school mathematics curricula: What are they? What do students learn?* Hillsdale, NJ: Lawrence Earlbaum Associates, Inc.
- Silver, E., Smith, M. S., & Nelson, B. S. (1995). The QUASAR Project: Equity concerns meet mathematics reforms in the middle school. In W. G. Secada, E. Fennema, & L. B. Adajian (Eds.), *New directions in equity in mathematics education* (pp. 9–56). New York: Cambridge University Press.
- 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.
- Stiff, L. (1990). African-American students and the promise of the *Curriculum and Evaluation Standards*. In T. J. Cooney & C. R. Hirsch (Eds.), *Teaching and learning mathematics in the 1990s* (Yearbook of the National Council of Teachers of Mathematics, pp. 152–157), Reston, VA: NCTM.
- Strutchens, M., Lubienski, S. T., McGraw, R. & Westbrook, S. K. (2004). NAEP findings regarding race/ethnicity: Students' performance, school experiences, attitudes/beliefs and family influences. In P. Kloosterman., & F. K. Lester, Jr. (Eds.) *Results and interpretations of the 1990 through 2000 mathematics assessments of the National Assessment of Educational Progress* (pp. 269–304). Reston, VA: NCTM.

- 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.
- Swanson, C. B., & Stevenson, D. L. (2002, Spring). *Standards-based reform in practice: Evidence on state policy and classroom instruction from the NAEP state assessments. Educational Evaluation and Policy Analysis, 24*(1), 1–27.
- Tate, W. F. (1997). Race-ethnicity, SES, gender, and language proficiency in mathematics achievement: An update. *Journal for Research in Mathematics Education, 28*, 652–679.
- Thompson, B. (2002). “Statistical,” “practical,” and “clinical”: How many kinds of significance do counselors need to consider? *Journal of Counseling and Development, 80*, 64–71.
- Weis, L. (Ed.) (1988). *Class, race, & gender in American education*, Albany: State University of New York Press.
- Wenglinsky, H. (2004, November 23). Closing the racial achievement gap: The role of reforming instructional practices. *Educational Policy Analysis Archives, 12*(64). Retrieved December 2, 2004, from <http://epaa.asu.edu/epaa/v12n64/>.

**About the Author**

**Sarah Theule Lubienski**

University of Illinois at Urbana-Champaign

Email: [stl@uiuc.edu](mailto:stl@uiuc.edu)

Sarah Theule Lubienski is an associate professor in the College of Education at the University of Illinois, Urbana-Champaign. Her scholarship centers around intersections of education and equity, with a particular focus on mathematics achievement, instruction, and reform. Her recent NAEP-related research has been supported by two Secondary Analysis Grants, funded by the National Center of Education Statistics. She is the chairperson of the NAEP Studies Special Interest Group of the American Educational Research Association.

**EDUCATION POLICY ANALYSIS ARCHIVES** <http://epaa.asu.edu>

**Editor: Sherman Dorn, University of South Florida**

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Sherman Dorn, [epaa-editor@shermadorn.com](mailto:epaa-editor@shermadorn.com).

**Editorial Board**

**Michael W. Apple**

University of Wisconsin

**Robert Bickel**

Marshall University

**Casey Cobb**

University of Connecticut

**Gunapala Edirisooriya**

Youngstown State University

**Gustavo E. Fischman**

Arizona State University

**Gene V Glass**

Arizona State University

**Aimee Howley**

Ohio University

**William Hunter**

University of Ontario Institute of Technology

**Benjamin Levin**

University of Manitoba

**Les McLean**

University of Toronto

**Michele Moses**

Arizona State University

**Michael Scriven**

Western Michigan University

**John Willinsky**

University of British Columbia

**David C. Berliner**

Arizona State University

**Gregory Camilli**

Rutgers University

**Linda Darling-Hammond**

Stanford University

**Mark E. Fetler**

California Commission on  
Teacher Credentialing

**Richard Garlikov**

Birmingham, Alabama

**Thomas F. Green**

Syracuse University

**Craig B. Howley**

Ohio University

**Daniel Kallós**

Umeå University

**Thomas Mauhs-Pugh**

Green Mountain College

**Heinrich Mintrop**

University of California, Berkeley

**Anthony G. Rud Jr.**

Purdue University

**Terrence G. Wiley**

Arizona State University

**EDUCATION POLICY ANALYSIS ARCHIVES**  
**English-language Graduate-Student Editorial Board**

**Noga Admon**  
New York University

**Jessica Allen**  
University of Colorado

**Cheryl Aman**  
University of British Columbia

**Anne Black**  
University of Connecticut

**Marisa Cannata**  
Michigan State University

**Chad d'Entremont**  
Teachers College Columbia University

**Carol Da Silva**  
Harvard University

**Tara Donahue**  
Michigan State University

**Camille Farrington**  
University of Illinois Chicago

**Chris Frey**  
Indiana University

**Amy Garrett Dikkers**  
University of Minnesota

**Misty Ginicola**  
Yale University

**Jake Gross**  
Indiana University

**Hee Kyung Hong**  
Loyola University Chicago

**Jennifer Lloyd**  
University of British Columbia

**Heather Lord**  
Yale University

**Shereza Mohammed**  
Florida Atlantic University

**Ben Superfine**  
University of Michigan

**John Weathers**  
University of Pennsylvania

**Kyo Yamashiro**  
University of California Los Angeles



## Archivos Analíticos de Políticas Educativas

### Associate Editors

**Gustavo E. Fischman & Pablo Gentili**

Arizona State University & Universidade do Estado do Rio de Janeiro

Founding Associate Editor for Spanish Language (1998—2003)

**Roberto Rodríguez Gómez**

### Editorial Board

**Hugo Aboites**

Universidad Autónoma  
Metropolitana-Xochimilco

**Dalila Andrade de Oliveira**

Universidade Federal de Minas  
Gerais, Belo Horizonte, Brasil

**Alejandro Canales**

Universidad Nacional Autónoma  
de México

**Erwin Epstein**

Loyola University, Chicago,  
Illinois

**Rollin Kent**

Universidad Autónoma de  
Puebla. Puebla, México

**Daniel C. Levy**

University at Albany, SUNY,  
Albany, New York

**María Loreto Egaña**

Programa Interdisciplinario de  
Investigación en Educación

**Grover Pango**

Foro Latinoamericano de  
Políticas Educativas, Perú

**Angel Ignacio Pérez Gómez**

Universidad de Málaga

**Diana Rhoten**

Social Science Research Council,  
New York, New York

**Susan Street**

Centro de Investigaciones y  
Estudios Superiores en  
Antropología Social Occidente,  
Guadalajara, México

**Antonio Teodoro**

Universidade Lusófona Lisboa,

**Adrián Acosta**

Universidad de Guadalajara  
México

**Alejandra Birgin**

Ministerio de Educación,  
Argentina

**Ursula Casanova**

Arizona State University,  
Tempe, Arizona

**Mariano Fernández**

**Enguita** Universidad de  
Salamanca. España

**Walter Kohan**

Universidade Estadual do Rio  
de Janeiro, Brasil

**Nilma Limo Gomes**

Universidade Federal de  
Minas Gerais, Belo Horizonte

**Mariano Narodowski**

Universidad Torcuato Di  
Tella, Argentina

**Vanilda Paiva**

Universidade Estadual Do  
Rio De Janeiro, Brasil

**Mónica Pini**

Universidad Nacional de San  
Martín, Argentina

**José Gimeno Sacristán**

Universidad de Valencia,  
España

**Nelly P. Stromquist**

University of Southern  
California, Los Angeles,  
California

**Carlos A. Torres**

UCLA

**Claudio Almonacid Avila**

Universidad Metropolitana de  
Ciencias de la Educación, Chile

**Teresa Bracho**

Centro de Investigación y  
Docencia Económica-CIDE

**Sigfredo Chiroque**

Instituto de Pedagogía Popular,  
Perú

**Gaudêncio Frigotto**

Universidade Estadual do Rio  
de Janeiro, Brasil

**Roberto Leher**

Universidade Estadual do Rio  
de Janeiro, Brasil

**Pia Lindquist Wong**

California State University,  
Sacramento, California

**Iolanda de Oliveira**

Universidade Federal  
Fluminense, Brasil

**Miguel Pereira**

Catedrático Universidad de  
Granada, España

**Romualdo Portella do**

**Oliveira**

Universidade de São Paulo

**Daniel Schugurensky**

Ontario Institute for Studies in  
Education, Canada

**Daniel Suarez**

Laboratorio de Políticas  
Públicas-Universidad de  
Buenos Aires, Argentina

**Jurjo Torres Santomé**

Universidad de la Coruña,  
España