DOCUMENT RESUME

ED 413 348 TM 027 656

AUTHOR	Mayer, Daniel P.
TITLE	New Teaching Standards and Old Tests: Dangerous Mismatch?
SPONS AGENCY	College Board, New York, NY.
PUB DATE	1997-03-00
NOTE	51p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Algebra; Educational Change; *High School Students; High Schools; Mathematics Achievement; *Mathematics Instruction; Middle Schools; *Standards; State Programs; *Test Results; *Testing Programs
IDENTIFIERS	*Middle School Students; *National Council of Teachers of Mathematics

ABSTRACT

As almost every state attempts to reform mathematics instruction by implementing new teaching standards, state testing practices remain largely unchanged. Is there a mismatch between these new standards and the old tests? This question is investigated by examining whether middle school and high school algebra students taught in a manner consistent with the National Council of Teachers of Mathematics (NCTM) "Professional Standards" performed differently on three standardized algebra assessments than students taught in traditional classrooms. The data come from 94 teachers, 2,369 students, and 40 schools in 1 of the nation's largest school districts. Results indicate that a mismatch does not exist between the "Standards" and the old tests. In fact, middle school algebra students whose teachers spent more time using the NCTM teaching approach had higher growth rates than students whose teachers spent less time using the approach. However, students with higher ability levels benefited more. The growth rates of the lowest achieving students, the high school students (who were disproportionately poor and black), were not helped or hindered by the NCTM teaching approach. This study provides policymakers with evidence that traditional multiple choice tests do not directly undermine the standards movement in this one school district. On the other hand, old tests will not provide teachers of low-achieving students with any incentive to adopt the "Standards." (Contains 2 figures, 8 tables, and 50 references.) (Author)

New Teaching Standards and Old Tests: Dangerous Mismatch?

Daniel P. Mayer
Mathematica Policy Research, Inc.

Paper presented at the 1997 Annual Meeting of the American Educational Research Association in Chicago.

Research for this paper was supported by The College Board. Opinions and conclusions are those of the author and do not necessarily reflect the views of the supporting agency. Please send comments to Daniel Mayer at 7215 Holly Ave., Takoma Park, MD 20912. E-mail: dmayer@mathematica-mpr.org.

New Teaching Standards and Old Tests: Dangerous Mismatch?

Abstract

As almost every state attempts to reform mathematics instruction by implementing new teaching standards, state testing practices remain largely unchanged. Is there a mismatch between these new standards and the old tests? This question is investigated by examining whether middle and high school algebra students taught in a manner consistent with the National Council for Teachers of Mathematics Professional Standards performed differently on three standardized algebra assessments than students taught in traditional classrooms. The data come from 94 teachers, 2,369 students, and 40 schools in one of the nation's largest school districts.

Results indicate that a mismatch does not exist between the Standards and the old tests. In fact, middle school algebra students whose teachers spent more time using the NCTM teaching approach had higher growth rates than students whose teachers spent less time using the approach. However, students with higher ability levels benefited more. The growth rates of the lowest achieving students, the high school students (who are disproportionately black and poor), were not helped or hindered by the NCTM teaching approach.

This study provides policy makers with evidence that traditional multiple choice tests do not directly undermine the standards movement in this one school district. On the other hand, old tests will not provide teachers of low-achieving students with any incentive to adopt the Standards.

Introduction

The latest wave of ambitious education reforms in the United States may be undermined by a potential discontinuity: as almost every state attempts to reform mathematics instruction by implementing new teaching standards (Blank & Pechman, 1995), state testing practices remain largely unchanged (Blank, Hemphill, Sardina, Langesen, & Brathwaite, 1995). Advocates for new standards claim that old tests "no longer suffice" (National Council of Teachers of Mathematics, 1989, p.192). The old tests may undermine the standards-based reform efforts in either of two ways. First, because the new standards ask teachers to emphasize skills not measured on the old tests, student test scores might show no improvement, or even decline. Even though students may be mastering important (albeit unmeasured) skills, a flat or declining test score trend on the old tests could derail the reform efforts. The students, teachers, and schools who choose to implement the standards may ironically be penalized for their efforts. Ultimately, this could lead to the unraveling of the standards movement. Second, teachers may anticipate this mismatch and therefore refuse to implement the new standards in the first place.

In this study, I examine the first scenario by looking over the course of one year at how algebra students taught by teachers using the new standards perform on traditional tests relative to their counterparts taught in a more traditional manner.

State curriculum frameworks, the federal government's Goals 2000 initiative, and the rapid-fire succession of newly-minted standards documents produced by professional curriculum associations are all testaments to the prominence of the standards movement. This movement reflects a profound shift in educational policy. Historically, education reforms have tinkered at the edges of the educational process; now policy makers are focusing on its very heart. "In its two and quarter centuries, the United States has never [until now] had explicit education content…goals" (Marshall, Fuhrman, & O'Day, 1994, p. 12). Even the extensive reform efforts of the 1970s and 1980s remained aloof from curriculum and teaching practices. During those decades policy makers tried to improve schooling by adjusting resource allocations (e.g. striving for racial balance and financial equity) and by setting outcome goals (e.g. setting minimum course requirements and implementing minimum competency tests). The perceived failures of these policies arguably has led to the country's current enthusiasm for educational standards aimed at influencing teaching practice and the curriculum.

Mathematics standards are playing a significant role in the standards movement. Most states have recently created or revised mathematics curriculum frameworks with explicit recommendations regarding teaching practices which are heavily influenced by the National Council of Teachers of Mathematics' (NCTM) Professional Standards for Teaching Mathematics (Blank & Pechman, 1995). This should come as no surprise as NCTM was one of the earliest and most important players in the development of curriculum and teaching standards (National Council of Teachers of Mathematics, 1989). The ideas presented in these standards undergird not only the state frameworks, but also other prominent science, mathematics, and technology education reform movements throughout the United States and other developed countries (Black & Atkin, 1996).

The Standards[1] argue that optimal learning of mathematics requires that teachers place less emphasis on memorization of facts and mastery of routine skills and greater weight on application, reasoning, and conceptual understanding. The Curriculum and Evaluation Standards state that for students to "understand what they learn, they must enact for themselves verbs that permeate the mathematics curriculum: 'examine,' 'represent,' 'transform,' 'solve,' 'apply,' 'prove,' 'communicate.' This happens most readily when students work in groups, engage in discussion, make presentations, and in other ways take charge of their own learning" (National Council of Teachers of Mathematics, 1989, pp. 58-59).

The successful implementation of the NCTM approach could provide future economic rewards for students because employers are currently requiring that their workers have higher mathematics skills than in the past and some of the most important conceptual skills which are emphasized by the NCTM approach (e.g. the ability to solve problems, to make conjectures, and to communicate both verbally and in writing) are increasingly valued in the workforce (Murnane & Levy, 1996).

But these benefits may not accrue if a mismatch between the Standards and old tests exists. The NCTM is clearly concerned about this possibility. It warned in its earliest Standards document that "[i]n an instructional environment that demands a deeper understanding of mathematics, test instruments that call for only the identification of single correct responses no longer suffice" (National Council of Teachers of Mathematics, 1989, p.192, emphasis added).

Since issuing this warning, only a few states and school districts have experimented with alternative assessments such as portfolios and open-ended question tests. Traditional assessments continue to represent the status quo (Blank et al., 1995) and there are few signs of change. In fact, some of the states that were experimenting with alternative assessments (e.g. California, Vermont, and Kentucky) are now (for technical, economic, and political reasons) moving back toward traditional assessments (Kirst & Masseo, 1996; Lawton, 1997). The current testing environment therefore stands in stark contrast to the one envisioned by the creators of the Standards. Given the prominence of the standards reform movement in general, and the NCTM Standards in particular, this makes the mismatch question all the more compelling.

Is there evidence that a mismatch exists? The NCTM does not present any, but critics of the Standards argue that research suggests that the NCTM teaching approach will lower standardized test scores of students, especially low-income and low-skilled students (e.g.Hirsch, 1996). Unfortunately, as I show below, the research base both supporting and countering this claim has serious design flaws and offers inconsistent findings. These limitations are unsettling given the prominence of the NCTM reform effort. Policy makers, educators, researchers, and parents need to know: Do students taught in NCTM-like classrooms perform differently on standardized assessments than students taught in traditional classrooms?

This study seeks to answer this question in order to provide insight into the broader policy question of whether or not standards-based educational reform initiatives can succeed in today's testing environment. For teachers to adopt the new approaches they must believe that they will prove beneficial, but arguments of long-term economic payoffs may not be an effective catalyst for change. The more immediate way teachers determine their teaching approach is by noting how their teaching impacts student performance on the assessments given in their school (Koretz, Linn, Dunbar, & Shepard, 1991; Resnick & Resnick, 1991). These assessments carry weight with teachers since they frequently determine a student's immediate opportunities (e.g. access to advanced mathematics courses and higher education) and, in some cases, even a teacher's (e.g. in cases where individual teachers or schools are rewarded or penalized for test score performance).

I explore the mismatch question within the context of eighth and ninth grade introductory algebra classrooms. Focusing on secondary mathematics instruction is important because it has traditionally been understudied within the context of this domain of research. Looking at algebra instruction is significant because algebra is recognized as a "gatekeeper" to future opportunities. Students taking algebra are more likely to take more advanced mathematics in high school and to have higher mathematics performance by the end of high school (e.g. Smith, 1996; Stevenson, Schiller, & Schneider, 1994), and taking algebra in high school increases the odds that students make it to college (e.g. Pelavin & Kane, 1990).

This study includes both middle school (eighth grade) and high school (ninth grade) algebra classrooms because the relationship between teaching and learning in algebra classrooms could differ dramatically depending upon grade level. This could happen for two related reasons. First, the academically advanced students take algebra in eighth grade, while most others take it in ninth. This is both a national pattern (National Center for Education Statistics, 1992) and a pattern in the sample of students included in this study. Second, this study, and prior research (e.g. Metz, 1978; Oakes, 1985; Raudenbush, Rowan, & Cheong, 1993), has found that teachers use different teaching approaches depending on the types of children they instruct. Classrooms consisting of advanced students tend to receive an instructional approach more in line with the Professional Standards, which emphasizes higher order thinking skills. This sorting of students into various instructional "tracks" could result in academic inequities (McDonnell, 1995; Oakes, 1985). If these differences (in both student mathematical ability and teacher instructional style) affect the relationship between teaching style and student performance on standardized assessments, then the answer to the mismatch question could well depend upon the grade level of the students.

Limitations of the existing research base

As noted above, the secondary mathematics instruction research base is both thin and flawed. It is thin, in large part, because most researchers have been interested in exploring whether the Professional Standards work. Answering the effectiveness question is entirely different from exploring whether a mismatch between the Professional Standards and traditional tests exits. Since the NCTM and many researchers argue that standardized tests are not a valid measure of the progress made in NCTM-like classrooms, answering the effectiveness question requires that students be measured with non-standardized tests. Researchers interested in this question have developed special assessments for their studies (e.g. Campbell, 1995; Hiebert & Wearne, 1993).

But only studies which look at the relationship between NCTM-type teaching practices and standardized test scores are relevant to this study and the findings from them are inconsistent. While some argue that the teaching approaches which mirror those endorsed by the NCTM lower standardized test scores (e.g. Hirsch, 1996) others have found that they boost them (e.g. Knapp & Associates, 1995). Hirsch bases his claims on findings from early "process-product research" (i.e. research which links classroom processes like teaching practice to products like test scores), one of the most influential research traditions in the study of teaching (Brophy & Good, 1986; Shulman, 1986). While numerous studies point to the same conclusion, generalizations from these early process-product studies should be made with caution because the research was conducted prior to the development of the learning theories embedded in the NCTM Professional Standards.

The researchers who developed the learning theories used by the NCTM provided useful insights about knowledge acquisition, but they offered no empirical evidence concerning the relationship between the NCTM-endorsed teaching approaches and standardized, or even non-standardized, tests (e.g. Case & Bereiter, 1984; Cobb & Steffe, 1983; Hiebert, 1986; Lampert, 1986; Lesh & Landau, 1983; Schoenfeld, 1987).

A handful of recent process-product studies do directly test these learning theories using traditional standardized tests. These studies suggest that students are not penalized if their teachers used an NCTM-type approach in their classrooms (Carpenter, Fennema, Peterson, Chiang, & Loef, 1989; Cobb et al., 1991; Knapp & Associates, 1995; Simon & Schifter, 1993). In fact, Knapp and Associates and Carpenter et al. found that the NCTM-taught students sometimes outperform their counterparts in more traditional classrooms.

How relevant are these studies to this study, and how much faith should be placed in their overall conclusions? On the one hand, their relevance is limited since none of them explicitly focused on secondary mathematics, let alone algebra, and two of them focused explicitly on the early elementary grades (e.g.Carpenter et al., 1989; Cobb et al., 1991). On the other hand, each of the studies did concentrate on themes at the heart of the NCTM teaching approach. Even if these studies were relevant, are their conclusions believable? I will show below that all four have their limitations.

For example, neither Simon and Schifter (1993), nor Cobb et al. (1991), established the validity of their study's most important variable, teaching practice. Their findings rely on the assumption that real differences exist between the teaching strategies employed in the treatment and control groups. Simon and Schifter (1993) establish what type of teaching occurred in the classrooms by using measures of student attitudes and beliefs toward mathematics. Common sense would suggest that this approach only provides a very rough approximation of teacher practice. The authors feed this skepticism by offering no information pertaining to the reliability or validity of their measures. Cobb et al. (1991) also use an indirect way of ascertaining what the teachers do in their classrooms. Though using teacher reports may seem to be an effective method at first blush, the survey questions they use only ask about pedagogic beliefs, not practice. Recent research suggests that when teachers discuss their teaching in the abstract it often fails to accurately capture what they do in practice (Burstein et al., 1995). Unfortunately, Cobb et al. offer no corroborating evidence to prove that beliefs and practice are correlated.

These studies are also limited by their lack of attention to confounding factors which might drive their findings. Cobb et al.'s treatment group volunteered to participate in a summer workshop and each volunteered to receive "extensive support" from the trainers throughout the school year. The control teachers consisted of the other teachers (those who did not volunteer). Were these teachers less interested in learning new teaching approaches? Were they more senior? Less senior? Better educated? Less well educated? Of a certain gender or race? Raudenbush, Rowan, and Cheong (1991) argue that a teacher's background and training affects the probability that she emphasizes an NCTM-type approach in her classroom. Thus ignoring the profiles of the teachers could be a major oversight if one wants to truly isolate the impact of teaching style.

Simon and Shifter used a design which controls for differences in teachers, yet their study has other limitations. These researchers used a historical design where the teachers received the intervention in the middle of the study. In the first year of the study teachers taught in their typical fashion. Then, over the summer, before the second year of the study, the teachers volunteered to receive training in the new teaching approach. After the second academic year the test results from the students from year one were compared to the test results from the students from year two. Were there differences between the year one and two students? Was there a "Hawthorne effect" (i.e. teacher expectations, not actions, changed student achievement)? These questions illustrate why historical research designs are often viewed as problematic (e.g. Cook & Campbell, 1979; Light, Singer, & Willett, 1990).

Another potential confounding factor that both Cobb et al. and Simon and Shifter ignore is that teachers use different teaching approaches depending on the overall makeup of their classrooms. Metz (1978) and Oakes (1985) found that classrooms comprised of more advanced students are provided with more opportunity to engage in critical thinking, which may in turn mean that their teachers use a more NCTM-like teaching approach. How were students sorted into the classrooms in both of these studies? Were the more advantaged students placed in classrooms with teachers who were more likely to use an NCTM approach? If this was the case, to what degree did this affect their conclusions?

Knapp et al. (1995) and Carpenter et al. (1989) successfully avoided some of the pitfalls which plague Cobb et al. and Simon and Shifter's research, but their work has significant limitations, too. These authors, as well as Cobb et al. and Simon and Shifter, use pre-test scores to help them measure student knowledge at one point in time, even though the limitations of this approach have been well documented (Willett, 1994). First, this approach does what it is supposed to do poorly. Most researchers use the pre-test to control for differences in the students' initial status. However, a pre-test can only imperfectly control for initial differences and this leads to biased parameter estimates (Rogosa, Brandt, & Zimowski, 1982). A second source of bias comes from the correlation between the pre-test score and any unobserved influences on student achievement, such as SES or parental involvement (Willett, 1994). Thus, by controlling for initial status with a pre-test, these authors use a measure of achievement which probably reflects more about student background characteristics than it does about student learning. To avoid this problem, researchers should study growth in achievement over time by gathering at least three measures of achievement (Rogosa & Saner, 1995).

Finally, all four of these studies are flawed in their analyses. Because students are nested within classrooms, it is likely that there are unobserved student characteristics within each classroom which are highly correlated with one another. It has been well documented that this situation, left unattended, results in biased estimates of the parameters' standard errors (Bryk & Raudenbush, 1992). Thus, these authors create some doubt about what they proclaim is, and is not, statistically significant. Each of these concerns is addressed in the design of this study, which is described next.

Methods

The Study Site

The target population consists of all Algebra 1 students and their teachers in one of the country's largest school systems. The Elm school district (a pseudonym for the actual district) rings a major city and has over 100,000 students. Seventy percent of the student population are black, 20 percent white, 4.6 percent Hispanic, and 4.4 percent Asian. Thirty-seven percent of the students receive free or reduced price school meals.

There are several compelling reasons why Elm made a superb location for this study. Many school districts offer lip service to implementing the Professional Standards, but few actually provide the necessary incentives and training . Elm is an exception. In 1989, just after the release of the first NCTM Standards document (National Council of Teachers of Mathematics, 1989), the Elm school committee recommended that the Standards be implemented in all mathematics courses. The incentive for teachers to adopt the Standards came from the state's testing program and the professional training offered by the College Board's EQUITY 2000 project.

The state in which Elm resides is one of the only states in the country using an "authentic" assessment program. The program has been in place for several years and in many respects is aligned with the Professional Standards. The College Board's EQUITY 2000 project may also have influenced whether teachers used the Professional Standards. In 1991, as part of this project, the College Board committed over $2 million to conduct five annual summer professional development institutes for all Elm algebra teachers. The Institutes focused on teaching the teachers to use the Professional Standards in their algebra classrooms (Choike, 1993) and the average teacher attended at least two institutes.

Population and the Attained Sample

The target population (shown in Table 1) for this study includes all black and white, eighth and ninth grade Algebra 1 students who remained in Algebra 1 and had the same teacher for the entire 1995-1996 school year.[2] The analytic sample includes the students who met the following three criteria: (1) their teachers completed a survey; (2) they completed at least two of the three algebra tests; and (3) they had complete data for all other background measures used in the analysis.

Table 1 shows that 67% of the students, 74% of the teachers, and 91% of the schools met the exclusion criteria and were therefore included in the analytic sample. Over 85% of the student, 98% of the teacher, and 100% of the school data loss is explained by missing survey data. This is not to say that the survey response rate was low; in fact, 80% of the teachers who taught algebra for the full year completed the survey. The teachers who completed a survey are not statistically different (in years of teaching experience, highest degree attained, gender, or ethnicity) from those who did not. The story is not so simple for the students. A two sample t-test comparing the means of high school and middle school students with missing data to the students without missing data across non-missing measures established that the attained high school sample is disproportionately black, has a slightly lower basic math skills score, and a slightly lower SES level. The attained middle school students have higher values on each of the academic ability measures used in the analysis. These differences indicate that generalizing this study's findings back to the target population should be done with caution.

Table 1 Here

Measures

In order to answer the primary research question (Is there a mismatch between traditional testing and new teaching approaches?), and control for important confounding factors, indicators on five factors were obtained. These factors are: (1) student mathematics achievement; (2) teaching style; (3) teacher background characteristics; (4) student background characteristics; and (5) the school environment.

Mathematics Achievement

To measure the impact of classroom practices on student learning, the Elm testing department administered three criterion referenced algebra tests. The first test was created explicitly for this study, consisted of 50 multiple choice items, and was administered in early September. The second and third tests are routinely given to all algebra students in early January and late May and each consisted of 25 multiple choice items. The estimated Cronbach's alpha reliability coefficient for each test was .79, .70 and .74, respectively. Students were given an unlimited amount of time to complete each exam and item response theory (IRT) was used to scale the test scores and thereby make them directly comparable across testing periods (Hambleton, Swaminathan, & Rogers, 1991).

In order to answer the primary question in this study learning was measured using traditional tests. These tests could not be construed as non-traditional because they are criterion referenced and rely exclusively on a multiple choice format (see sample questions presented in Table 2)

Table 2 Here

Teaching Style

All of the measures of the level of implementation of the NCTM teaching approach, and teacher background characteristics, and some of the measures of the school implementation environment come from the teacher survey developed for this study. All of the questions on this survey come from one of five prominent recent surveys (McLaughlin & Talbert, 1993; Pallas, 1988; Porter et al., 1993; Weiss, Matti, & Smith, 1994; Talbert, 1994) which investigated themes at the heart of this study.

Table 3 Here

All three of these model surveys only requested that teachers report how often they used various teaching methods (e.g. every day, twice a week, etc.) and not the duration over which the activities were used (e.g. five minutes at time, 15 minutes, etc.). By adding duration response options to my survey I could estimate the average percent of class time (= frequency x duration) teachers spend using each of the 17 teaching activities. Table 3 lists the 17 approaches used to gauge the percent of time each teacher used NCTM teaching activities in his or her classroom.

While evaluating the amount of time teachers use each of the 17 activities is informative, it can also be somewhat misleading because the activities are not mutually exclusive. Teachers could require students to be engaged in activities that include both NCTM and traditional components (e.g. students use calculators when they work on textbook problems). Consequently, I identified the preferred pedagogical style of the teachers by creating an indicator of the percent of time that teachers spend using the 13 NCTM approaches (relative to all 17 approaches). The composite measure of these 13 variables has a very high internal reliability rating (α=.85).

Reliability and validity of self-reported teaching data

As noted above, the state curriculum frameworks movement reflects a new direction for educational policy. This explains why it was only in the late 1980s that researchers and policy makers began to push for the routine collection of data which provided information on the schooling process (e.g. Murnane & Raizen, 1988; OERI, 1988; Porter, 1991; Shavelson, McDonnell, Oakes, Carey, & Picus, 1987). In response, some of the major national research organizations, such as the National Assessment of Educational Progress and the National Center for Education Statistics, began adding questions to their student, teacher, and school administrator surveys. However, Burstein et al. (1995) pointed out that the validity of the questions used on these surveys might be limited. They noted, "Little effort has been made to validate these measures by comparing the information they generate with that obtained through alternative measures and data collection procedures" (1995, p. 8). Why should there be skepticism about teacher reported data pertaining to the instructional process? Burstein et al. (1995) argue that all surveys are

> limited in their ability to portray a valid picture of the schooling process... [S]ome aspects of the curricular practice simply cannot be measured without actually going into the classroom and observing the interactions between teachers and students. These interactions include discourse practices that evidence the extent of students' participation and their role in the learning process, the specific uses of small-group work, the relative emphasis placed on different topics within a lesson, and the coherence of teachers' presentations (p. 7).

Surely, this perceived limitation of surveys, combined with policy makers' and researchers' historical emphasis on input-output studies, helps explain why, to date, much of what is known about the instructional process comes from in-depth studies done in only a handful of classrooms. The problem with these in-depth studies is that their generalizability to other classrooms is unknown (Burstein et al., 1995).

Consequently, Burstein et al. (1995) and Porter et al. (1993) began studying the validity of using surveys to examine classroom instruction. They found that surveys can provide accurate information about the teaching strategies most frequently used in classrooms. But because only a limited amount of this type of research exists, and because the quality of the teaching style measure is critical to this study, I conducted my own study of the NCTM measure's reliability and validity.

The survey questions pertaining to teaching style were re-administered four months later to a random subset of 20 teachers in order to see if the survey elicited consistent responses between two time points. For this subset, the correlation between the first and second NCTM composite was .69 ($p = .0013$) thus providing assurance that the survey is quite reliable.

For the validity study I selected a random sample of nine teachers and observed them each for three class periods. These teachers were selected by an independent researcher in order to ensure that my observations would not be colored by knowing the teachers' self-reported NCTM scores in advance of collecting and coding the data. I found that the observed and self-reported NCTM scores are strongly correlated ($r=.85$, $p=.004$).

Other Important Measures

Student, teacher, and school level phenomena were accounted for because not accounting for these factors would most likely lead to a biased estimate of the relationship between the NCTM teaching approach and learning.

The student demographics include race/ethnicity (white or black), gender, and socioeconomic status. SES is measured on a three point scale (2=free lunch, 1=reduced priced lunch, 0=no assistance).

In order to avoid having selection bias drive my findings (which could happen if the most talented students both receive the teachers who place the most emphasis on the NCTM teaching approach and have the fastest growth rates), three measures of prior academic ability are used as controls. The students' prior year's grade point average (GPA) from their core courses (English, mathematics, social studies, and science) gauges overall academic ability. Each student's most recent score on a state-administered criterion-referenced mathematics basic skills tests measures mathematics ability. (All students in Elm take this test beginning in sixth grade and must take it annually until they pass it. The time at which the students pass it is accounted for in the analysis.) The third ability measure is the fall criterion-referenced algebra test which is described above.

Several teacher variables were included in the analysis. The basic demographic variables consist of race/ethnicity, gender, highest educational degree attained (0=B.A. in any subject, 1=M.A. in any subject other than mathematics or mathematics education, 2=M.A. in mathematics or mathematics education), and the number of years teaching. Three variables assess the teacher's attitude toward embracing the NCTM reform: EQUITY 2000 Summer Mathematics Institute (SMI) attendance, a commitment to teaching composite (α=.72), and a gauge of the teacher's faith in their students' ability to pass algebra.

Seven school environment variables were used. Three are equally weighted composites created from teacher self-reports. These variables measure the degree of collegiality (α=.88), principal leadership (α=.94), and building-level problems (α=.78). The other variables come from central office files and include school size, the percent of students who are black, the percent of students receiving free or reduced lunch, and the percent of students absent more than 20 days out of the year.

Data Analysis

Hierarchical linear modeling (HLM) (Bryk & Raudenbush, 1992) can account for the hierarchical nature of these data (i.e. test scores nested within students and students nested within teacher's classrooms) and is therefore an appropriate statistical approach for estimating how student learning is influenced by student, classroom, and school processes. The models I fit linked student test scores, student background characteristics, and classroom and school predictors using three levels of statistical models. The "level 1" model expresses a student's observed knowledge as a function of time and takes the form:

$$Y_{tij} = \underset{\text{"initial status"}}{\pi_{0ij}} + \underset{\text{"growth rate"}}{\pi_{1ij}(\text{TIME})_{tij}} + e_{tij}, \quad \text{where} \quad e_{tij} \sim N(0, \sigma^2)$$

This growth trajectory model describes the observed status of student i at time t in teacher j's classroom as a function of time plus random error. Y_{tij} is the test score at time t for child i in teacher j's classroom. TIME is coded 0, 1, 2. Zero represents the fall test score, 1 represents the winter test score, and 2 represents the spring test score. By scaling TIME in this way, π_{0ij} represents the initial status for person i (i.e. their test score at the beginning of the school year) and π_{1ij} represents the growth rate for person i over the course of the 1995-1996 academic year.

π_{1ij} depicts predicted learning during a four month interval (the amount of time between each of the three exams).

The "level 2" models express the parameters from the level 1 model as a function of student characteristics in order to test whether the trajectories vary across individuals. The "level 3" models express parameters from the level 2 model as a function of NCTM (and other teacher and school characteristics) and allows me to determine whether the parameters in "level 2" models differ across classrooms. For a more complete description of this commonly used statistical technique see Bryk and Raudenbush (1992).

The NCTM Standards & Growth In Algebra Learning

Table 4 presents the means and standard deviations for all variables included in the study. The table illustrates just how different the middle school students are from the high school students. The means for the two groups are statistically different on each of the measures used in the analysis. The difference stems, in large part, from the fact that the "advanced" students in Elm take algebra in the eighth grade.

Figure 1 provides information about the preferences teachers give to <u>each</u> of the 13 NCTM tasks included in the composite. For example, both middle and high school teachers frequently use the most generic NCTM tasks (e.g. working in small groups and using calculators) while giving scant attention to some of more innovative approaches (e.g. working on individual projects, group investigations, and writing about mathematical problems). A striking difference between the middle and high school teachers is that seven of the 13 NCTM tasks listed in the figure are used less than eight percent of the time in the high schools, while only one of the NCTM tasks is used less than eight percent of the time in the middle schools. This indicates that, on average, when middle school teachers use the NCTM approach, they tend to use a variety of teaching techniques, while the high school teachers have a much more limited repertoire. This finding implies that middle and high school teachers who report using NCTM approaches the same percent of time are not actually teaching in the same manner.

Table 4 Here

The NCTM composite offers a more sharply focused look at how the middle and high school teachers differ in terms of their enthusiasm for the NCTM teaching approach as a whole. High school teachers spend, on average, 63% of their time using NCTM tasks, while middle school teachers spend 80%. These differences between the middle and high schools teachers' pedagogical approaches lends support to the "sorting" hypothesis advanced by Metz, Oakes, and Raudenbush et al. (1978; 1985; 1993). The middle school algebra students in Elm are certainly the more academically advanced and they definitely receive much more exposure to the NCTM approach than their less advanced counterparts in the high schools.

Figure 1 Here

Variation Partitioned

A fully unconditional three level growth model (e.g. the only predictor of test scores was "time" at level-1) indicates that while a large amount of variance in growth rates exists within students a substantial amount of potentially explainable between-student, and between-teacher variation exists in both the middle and high schools (see Table 5). <u>This implies that student growth trajectories differ substantially among teachers and could potentially be related to teaching style</u>.

Table 5 Here

High School Growth Models

The first model presented in Table 6 illustrates that a significant, and large, sorting effect exists even within the high schools (not just between the middle and the high schools). The model indicates that the predicted fall test score for students whose teacher's NCTM score is two standard deviations (sd=.21) below the mean is 463, while students in classrooms where the NCTM score is two standard deviations above the mean would have a fall test score of 475, an almost one-half a standard deviation difference.

Model 2 shows that this 12 point gap in the predicted fall test scores can be accounted for by adding other significant student, teacher, and school variables. Black students are predicted to have a 8.4 lower initial test score, while higher SES levels and higher math skills are associated with higher math scores. Students who pass the state math skills test in sixth grade scored 9 points higher on the fall test than other students. The number of problems in the school and the percent of black students are negatively related to fall test scores, while school absentee rates are positively associated with them. The absenteeism finding is surprising given that the higher the absentee rate the worse the academic climate. This unusual pattern appears to represent a skimming effect whereby the best students in these disadvantaged schools remained in the final sample.

Model 3 adds NCTM to the growth rate portion of the model, and Model 4 adds additional significant predictors. These models reveal that no significant relationship exists between NCTM and a student's rate of growth. Since E.D. Hirsch and the early process-product studies claim that the NCTM teaching approach should negatively interact with a student's ability and SES level, I tested for this interaction. None exists. Thus, Model 4 illustrates that the NCTM approach does not have an effect on growth rates at the high school level, though some other variables are related to growth. Students in larger schools are predicted to gain at a faster rate than students in smaller schools, high-GPA students grow faster than low-GPA students, and black students grow faster than whites (a very important finding which is beyond the scope of this article).

Table 6 Here

Middle School Growth Models

A series of models displaying the relationship between NCTM and algebra learning in the middle schools are presented in Table 7. As in the high school example, the first model illustrates that a significant and large sorting effect exists within the middle schools. The predicted fall test score for a student who has a teacher whose NCTM score is two standard deviations (sd=.26) below the mean would be 482, while a student in a classroom where the NCTM score is two standard deviations above the mean would have a fall test score of 496, a difference of over two-thirds of a standard deviation.

Model 2 illustrates that black students are predicted to have a lower fall test score than white students, while high SES and high achieving students (i.e. high-GPA, high math skills, and students who passed the math skills test in sixth grade) are predicted to have higher fall test scores than their respective low SES and low achieving counterparts. The number of building-related problems is negatively related to fall test scores, while a teacher's attendance at the SMIs, the size of the student body, and the level of absenteeism are positively related to the pre-test. Model 3 illustrates that the main effect of NCTM is not significant (p-value = .20). But the addition of other predictors in Model 4 changes things. NCTM is positive and marginally significant (p-value

= .09), as is an interaction between NCTM and the student's prior year's GPA (p-value = .09). This model indicates that all middle school students are predicted to grow more rapidly in classrooms where the teacher emphasizes an NCTM approach, but the higher the student's prior year GPA, the more they will respond to the approach.

Model 4 also indicates that other student, teacher, and school variables are significantly related to growth. Prior student ability is positively related to growth, as is the level of faith that teachers have in their students and the level of commitment teachers have in their jobs. Both faith and commitment interact with student GPA, but they do so in different ways. As student GPA increases, the magnitude of the effect of faith decreases, but the magnitude of the effect of commitment increases. School size is positively related to student growth rates, but for some reason principal leadership is negatively related to growth.

The finding of a positive effect of NCTM and its interaction with GPA is sensitive to the presence of a high leverage point (one teacher's NCTM score is three standard deviations below the mean). Refitting the model with the leverage point set aside (Model 5) results in the NCTM*GPA term becoming highly significant (p-value < .002) and the coefficient's magnitude increasing by one-third. In addition, the p-value for the main effect of NCTM raises to .37 and the magnitude of the coefficient is almost cut in half. The removal of the high leverage point suggests that students with low GPA's (at least one standard deviation below the mean) receive close to no benefit from the NCTM teaching approach. This suggests that the reason the high school students do not respond to the NCTM teaching approach may be due to the fact that they are a relatively low skilled group of students.

Table 7 Here

Figure 2 illustrates the magnitude of the effect of NCTM in the middle schools. These results are based on Model 4. The figure presents the predicted growth trajectories for two types of middle school black students (a high and a low-GPA student, defined as one standard deviation above and below the mean GPA) under two different scenarios (placed in a high or low NCTM classroom, defined as 20 percentage points above and below the mean NCTM level). (All other variables were set to their mean values.) The most important message is that the algebra knowledge of a low-GPA, as well as high-GPA, student is predicted to grow faster in the middle schools if they are in high NCTM classrooms. A typical low-GPA student is predicted to be nine points (one-third of a standard deviation) better off being in a high versus a low NCTM classroom. High-GPA students clearly reap more benefits from the NCTM approach: they are predicted to be 17 points (over one-half a standard deviation) better-off by being in a high versus a low NCTM classroom.

The figure also illustrates important information about the performance gap between high and low-GPA students. The low-GPA students score, on average, three points lower than their high-GPA counterparts in the fall. If a low-GPA student were to be in a high NCTM classroom and the high-GPA student were to be in a low NCTM classroom, the low-GPA student might actually score three points higher on the spring test. But this scenario is unlikely given that the high-GPA students are sorted into the high NCTM classrooms. In fact, the more likely scenario is that the high-GPA students receive high NCTM teachers and the low-GPA students receive the low NCTM teachers. The predicted result of this sorting is that the three point gap between these students in the fall balloons into a gap of 23 points (over two-thirds of a standard deviation) by the spring.

Figure 2

Variation Explained

Table 8 illustrates that almost 45% of between-teacher growth variation was explained in the middle schools and that only about four percentage points of this reduction was explained by the NCTM composite. In the high schools, only 20% of between-teacher growth variation was explained and none was explained by the addition of the NCTM variable.

Table 8 Here

Discussion

My results indicate that a dangerous mismatch does not exist in the Elm schools. Student test scores are not hurt, and for some they are helped, by using the Professional Standards with old tests. This suggests that critics of the NCTM teaching approach (e.g. E.D. Hirsch) may not be justified in their claims that the Professional Standards lead to declines in standardized test scores. Stakeholders who believe that a back to basics teaching style must be used because that's what the tests measure, are provided with no evidence that such a teaching approach will benefit high school algebra students in Elm, and this study suggests that a back to basics approach will even harm Elm middle school students.

Limitations of this Study

In carrying out further research in this area, and in interpreting my findings, some limitations of this study should be considered. One limitation is that the study's findings only generalize to the white and black eighth and ninth grade algebra students who remained with one teacher for the whole year in Elm. A second limitation is that this study, and any study which does not randomly assign its subjects to the treatment, is that it only imperfectly controls for student, teacher, and school differences.

A third limitation involves the teaching style measure. Even though I found that this measure presents a valid and reliable measure of the amount of time teachers spend using the NCTM approach, the measure is still limited by its inability to gauge the quality of the interactions between the teacher and her students. The lack of finding in the high school may be due to the fact that the NCTM measure does not adequately account for subtle differences in the quality of the implementation of the NCTM approach. A more discriminating measure that could pick up these differences could potentially unravel the high school-middle school differential. The problem for researchers is to identify a way to create a measure, or series of measures, that can be gathered relatively inexpensively from a number of classrooms. In-depth studies of the teaching process are informative, but, as noted above, generalizing from them is difficult because they can only be conducted in a very small number of classrooms.

This study's NCTM measure, however, does clearly measure something relevant. The patterns of NCTM exposure are consistent with what has been found in prior research (i.e. the most talented students get more exposure to an NCTM-like teaching style). If the measure were meaningless, I would not have identified statistically significant relationships between it and the student's initial test scores and growth rates.

Research Implications

These findings contradict the findings from the early process-product research and corroborate the findings of later process-product research. The early studies suggested that all students, but especially low-income low-achieving students, would be harmed by using some of the teaching approaches recommended by the Professional Standards (Brophy & Good, 1986). Hirsch recently argued that this research has "consistently shown" that an NCTM type approach is "the least effective approach" to teaching and that no other "mainstream research" exists to refute these findings (Hirsch, 1996, p. 216). Using the early studies to pass judgment on the Professional Standards is problematic because the teaching measures used in those studies predate the development of the learning theories used in developing the Professional Standards and thus are not tightly coupled with the Professional Standards.

Each of the few post-Standards process-product studies (i.e. this study and Carpenter et al., 1989; Cobb et al., 1991; Knapp & Associates, 1995; Simon & Schifter, 1993) has a more precise definition of the teaching approach emphasized by NCTM, and each of these studies finds that students taught in a manner consistent with the NCTM approach do at least as well, if not better, on a traditional multiple choice test as their peers in more traditionally taught classrooms. This should give pause to those who cite the pre-Standards process-product studies as evidence that the Professional Standards are harmful.

But critics of the Professional Standards may have justifiably turned to the early process-product studies for answers because the post-Standards process-product studies are relatively limited in both their number and quality. For this reason, future studies looking at the relationship between the NCTM Professional Standards and standardized tests are badly needed. New studies need to be designed to both avoid the pitfalls earlier researchers fell into (see above), and to answer questions raised by this study.

One of the most pressing questions my study provokes is the following: why do the high ability students, on average, benefit most from the NCTM approach? This question is particularly relevant given NCTM's (1991) claim that the approach is appropriate for students of all ability levels, and Hirsch's (1996) counterclaim that all students, especially low-achieving students, will be harmed by an NCTM approach.

Is this finding specific to algebra instruction? Unfortunately, both the early and late process-product studies have focused on the primary grades and given scant attention to the secondary grades, let alone algebra. More studies focused exclusively on algebra are needed.

Could it really be that only the more "talented" students respond to the NCTM teaching approach? The fact that an effect only exists in the middle schools, and that it is more pronounced for the more academically talented students in the middle schools, suggests that this is possible. As noted above, some of the recent process-product studies have virtually ignored the ability level of the students in their analyses (Cobb et al., 1991; Simon & Schifter, 1993). This is a lost opportunity for two reasons. First, historically, teaching approaches have been found to be more or less effective with students of different abilities (Brophy & Good, 1986). Second, students in this study do not appear to be assigned randomly to teaching approaches. The most talented students in both the Elm high and middle schools received teachers who were most likely to use the NCTM approach. Though prior research has documented this type of phenomena (e.g. Metz, 1978; Oakes, 1985; Raudenbush et al., 1993), almost none of the post-Standards process-product studies investigated or adequately controlled for this sorting process. Future research should be sure to do so.

But is the difference in the impact of NCTM on growth rates really due to differences in student ability level? Perhaps, but other plausible explanations exist at the school, teacher, and student level and these should be explored. At the school level, a plausible explanation could be the following: Because middle schools in Elm have student populations and levels of absenteeism that are half that of the high schools, dramatically different teaching and learning environments could exist in these institutions. Because the NCTM approach demands a kind of intense teacher-student interaction, it might be that the approach either works better or is easier to implement in the more intimate middle school setting.

A teacher-level explanation for the differential in the effectiveness of the Professional Standards might focus on the qualitative differences in the implementation of the NCTM approach between the middle and high schools. Perhaps the high school teachers are less effective because they embrace the more generic NCTM tasks (e.g. working in small groups and using calculators) rather than the more precise and innovative ones (e.g. working on individual problems, group investigations, and writing about mathematical problems). Possible reasons for this include state level policies, institutional operating procedures, and student factors. The state's eighth grade, not ninth grade, testing program uses an "authentic" assessment to hold schools accountable. This could explain why middle school teachers are more likely to use the NCTM teaching approach in their classrooms. An institutional operating procedures argument would be that high school teachers are given less time to master how to use the NCTM tasks. A student argument would be that because high school algebra students are less academically talented, their teachers assume that only certain types of NCTM approaches could be used with them effectively.

A student-level explanation for the difference in effectiveness might focus on developmental issues. Does the stage of psychological and academic development of the middle and high school students affect their receptivity to the NCTM approach? Though only one year in age separates most eighth and ninth grade students, it is conceivable that, developmentally, the eighth graders are more responsive to working in groups, discussing ideas, and engaging in hands-on learning, while this teaching approach is distracting or uninteresting to ninth graders.

While these questions remain to be answered by further research there are some important relevant findings which this study produced.

Policy Implications

The United States has launched an unprecedented educational reform movement in which the NCTM Professional Standards play a critical role. The movement is visible in the spate of new state curriculum frameworks, the federal government's Goals 2000 initiative, and the professional curriculum associations' newly created standards documents. This reform effort stands in sharp contrast to earlier efforts because it intends to change curriculum and teaching practices. In the past, policy makers have tinkered at the edges of the classroom, trying to foster improvement by reallocating resources and by setting outcome goals. The perceived failure of these policies has led to the country's current interest in educational standards aimed at carefully shaping what happens in the classroom.

This study provides policy makers with evidence that the traditional multiple choice tests that characterize the current testing environment may not directly undermine the standards movement in Elm, but they might not help move it forward, either. Given that tests influence teaching behavior (Koretz et al., 1991; Resnick & Resnick, 1991), as long as standardized tests are used to gauge student growth in algebra learning, the high school teachers--teachers of the lowest achieving students in Elm--will not receive any incentive from the test results to change their teaching style. Middle school algebra teachers, on the other hand, do receive some incentive to adopt the NCTM teaching approach.

If policy makers want both to continue to use traditional standardized tests to measure student performance and to see the new mathematics standards implemented, they should consider that high school teachers may need more coaxing to embrace the NCTM approach than their middle school counterparts.

Those who see algebra as a gatekeeper to future opportunities may interpret these findings in two distinctly different ways. Because the disadvantaged students that algebra advocates want to help (e.g. the high school students who are the predominantly black and poor students in this study) do not receive any boost in their test scores from the NCTM approach, this suggests that experimentation with a yet to be identified, more effective teaching approach might be warranted. On the other hand, if, as the NCTM advocates argue, students are learning some important additional skills which are not measured by traditional standardized tests (e.g. the ability to solve problems, to make conjectures, and to communicate both verbally and in writing), then sticking with the NCTM approach would prove beneficial in the long run, since these skills have been identified as economically beneficial (Murnane & Levy, 1996). The trick will be finding a way to encourage the high school teachers to use the NCTM approach even if their students do not perform any better on the old tests.

The significant sorting effect between and within the middle and high schools should give pause to all who are concerned with educational equity. This study corroborates prior research that teachers use different teaching approaches depending on the skill level of the children they instruct (e.g. Metz, 1978; Oakes, 1985; Raudenbush et al., 1993). In particular, the academically advanced students are more likely to receive more exposure to the NCTM teaching approach.

The sorting of students into various instructional "tracks" using varying levels of the NCTM approach causes performance inequities within the middle schools because all middle school students benefit from the approach. In the high school, because neither low- nor high-achieving students appear to benefit from the Professional Standards (at least according to their performance on traditional multiple choice tests), the inequity is more subtle. Because the reform initiative was conceptualized to benefit everyone, its uneven implementation undermines this original vision.

For the standards movement to truly improve the quality of education in the United States, researchers, policy makers, and practitioners must pay close attention to the implementation environment. In this study I focused on the relationship between new teaching standards and old tests because the advocates of the new standards claim that the continued use of old tests will create a disincentive for teachers to embrace the new standards. Prior research exploring this issue has serious design flaws, offers inconsistent findings, and has focused almost exclusively on the primary grades. This study corrected for these limitations and found that that the fears of a mismatch between the new standards and the old tests are not entirely warranted in algebra classrooms. Middle school algebra students learn algebra at a faster rate when their teachers use the NCTM Professional Standards, but the learning rates of the disproportionately low-achieving, black, and poor high school students are unaffected. If, as the reformers claim, there are benefits that students receive from the NCTM approach which are not measured by the old tests, then the current implementation environment may be fostering inequities by failing to offer the teachers of the low-achieving students any incentive to adopt the approach.

TABLE 1: The target population and the analytic sample (the percent of target population included in the analytic sample is shown in parentheses).

	Students	Teachers	Schools
Target population			
Middle School	2070	49	25
High School	1444	78	20
Total	3514	127	45
Analytic sample			
Middle School	1400 (68%)	37 (76%)	22 (88%)
High School	969 (67%)	57 (73%)	19 (95%)
Total	2369 (67%)	94 (74%)	41 (91%)

TABLE 2: Sample questions from the criterion referenced algebra tests.

1) Factor: $a^2 - 7a + 6 =$

A. (a-1) (a + 6)
B. (a + 1) (a + 6)
C. (a - 1) (a - 6)
D. (a + 1) (a - 6)

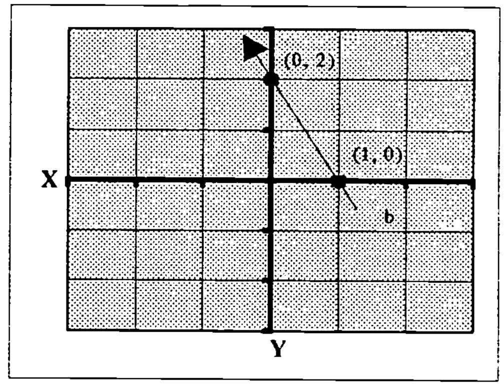
2) If the ratio of x to y is 3 to 4 and the ratio of y to z is 2 to 1, then the ratio of x to z is:

A. 3 to 2
B. 3 to 1
C. 2 to 4
D. 2 to 3

3) An automobile is moving at "r" miles per hour, and an airplane is moving four times as fast. How many hours will the plane require for a 600 mile flight?

A. 1600/r
B. 4r/600
C. 600 - 4r
D. 600/4r

4) Consider line b below. Two points on the line are indicated. Determine the equation of the line.



A. y = -2x + 2
B. y = 3x + 2
C. y = 2x + 2
D. y = -x + 2

BEST COPY AVAILABLE

36

TABLE 3: The 17 teacher practice variables used to assess the percent of time teachers spend using the NCTM teaching approaches with their students.

Traditional Approaches	
	Students...
	Listen to lectures
	Work from a textbook
	Take computational tests
	Practice computational skills
NCTM Approaches	
	Students...
	Use calculators
	Work in small groups
	Use manipulative materials
	Making conjectures
	Engage in teacher led discussion
	Engage in student led discussion
	Work on group investigations
	Write about problems
	Solve problems with more one correct answer
	Work on individual projects
	Orally explain problems
	Use computers
	Discuss different ways to solve problems

TABLE 4: Means and standard deviations for all student (n= 2369), teacher (n=94), and school (n=40) variables in the analytic sample and test statistics on the differences in means between the middle and high school variables.

Variables	Full Sample Mean	Full Sample (SD)	Middle School Mean	Middle School (SD)	High School Mean	High School (SD)	Test Statistic (df)	P-value
Student	n=2369		n=1400		n=969			
Male	45%		43%		49%		χ2 statistic = 9.83 (1)	0.002
Female	55%		57%		51%			
White	25%		31%		16%		χ2 statistic = 70.04 (1)	0.001
Black	75%		69%		84%			
SES	1.50	(0.81)	1.60	(0.73)	1.30	(0.87)	t-statistic = -8.82 (1810)	<0.001
GPA	2.66	(0.85)	2.90	(0.75)	2.20	(0.8)	t-statistic = -24.51 (2367)	<0.001
Math Basic Skills	354	(21)	360	(21)	345	(16)	t-statistic = -19.48 (2343.1)	<0.001
Last Took Math Basic Skills	6.6	.77	6.2	(.39)	7.2	(.79)	t-statistic = 36.54 (1303)	<0.001
Fall Test	482	(25)	489	(21)	471	(27)	t-statistic = -16.07 (1539.1)	<0.001
Winter Test	513	(30)	523	(29)	500	(27)	t-statistic = -19.10 (2139)	<0.001
Spring Test	546	(34)	555	(33)	532	(31)	t-statistic = -17.43 (1995.1)	<0.001
Teacher	n=94		n=37		n=57			
Percent of time using NCTM approaches	70%	(20%)	80%	(13%)	63%	(20%)	t-statistic = -4.10 (92)	0.003
Black	37%		30%		42%		χ2 statistic = 1.49 (1)	0.222
White	63%		70%		58%			
Female	53%		59%		49%		χ2 statistic = 0.96 (1)	0.326
Male	47%		41%		51%			
Years teaching	12.00	(10)	14.46	(11)	11.05	(9.97)	t-statistic = -1.58 (72)	0.118
Education	0.72	(0 66)	0.65	(0.63)	0.77	(0.68)	t-statistic = 0.88 (92)	0.380
Number of Summer Math Institutes attended	2.33	(1.59)	2.54	(1.64)	2.19	(1.55)	t-statistic = -1.03 (92)	0.302
Problems in school	0.00	(1)	-0.4	(0.94)	0.30	(0.95)	t-statistic = 3.30 (92)	0.001
Principal leadership	0.00	(1)	0.10	(1)	-0.07	(0.99)	t-statistic = -0.79 (92)	0.430
Collegiality among staff	0.00	(1)	-0.07	(1)	0.05	(0.99)	t-statistic = -0.57 (92)	0.570
Commitment to teaching profession	0.00	(1)	0.07	(0.9)	-0.04	(1)	t-statistic = -0.52 (92)	0.600
Percent of students who will pass algebra	79%	(0.19)	85%	(0.15)	75%	(0.2)	t-statistic = -2.57 (92)	0.012
School	n=40		n=22		n=18			
Percent white	17%	(15%)	18%	(14%)	16%	(16%)	t-statistic =-0.46 (33.4)	0.647
Percent black	76%	(17%)	75%	(15%)	77%	(19%)	t-statistic =-0.38 (38)	0.703
Percent receiving free or reduced lunch	38%	(16%)	40%	(16%)	35%	(16%)	t-statistic =-1.12 (38)	0.270
Percent absent more than 20 days	25%	(14%)	16%	(7%)	36%	(13%)	t-statistic =5.80 (24.8)	<0.001
Number of students	1103	(607)	737	(152)	1550	(656)	t-statistic =5.15 (18.5)	<0.001

TABLE 5: The amount of within-student variance, and the amount of between-student and between-teacher variance at initial status and growth.

	Middle Schools		High Schools	
	Variance	% of total variance	Variance	% of total variance
Within Student Variance[a]	373.4	100	552.5	100
Initial Status				
Between Students	53.5	37.8	47.8	24.0
Between Teachers	88.0	62.2	151.5	76.0
Total	141.5	100.0	199.3	100.0
Growth				
Between Students	40.8	37.2	12.4	13.4
Between Teachers	69.0	62.8	79.9	86.6
Total	109.8	100.0	92.3	100.0

[a]This represents the variance in observed scores about an individual's growth trajectory.

40

TABLE 6: High school growth models (N= 969 students, and 57 teachers).

	Model 1		Model 2		Model 3		Model 4	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE
Initial Status								
Intercept	469.00***	1.93	471.80***	2.45	471.77***	2.47	475.05***	2.22
NCTM	**14.52***	**5.85**	**4.84**	5.03	**6.77**	**8.07**	**2.66**	**7.68**
Math skills			0.24***	0.04	0.24***	0.04	0.23***	0.04
Math skills 6th grade			8.83**	1.28	8.83***	1.27	8.82***	1.28
Black			-5.02*	2.00	-5.02*	2.01	-8.36***	1.96
Problems			-3.19*	1.33	-3.18*	1.33	-3.28*	1.33
Absenteeism			0.29*	0.13	0.29*	0.13	0.35*	0.13
Size[a]			0.01	0.00	0.01	0.00	-0.69*	0.00
Percent black			-16.33*	8.13	-16.31*	8.12	-17.35*	7.96
Growth Rate								
Slope	30.14***	1.48	30.01***	1.46	30.06***	1.49	25.82***	1.43
NCTM					**-2.76**	**7.81**	**1.89**	**7.13**
Black							4.55**	1.29
GPA							1.44**	0.54
SES							1.02*	0.45
Size[a]							0.9**	0.00

[a] In this table, size has been rescaled to be in the 100's.

~ p<.10 ** p.<.01
* p<.05 *** p.<.001

BEST COPY AVAILABLE

TABLE 7: Middle school growth models (N= 1400 students, and 37 teachers).

	Model 1		Model 2		Model 3		Model 4		Model 5[b]	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE
Initial status										
Intercept	488.90***	1.59	487.58***	1.62	487.55***	1.62	487.64***	1.58	487.66***	1.58
NCTM	**27.34***	**10.23**	**6.92**	**6.54**	**3.08**	**7.72**	**6.31**	**7.33**	**9.52**	**9.15**
GPA			3.67***	0.75	3.67***	0.75	1.93*	0.85	2.06*	0.86
Math skills			0.28***	0.04	0.28***	0.04	0.21***	0.04	0.20***	0.04
Math skills*6th grade			3.89***	1.38	3.88**	1.38	3.89**	1.37	4.03**	1.36
Black			-3.24**	1.22	-3.24**	1.22	-3.17*	1.23	-3.24*	1.25
Problems			-2.06*	0.90	-2.08*	0.91	-1.97~	0.99	-1.97~	0.99
SMI attendance			2.46**	0.69	2.44**	0.70	2.28**	0.72	2.33**	0.72
Absenteeism			0.57**	0.19	0.57**	0.19	0.56*	0.21	0.55*	0.22
Size[a]			1.78*	0.01	1.76*	0.01	1.70*	0.01	1.78*	0.01
Growth Rates										
Slope	32.64***	1.46	32.49***	1.46	32.52*	1.45	31.90***	1.13	32.18***	1.13
NCTM					**13.53**	**10.44**	**15.17~**	**8.78**	**8.40**	**9.24**
NCTM*GPA							**6.20~**	**3.64**	**9.70****	**3.00**
GPA							2.34***	0.46	2.23***	0.42
Math skills							0.10***	0.02	0.10***	0.02
Faith							9.04	5.91	9.07	5.60
Faith*GPA							-5.05**	1.79	-5.10**	1.85
Commitment							2.93**	0.94	2.67**	0.89
Commitment*GPA							0.77*	0.32	1.01**	0.29
Principal							-3.98**	1.01	-4.03***	0.91
Size[a]							1.65**	0.01	1.53*	0.01

[a] In this table size has been rescaled to be in the 100's.

[b] Model 5 represents what happens to the findings reported in Model 4 when one teacher whose NCTM score was three standard deviations below the mean is set aside from the analysis. Note the changes in NCTM and NCTM*GPA between the two models.

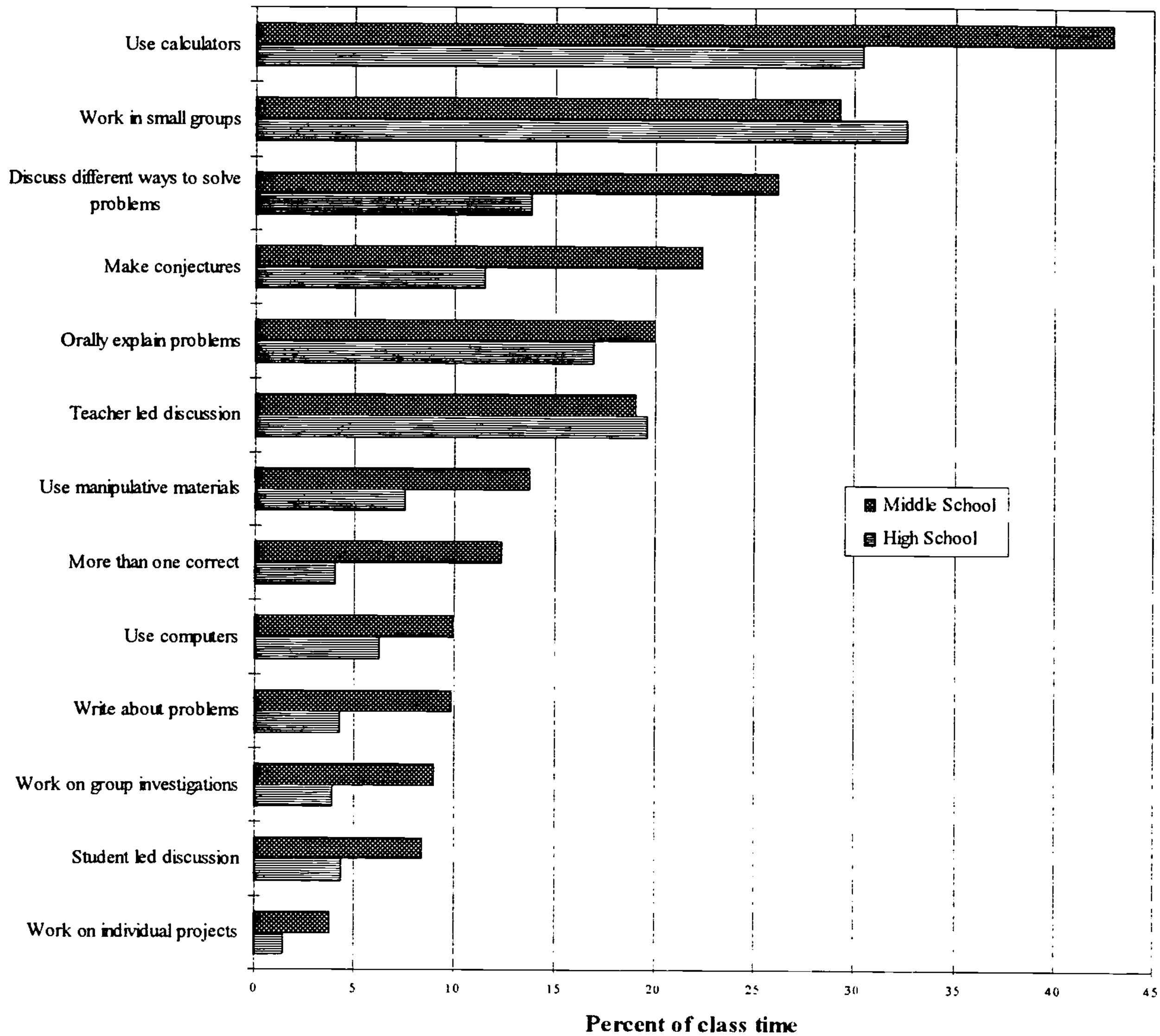
~ p<.10 ** p.<.01
* p<.05 *** p.<.001

BEST COPY AVAILABLE

TABLE 8: Variances and variance accounted for in student growth rates between teachers.

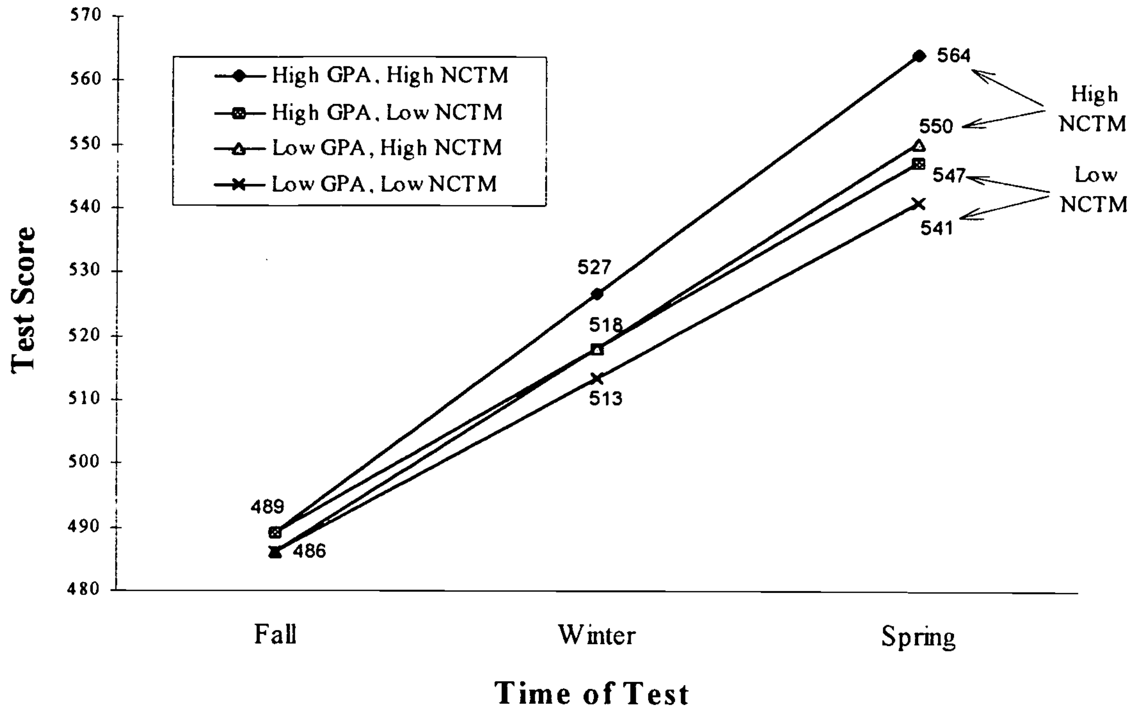
	Middle School		High School	
Model	Variance	Percentage reduction in variance	Variance	Percentage reduction in variance
1) No predictors	66.8		78.9	
2) Student predictors	60.5	9.4	74.6	5.4
3) Student & Teacher predictors	52.6	21.3	74.6	5.4
4) Student & Teacher & School predictors	39.7	40.6	62.8	20.4
5) Student & Teacher & School predictors & NCTM	37.0	44.6	62.8	20.4

FIGURE 1: The average percent of time middle and high school teachers use the NCTM teaching approaches.



BEST COPY AVAILABLE

FIGURE 2: Projected growth in middle school algebra learning.



References

Black, P. J., & Atkin, J. M. (Eds.). (1996). *Changing the subject: Innovations in science, mathematics, and technology education.* New York City: Routledge.

Blank, R. K., Hemphill, C., Sardina, S. L., Langesen, D., & Brathwaite, B. (1995). *State education policies on k-12 curriculum, student assessment, and teacher certification: 1995* . Washington, DC: Council of Chief State School Officers.

Blank, R. K., & Pechman, E. M. (1995). *State curriculum frameworks in mathematics and science: How are they changing across the states?* . Washington, DC: Council of Chief State School Officers.

Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328-375). New York: Macmillan.

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

Burstein, L., McDonnell, L. M., Van Winkle, J., Ormseth, T., Mirocha, J., & Guitton, G. (1995). *Validating national curriculum indicators* . Santa Monica: RAND Corporation.

Campbell, F. P. (1995,). *Project IMPACT: Mathematics achievement in predominately minority elementary classrooms attempting reform.* Paper presented at the Annual meeting of the American Educational Research Association, San Francisco, CA.

Carpenter, T. P., Fennema, E., Peterson, P., Chiang, C. P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26*(4), 499-531.

Case, R., & Bereiter, C. (1984). From behaviorism to cognitive development. *Instructional Science, 13*, 141-58.

Choike, J. R. (1993). *Math as a lever for reform: Excerpts from the Equity 2000 National Committee Report* . New York: The College Board.

Cobb, P., Wood, T., Yackel, E., Wheatley, G., Trigatti, B., & Perlwitz, M. (1991). Assessment of a problem-centered second-grade mathematics project. *Journal for Research in Mathematics Education, 22*(1), 3-29.

Cobb, R., & Steffe, L. P. (1983). The constructivist researcher as teacher and model builder. *Journal for Research in Mathematics Education, 14*, 83-94.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings.* Boston: Houghton Mifflin.

Hambleton, R. K., Swaminathan, H., & Rogers, J. H. (1991). *Fundamentals of item response theory.* (Vol. 2). Newbury Park: Sage.

Hiebert, J. (Ed.). (1986). *Conceptual and procedural knowledge: The case of mathematics.* Hillside: Lawrence Erlbaum Associates.

Hiebert, J., & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. *American Educational Research Journal, 30*, 393-425.

Hirsch, E. D. (1996). *The schools we need: Why we don't have them.* New York: Doubleday.

Kirst, M., & Masseo, C. (1996, December 1996). The rise, fall, and rise of state assessment in California 1993-96. *Phi Delta Kappan, 78*, 319-323.

Knapp, M. S., & Associates. (1995). *Teaching for Meaning in High-Poverty Classrooms*. New York: Teachers College Press.

Koretz, D., Linn, R. L., Dunbar, S. B., & Shepard, L. A. (1991,). *The effects of high-stakes testing on achievement: Preliminary findings about generalization across tests.* Paper presented at the American Educational Research Association and the National Council on Measurement in Education, Chicago.

Lampert, M. (1986). Knowing, doing and teaching mathematics. *Cognition and Instruction, 3*, 305-42.

Lawton, M. (1997, January 6, 1997). States turn to a mix of tests in hopes of a clearer picture., pp. 6.

Lesh, R., & Landau, M. (Eds.). (1983). *Acquisition of mathematics concepts and processes*. New York: Academic Press.

Light, R. J., Singer, J. D., & Willett, J. B. (1990). *By design: Planning research on higher education*. Cambridge: Harvard University Press.

Marshall, S., Fuhrman, S., & O'Day, J. (1994). National curriculum standards: Are they desirable and feasible? In R. Elmore & S. Fuhrman (Eds.), *The governance of curriculum: 1994 yearbook of the Association for Supervision and Curriculum Development* . Alexandria: ASCD.

McDonnell, L. M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis, 17*(3), 305-322.

Metz, M. H. (1978). *Classrooms and corridors: The crisis of authority in desegregated secondary schools*. Berkeley: University of California Press.

Murnane, R. J., & Levy, R. (1996). *Teaching the new basic skills: Principles for educating children to thrive in a changing economy.* New York City: The Free Press.

Murnane, R. J., & Raizen, S. A. (1988). *Improving indicators of the quality of science and mathematics education in grades K-12* . Washington, D.C: National Academy Press.

National Center for Education Statistics. (1992). *A profile of American eighth-grade mathematics and science instruction*. Washington, D.C.: U.S. Department of Education.

National Council of Teacher of Mathematics. (1995). *Assessment standards for school mathematics*. Reston: National Council of Teacher of Mathematics.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics* . Reston: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics* . Reston: National Council of Teachers of Mathematics.

Oakes, J. (1985). *Keeping track: How schools structure inequality.* New Haven: Yale University Press.

OERI, S. A. S. G. (1988). *Creating responsible and responsive accountability systems* . Washington, D.C.: Department of Education.

Pelavin, S. H., & Kane, M. K. (1990). *Changing the odds: Factors increasing access to college* . New York: The College Entrance Examination Board.

Porter, A. C. (1991). Creating a system of school process indicators. *Educational Evaluation and Policy Analysis, 13*(1), 13-29.

Porter, A. C., Kirst, M. W., Osthoff, E. J., Smithson, J. L., & Schneider, S. A. (1993). *Reform up close: An analysis of high school mathematics and science classrooms* . Madison: Wisconsin Center for Education Research.

Raudenbush, S. W., Rowan, B., & Cheong, Y. F. (1993). Higher order instructional goals in secondary schools: Class, teacher, and school influences. *American Educational Research Journal, 30*, 523-553.

Resnick, L. B., & Resnick, D. P. (1991). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative view of aptitude, achievement, and instruction* (pp. 37-75). Boston: Kluwer Academic Publishers.

Rogosa, D., & Saner, H. (1995). Longitudinal data analysis examples with random coefficient models. *Journal of Educational and Behavioral Statistics, 20*(2), 149-170.

Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 90*, 726-748.

Rowan, B., Raudenbush, S. W., & Kang, S. J. (1991). Organizational design in high schools: A multilevel analysis. *American Journal of Education, 99*, 238-266.

Schoenfeld, A. H. (1987). *Cognitive science and mathematics education*. Hillside: Lawrence Erlbaum Associates.

Shavelson, R. J., McDonnell, L. M., Oakes, J., Carey, N., & Picus, L. (1987). *Indicator systems for monitoring mathematics and science education* . Santa Monica, CA: RAND.

Shulman, L. S. (1986). Paradigms and research programs in the study of teaching: A contemporary perspective. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 3-36). New York: Macmillan.

Simon, M. A., & Schifter, D. (1993). Toward a constructivist perspective: The impact of a mathematics teacher inservice program on students. *Educational Studies in Mathematics, 25*, 331-340.

Smith, J. B. (1996). Does an extra year make any difference? The impact of early access to algebra on long-term gain in mathematics attainment. *Educational Evaluation and Policy Analysis, 18*(2), 141-153.

Stevenson, D. L., Schiller, K. S., & Schneider, B. (1994). National sequences of opportunities for learning. *Sociology of Education, 67*, 184-198.

Willett, J. B. (1994). Measuring Change More Effectively by Modeling Individual Change Over Time. In T. Husen & T. N. Postlethwaite (Eds.), *The International Encyclopedia of Education* (2nd ed.,). Oxford: Pergamon Press.

---

1. The National Council of Teacher of Mathematics has published three separate standards documents but this study's focus is on the *Professional Standards* because this document is the NCTM's most comprehensive standards document regarding pedagogy (National Council of Teacher of Mathematics, 1995; National Council of Teachers of Mathematics, 1989; National Council of Teachers of Mathematics, 1991). References to the other documents are made where appropriate.

2. The population was restricted to black and white students and eighth and ninth grade students for two separate reasons. The ethnic restrictions were made because: the number of Asian, Hispanic, and American Indian students in the target population are both limited in terms of sheer numbers (204, 105, and 14 respectively) and they are clumped into only a handful of teachers' classrooms. This restricted my statistical power and caused me to call into question my ability to make reliable estimates for these ethnic groups. The sample was restricted to eighth and ninth graders (who make-up about 85 percent of the target population) because the students in the other grades (seventh, tenth, eleventh, and twelfth) were qualitatively different and had severe missing data problems which prevented me from adequately accounting for their differences.

TM027656

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

ERIC®

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: New Teaching Standards and Old Tests: Dangerous Mismatch?	
Author(s): Daniel Mayer	
Corporate Source: The College Board	Publication Date:

II. REPRODUCTION RELEASE:

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

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

[X] **Check here For Level 1 Release:** Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical) ***and*** paper copy.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 1

The sample sticker shown below will be affixed to all Level 2 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 2

[ ] **Check here For Level 2 Release:** Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical), but ***not*** in paper copy.

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

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

Sign here→ please

Signature: [signature]	Printed Name/Position/Title: Daniel Mayer	
Organization/Address: Mathematica Policy Research, 7215 Holly Ave, Takoma Park, MD 20912	Telephone: (301) 563-6930	FAX: 563-6940
	E-Mail Address:	Date: 8/1/97

CUA

THE CATHOLIC UNIVERSITY OF AMERICA
Department of Education, O'Boyle Hall
Washington, DC 20064

800 464-3742 (Go4-ERIC)

April 25, 1997

Dear AERA Presenter,

Hopefully, the convention was a productive and rewarding event. We feel you have a responsibility to make your paper readily available. If you haven't done so already, please submit copies of your papers for consideration for inclusion in the ERIC database. If you have submitted your paper, you can track its progress at http://ericae2.educ.cua.edu.

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

We are soliciting all the AERA Conference papers and will route your paper to the appropriate clearinghouse. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality.

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

Mail to:

AERA 1997/ERIC Acquisitions
The Catholic University of America
O'Boyle Hall, Room 210
Washington, DC 20064

Sincerely,

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