Journal for Research in Mathematics Education
2001, Vol. 32, No. 4, 368–398

The Impact of Two Standards-Based Mathematics Curricula on Student Achievement in Massachusetts

Julie E. Riordan and Pendred E. Noyce, *The Noyce Foundation*

Since the passage of the Education Reform Act in 1993, Massachusetts has developed curriculum frameworks and a new statewide testing system. As school districts align curriculum and teaching practices with the frameworks, standards-based mathematics programs are beginning to replace more traditional curricula. This paper presents a quasi-experimental study using matched comparison groups to investigate the impact of one elementary and one middle school standards-based mathematics program in Massachusetts on student achievement. The study compares statewide standardized test scores of fourth-grade students using *Everyday Mathematics* and eighth-grade students using *Connected Mathematics* to test scores of demographically similar students using a mix of traditional curricula. Results indicate that students in schools using either of these standards-based programs as their primary mathematics curriculum performed significantly better on the 1999 statewide mathematics test than did students in traditional programs attending matched comparison schools. With minor exceptions, differences in favor of the standards-based programs remained consistent across mathematical strands, question types, and student sub-populations.

Key Words: Achievement; Curriculum; Elementary, K–8; Middle grades, 5–8; Reform in mathematics education

Controversy has erupted recently in several states, including California, Massachusetts, and Texas, regarding the use of “standards-based curriculum” in mathematics. Standards-based mathematics curriculum is usually interpreted to mean curriculum aligned with the content standards prepared by the National Council of Teachers of Mathematics (NCTM) (Bay et al., 1999). Otherwise known as the NCTM *Standards*, these documents are considered the most widely used of all school subject standards in the United States. The NCTM *Standards* recommend that the curriculum should place an emphasis on problem solving, reasoning, making connections between mathematical topics, communicating mathematical ideas and providing opportunity for all students to learn (NCTM, 1989, 1991, 1995, 2000). The *Standards* also encourage the teaching of certain mathematical content, including algebra, geometry, trigonometry, statistics, probability, discrete mathematics and calculus (NCTM, 1995).

We would like to thank the anonymous reviewers for their perceptive and constructive comments on an earlier draft of this article.

Standards-based programs are those written specifically to fulfill not only the content standards but also the pedagogical approaches that the standards advocate. Compared to mathematics instruction commonly observed in American classrooms today, standards-based curriculum programs place less emphasis on memorization, on manipulating numbers (e.g., long division, factoring polynomials), and less time devoted exclusively to skills development (Goldsmith, Mark, & Kantrov, 1998; Schoenfeld, 1992). Teachers are encouraged to spend less time on formal lecture and demonstration because these programs are based on the notion that "students learn by creating mathematics through their own investigations of problematic situations, and [therefore] teachers should set up situations and then step aside so that students can learn" (Kilpatrick, Swafford, & Findell, 2001, p. ix).

Although concepts about number and operations on numbers still constitute the core of mathematics learning, students in a standards-based program also spend substantial time on topics such as spatial reasoning, estimation, measurement, probability, and exploration of data (NCTM, 2000; Steen, 1990). Students using these programs are often asked to work in small groups, to come up with alternate methods for solving problems, and to describe their reasoning verbally, in writing and through multiple representations (e.g., charts, tables, diagrams). Students tend to work on fewer but more complex problems than in traditional programs, with the problems often based on real-life situations and applications. Basic skills practice tends to be embedded in real-life problems or addressed through games and activities (Goldsmith, 1998).

Calculators, computers, and other tools are identified more frequently with standards-based curricula than with traditional programs. The *Principles and Standards for School Mathematics* (NCTM, 2000) contains the clarification that calculators or other technological tools should not be used for all computations, nor should the teaching of basic skills be eliminated. In particular, the *Principles and Standards* document stresses the importance of students attaining computational fluency with whole numbers.

A contrasting view of mathematics education, which may be characterized as the traditional view, holds that "students learn by absorbing clearly presented ideas and remembering them, and that teachers offer careful explanations followed by organized opportunities for students to connect, rehearse, and review what they have learned" (Kilpatrick et al, 2001, p. ix). Academic mathematicians who adhere to this view tend to criticize the NCTM *Standards*, not necessarily because of a desire to defend the status quo of American mathematics education but because of a concern that the *Standards* are taking mathematics in the wrong direction. Such critics contend that the standards sacrifice rigor to accessibility and fail to address such fundamental problems as the insufficient mathematical knowledge of many American teachers (Askey, 1999) and the lack of mastery of arithmetic by students (Loveless, 2000; Wu, 1999).

Evidence from research studies suggests that the type of teaching prevalent in many American classrooms today more closely resembles the traditional model than the model presented in the various standards documents (Schoenfeld, 1985). For example, a recent NRC report (Kilpatrick et al, 2001) noted that K–8 instruction continues to emphasize the teaching of basic arithmetic, an emphasis that has not changed markedly since before the publication of the NCTM *Curriculum and Evaluation Standards* in 1989 (Burrill, 1997).

One rich source of data about instructional practice in middle school classrooms is the TIMSS Videotape Classroom Study of 1994–95, which compared a set of eighth-grade classrooms in Germany, Japan, and the United States. Stigler and Hiebert (1997) found that the most common goal of U.S. lessons was to teach students how to do something, whereas the goal of Japanese lessons was to enhance student understanding of mathematical concepts. Compared to Japanese teachers, American teachers spent more time reviewing and less time presenting new material. In American classrooms, a majority of mathematics topics (78%) tended to be stated or rather than explained or developed. American teachers asked more yes/no questions and fewer questions requiring students to describe or explain their work than did their Japanese colleagues. Furthermore, for 96% of the time American students were doing work at their seats, they were practicing procedures that they had already been shown how to do (Stigler & Hiebert, 1997; see also Baker & Schaub, 1991).

Changing teacher practice is difficult (Hiebert, 1999; Richardson, 1990). One goal of the developers of standards-based curriculum has been to provide teachers with materials that will allow and encourage them to align their teaching practice with the principles of the NCTM *Standards*. For example, the materials may direct the teacher to have students work in small groups discussing mathematics, explaining their reasoning, and coming up with multiple approaches to solving a problem. In standards-based programs, lessons are often introduced by presenting students with an unfamiliar problem rather than a worked example (Goldsmith et al., 1998). Although materials alone cannot change teacher practice, they can provide scaffolding for teachers trying to create a classroom environment different from that observed in the TIMSS Videotape Classroom Study. Nevertheless, introduction of standards-based programs has not proceeded without controversy.

Controversy over the mathematics curriculum increased following the U.S. Department of Education's release of its listing of "Exemplary and Promising Mathematics Programs." The programs were reviewed by a panel of educators, scientists, and policymakers, including representation from one college, two universities, a state department of education, the National Research Council, the National Science Foundation, the National Science Teachers Association, and the National Alliance of State Science and Mathematics Coalitions. The panel's purpose was to establish a reliable, research-based process for assessing mathematics and science programs, and then to identify exemplary and promising programs (U.S. Department of Education, 1999). Among the programs submitted were the two programs considered in this article, *Everyday Mathematics* (Bell, 1988–1996) and *Connected Mathematics* (Lappan et al, 1991–1997). The review process encompassed a rating of submitted programs by two field-based teams on the basis of program quality, usefulness to others and educational significance. Designation of a program as either "promising" or "exemplary" required further review by program evaluation experts of data showing evidence of gains in student achievement. The distinction between exemplary and promising programs was that exemplary programs provided convincing evidence of effectiveness in multiple sites with multiple populations, whereas promising programs demonstrated preliminary evidence of effectiveness in different sites. *Everyday Mathematics* received a promising rating, and *Connected Mathematics* was identified as exemplary**.**

Soon after the report was released, a group of mathematicians and others wrote an open letter to then Secretary of Education Richard Riley protesting the publication of this report. The letter, appearing as an advertisement in *The Washington Post* and endorsed by over 200 mathematicians, scientists, and educators, denounced the listing as premature and criticized it for promoting programs with "serious mathematical shortcomings" (An Open Letter …, 1999, p. A-5).

Also published in 1999 was a set of reviews conducted by an organization called Mathematically Correct, advocating for improved mathematics education in America's schools. The cofounders of the group reviewed approximately 30 mathematics programs in Grades 2, 5, and 7. These reviews addressed depth of content, quality of presentation, quality of student work, and the depth of student learning likely to occur, but they did not examine any student performance data. In the reviews, *Everyday Mathematics* and *Connected Mathematics* received low ratings, in contrast with the high ratings they had received from the U.S. Department of Education's panel.

As noted above, all programs designated as promising or exemplary by the U.S. Department of Education panel provided evidence of student achievement gains that met the standards of a panel of experts in program evaluation. However, studies submitted to the panel were based on limited samples of early implementers and therefore may have focused on schools where conditions for program implementation were especially favorable. In addition, many of the studies were carried out or reported by the developers or evaluators of the programs. Therefore, some degree of research or selection bias may have entered into the selection of school samples.

The study reported here seeks to determine the impact of two standards-based mathematics programs on student achievement in Massachusetts. The two programs are *Everyday Mathematics* at the elementary school level and *Connected Mathematics* at the middle school level, and the outcome measure was student performance on the Massachusetts statewide test in mathematics. The hypothesis tested in this study was that students in schools adopting certain standards-based programs perform better than those in matched comparison schools on standardized tests aligned with national content standards and, in addition, that these schools demonstrate greater gains in student performance over time.

Schools using either *Everyday Mathematics* or *Connected Mathematics* were selected for the study on the basis of how long each program had been used (and in the case of *Connected Mathematics*, how many units had been implemented). Comparison schools were then selected on the basis of past school performance on an earlier statewide test and student socioeconomic status. The selection process used in this study attempted to make the target curriculum and comparison groups as similar as possible so that few, if any, statistically significant differences existed between them. We then performed a series of analyses comparing test score performance between the target curriculum and comparison student populations.

Before discussing the results of these analyses, we describe both the testing system in Massachusetts and the two standards-based programs examined in this study. We then provide a summary of the literature pertaining to the effect of these programs on student achievement before explaining our methods and results.

The Massachusetts Testing System

Biennially from 1988 to 1996, fourth- and eighth-grade students in regular classrooms in Massachusetts schools took a standardized test, the Massachusetts Educational Assessment Program (MEAP). Tenth grade was added to the testing schedule in 1994. MEAP used a matrix-sampling technique that provided school-level and district-level results but no individual student scores. In 1993, the state legislature passed a sweeping education reform act requiring, among many other provisions, the establishment of curriculum frameworks and new assessments based on state standards in core academic areas. After being circulated in draft form for two years, the Massachusetts Mathematics Curriculum Framework, *Achieving Mathematical Power*, was approved by the Board of Education in 1995 and distributed the following year (Massachusetts Department of Education [MA DOE], 1996). The Framework was based in large part on the NCTM *Curriculum and Evaluation Standards* (1989) and emphasized student reasoning and conceptual understanding rather than a more skills-based approach. By 1997, new assessment questions in mathematics were being field-tested, and in 1998 the new assessment system, the Massachusetts Comprehensive Assessment System (MCAS), was administered for the first time to all students in Grades 4, 8, and 10. Like MEAP, MCAS consists of both multiple-choice and constructed-response items, the latter including both short answer questions and those requiring a more extended response. MCAS is a criterion-referenced test, based directly on the Framework and its accompanying Assessment Guide (The Framework and the Assessment Guide are available at the following website: http://www.doe.mass.edu/frameworks). Unlike MEAP, MCAS is administered every year and provides individual student results as well as school and district-level results.

The mathematics section of MCAS covers four strands of mathematics: Number Sense; Patterns, Relations and Functions; Geometry and Measurement; and Statistics and Probability. Each strand is tested with open-response, short-answer, and multiple-choice items, and each strand accounts for at least 20% of total points in both Grade 4 and Grade 8. Approximately half of a student's score is based on multiple-choice or short-answer questions and half on open-response questions. Raw scores are converted to scaled scores that range from 200–280. A score below 220 is rated as Failing; a score between 220–239 is rated as Needs Improvement; a score between 240–259 is considered Proficient and a score of 260 or more is considered Advanced.

Student scores and overall school scores are based on common items that are administered to all students. Common test items are those that are identical in all 12 forms of the test administered at each grade level. Approximately 80% of the questions on any given test are common questions, and all common questions are released to the public following test administration each year. The remaining 20% of the MCAS questions on each test form are matrix-sampled items, which differ across the 12 test forms at each grade level tested. Matrix-sampled items serve as the basis for equating tests from year to year, in order to allow tracking of school and district performance over time. Matrix sampled items, along with the common items, allow the calculation of sub-scores on different strands of mathematics and different question types (open response, short answer, multiple choice), reported at the school and district level only.

MCAS is scored on an 80-point scale from 200 to 280, whereas MEAP was scored on a 600-point scale from 1000 to 1600. Despite these changes, correlation studies show considerable stability in school performance as the state moved from one assessment system to the other. Thus, the Pearson correlation coefficient between school 1996 MEAP and 1998 MCAS scores in Grade 4 mathematics is .737, whereas for Grade 8 mathematics, the correlation coefficient between school performance on the two tests is .901.

A technical review of the MCAS completed after the first year of administration concluded that the test was valid, reliable, and challenging, and that it was accurately and consistently scored (MA DOE, 1999). Reliability of the mathematics score was estimated at .87 for Grade 4 and .91 for Grade 8. A supplementary study conducted in one large district found that the correlation between students' eighth-grade mathematics MCAS scores and their eighth-grade mathematics scores on a national standardized test, the Stanford-9, was .84.

Description of Two Standards-Based Curriculum Programs

Everyday Mathematics was developed by the University of Chicago School Mathematics Project. This K–6 curriculum involves students working in small groups to explore mathematical concepts using manipulatives, calculators, and other mathematical tools. Students are encouraged to invent and share multiple methods of solving problems. Along with computation and number operations, mathematical topics taught in the curriculum include data and probability, geometry and spatial sense, measures and measurement, numeration and order, and patterns, functions and sequences. Preliminary study of algebra and the use of variables begin in the third grade. As mentioned earlier in this article, the U.S. Department of Education (1999) listed *Everyday Mathematics* as a promising program. In contrast, a review by the more traditional California-based Mathematically Correct group assigned *Everyday Mathematics* a grade of C at second grade and a grade of C– at fith grade (Clopton, McKeown, McKeown, & Clopton, 1999).

Everyday Mathematics curriculum materials for kindergarten through Grade 3 were published in 1993, with materials for subsequent grades appearing approximately yearly with the Grade 6 materials published in 1996. The first schools to implement *Everyday Mathematics* in Massachusetts selected the program because it was perceived to be more challenging and advanced than traditional textbook programs, introduced the materials grade by grade as they became available, and began using fourth-grade materials in pilot form during the 1993–94 school year. Thus, the last MEAP test administered in these schools to fourth-grade classes who had not been using *Everyday Mathematics* was in 1992. Therefore, in our study we used the 1992 MEAP as the baseline score for early implementers of *Everyday Mathematics* and their comparison schools. For later implementers of *Everyday Mathematics*, 1996 is used as the baseline year, the last year before fourth graders in these schools used the program. By 1999, approximately 7% of elementary schools in Massachusetts had been using this program as their core curriculum for at least two years.

The *Connected Mathematics* curriculum was developed at Michigan State University from 1991 to 1997 under a grant from the National Science Foundation. *Connected Mathematics* is a problem-centered curriculum for students in Grades 6–8. The program divides a mathematics lesson into three parts—launching the problem, exploring the problem, and summarizing the problem with extensions to help students practice what they have learned. Students demonstrate what they have learned through assessment, including partner quizzes, projects, unit tests, and self-assessment. The program includes a recommendation that students keep journals and that they have access to scientific or graphing calculators at all times. In each of the three grades, topics in number, algebra, geometry, and probability/statistics are covered in an increasingly sophisticated manner. As in *Everyday Mathematics*, group explorations and the use of technology are encouraged. As mentioned previously, *Connected Mathematics* was listed as exemplary by the U.S. Department of Education, and received a top rating by the American Association for the Advancement of Science (AAAS) in its review of middle school mathematics programs (AAAS, 2000); Mathematically Correct, however, assigned it a grade of F.

Beginning in 1994, *Connected Mathematics* and other standards-based science and mathematics curricula were introduced in Massachusetts through a series of curriculum showcases organized by the Center for the Enhancement of Science and Mathematics Education (CESAME) at Northeastern University. CESAME also provided initial funding, professional development, and technical assistance to a number of school districts wanting to use *Connected Mathematics*, including one school that began implementation in all three grades during the 1995–96 school year and several of those that began implementation during the 1997–98 school year. In our study, we considered 1994 as the baseline year for eighth-grade students in the first school that implemented the program in Massachusetts and its comparison schools. For all schools that followed, 1996 is considered the baseline eighth-grade MEAP year. By 1999, approximately 5% of middle schools in the state had been using *Connected Mathematics* as their core program for at least two years.

Summary of Relevant Literature

Although many evaluations of standards-based mathematics programs have been limited to field studies conducted by the developers of the curricula, these studies provide initial trend data concerning student achievement. For example, Carroll (1997) reported positive results on the Illinois statewide standardized test for students using the *Everyday Mathematics* curriculum in the Chicago area, compared to students in a suburban county not using the program and also compared to state scores, even though the test format (consisting of 60 multiple choice questions) was not aligned with the curriculum. Schools with the greatest number of disadvantaged students scored above both the comparison schools and the state. Moreover, Carroll reported higher test scores for students who had experienced *Everyday Mathematics* since kindergarten than for students who had been in the program for only one or two years.

In the same study, another comparison was made between one district using the *Everyday Mathematics* program and similar districts (matched on the basis of school size, per pupil spending, and student demographics) not using the program. Third-grade students in the *Everyday Mathematics* district scored significantly higher on the Illinois statewide test than three of the four comparison districts and did not differ significantly from the fourth district (see also Carroll, 1995). In another study, Carroll and Porter (1994) found that fourth-grade *Everyday Mathematics* students scored as well on traditional items as they did on reform-oriented items on the Illinois standardized test. Carroll (1996) also found that in a comparison of performance on 25 mental computation problems administered at the fifth grade, students in a class using *Everyday Mathematics* outperformed the students in traditional classes on all but one of the problems.

A study done by Briars and Resnick (1999) examined the impact of *Everyday Mathematics* on achievement when implemented as part of systemic change in the Pittsburgh (PA) public schools. Their study compared scores on a statewide test from 1996, 1997, and 1998 for all fourth-grade students in the district, and they found overall improvement during this time in all competency levels (designated as skills, concepts, and problem solving). Schools were classified either as strong or weak implementers on the basis of how many teachers in the school were using all of the *Everyday Mathematics* components and were also providing student-centered instruction. Strong implementers demonstrated significantly higher gains than weak ones, even in schools with large numbers of poor and minority students.

A number of researchers report the same type of positive results for students using *Connected Mathematics*. Ben-Chaim et al. (1997) found evidence that students using *Connected Mathematics* performed better on proportional reasoning tasks presented in different contexts than students not using *Connected Mathematics*. Over all types of questions, *Connected Mathematics* students performed approximately 50% better than students in non-*Connected Mathematics* classes. In addition, the results from this study suggest that seventh-grade students using *Connected Mathematics* increased their proportional reasoning abilities by the end of the eighth grade with no further formal study of proportional reasoning. In a similar study, Lapan, Reys, Barnes, and Reys (1998) examined the impact of two standards-based curricula (one of which was *Connected Mathematics*) on mathematics achievement. Mathematical problem solving skills were significantly higher for students using standards-based curricula than for those not using either of the curricula examined in the study.

In five Minneapolis schools fully implementing *Connected Mathematics*, Winking, Barel, and Ford (1998) found that most eighth-grade students significantly outscored their counterparts in comparison sites on the State Basic Standards Tests. Results for partially implementing schools were modest or neutral in gain compared to students not using *Connected Mathematics*. Finally, in an evaluation report for the Arkansas Statewide Systemic Initiative, O'Neal and Robinson-Singer (1998) examined the progress of students in eight *Connected Mathematics* pilot school districts one year after implementation and found that statewide standardized test score gains in mathematics were positive and statistically significant for *Connected Mathematics* students. In addition, students in almost all participating districts made gains in mathematics test scores on the Stanford-9 test.

Other research studies report the success of certain features and practices common among standards-based programs (not just *Everyday Mathematics* or *Connected Mathematics)*. One study (Wood & Sellers, 1997) involved a comparison between students who had received two years of problem-centered instruction (e.g., working in small groups to solve mathematics problems, discussing solutions with the whole class, etc.) and those who had used a more traditional textbook in class. Not only did students in reform-oriented classroom perform better on norm-referenced standardized tests in Grades 1 through 4, they demonstrated greater conceptual understanding in numeration, place value, and multiplication. Similarly, in her comparison study of one traditional and one reform-oriented school in England, Boaler (1998) found that students in the traditional school were much less able to apply their mathematical knowledge to novel or real-life situations than students in the reform-oriented school. Although students in the reform-oriented school spent less time working on mathematics (and this was seen as a drawback), these students nonetheless outscored the comparison school students on tests and applied mathematics problems.

Student achievement can also be linked to the instructional practices in schools that have adopted a standards-based approach to teaching mathematics. The QUASAR Project, a national study of middle school mathematics reform in economically disadvantaged communities, provides evidence that the nature of mathematical tasks used in the classroom influences student learning outcomes. In particular, Stein, Grover, and Henningsen found that "the construct of the mathematical task was found to be a useful focusing device—one that served to highlight mathematical content and processes" (1996, p. 484). This study presented data about students using a range of reform-oriented curricula and concluded that even teachers whose background characteristics did not differ from most middle school teachers' could be successful in setting up and delivering tasks that required high-level mathematical reasoning. In another study, Stein, Lane, and Silver (1996) examined several ethnically diverse middle schools and found that instructional programs characterized by tasks that required nonstandard, complex, or conceptual forms of thinking, rather than an emphasis on procedural drill and well-rehearsed algorithms, led to greater gains in student learning as measured by an assessment developed for the QUASAR project. This assessment consisted of open-ended tasks that measured performance in problem solving, reasoning, and communication and included tasks on number and operation, estimation, patterns, pre-algebra, geometry, measurement, probability, and statistics.

METHOD

In this section, we describe data used in this study, our method of identifying schools using either *Everyday Mathematics* or *Connected Mathematics* (hereafter referred to as the target curriculum group) and the process of selecting comparison schools. We also present a summary of the teachers' characteristics in both the target curriculum and comparison schools.

Data

The primary sources of data for this study were school test results from the Massachusetts Educational Assessment Program (MEAP) administered between 1992 and 1996, and both school and individual student results from the 1999 Massachusetts Comprehensive Assessment System (MCAS). MCAS test results can be disaggregated by race, gender, regular or special education status, free and reduced price lunch status, and English proficiency. For each school, the average percentage of possible points attained for four mathematical strands and three question types is also reported and can be used to compare school performance on different kinds of mathematical tasks.

We obtained enrollment data at the school level concerning eligibility for free and reduced price lunch and race from the 1999 Individual Public School Report, a database managed by the Massachusetts Department of Education. Information regarding curricula used in Massachusetts was taken from another database—the Mathematics, Science and Technology Survey administered in 1999 by the Massachusetts Department of Education**.** District personnel were required to identify the mathematics and science curriculum programs used in their schools and to indicate whether the program(s) were being explored, piloted, or implemented . For the purposes of this study, programs that were identified by schools as being implemented (rather than just explored or piloted) were considered the primary curriculum. The 1999 Teacher Questionnaires administered with the MCAS by the Massachusetts Department of Education provided data about teacher characteristics and instructional practices.[1]

---

[1] This questionnaire was not specifically designed for this study but, rather, was administered by the Massachusetts Department of Education in 1998 and 1999 to fourth- and eighth- grade teachers in Massachusetts. Items on the questionnaire relating to teaching practice were derived from a national study performed by RAND, which classified the items as reflecting either "reform" or "traditional" teaching practice. These items can be found in Appendix 1.

Identification of the Target Curriculum Group

We used several sources of information to identify all schools in Massachusetts using either *Everyday Mathematics* or *Connected Mathematics* as their primary mathematics program. The publishers of *Everyday Mathematics* provided a list of all schools that had purchased the program, along with the date of purchase. The Center for the Enhancement of Science and Mathematics Education (CESAME) provided a similar list of middle schools known to be implementing *Connected Mathematics*. Both lists were cross-checked against the statewide curriculum survey, the 1999 Mathematics, Science and Technology Survey. Use of the curriculum had to be confirmed by both the survey and either by CESAME or by the publisher in order for the school to be included in the study.

In addition, because *Connected Mathematics* is a modular program often implemented a few units at a time, we conducted telephone interviews with school personnel to determine the extent to which the curriculum was used. Among schools surveyed using *Connected Mathematics* in Massachusetts, we found that the maximum number of units that could be reasonably implemented during the course of a school year was six units per grade, even though there are eight units available for each grade. Only schools implementing at least 11 *Connected Mathematics* units in Grades 6 through 8 by 1998–99 were included in this study.

A total of 67 schools were identified as using *Everyday Mathematics,* with 48 of them having implemented this program for four or more years. This group is identified as *EM* Group I in this study and also referred to as "early implementers" because the schools in this group were among the first to use *Everyday Mathematics* in Massachusetts. By 1999, the remaining 19 schools had implemented the program for two or three years and are identified as *EM* Group II, also referred to as "later implementers" in this article.

Altogether, 21 middle schools were identified as using *Connected Mathematics*. One school had implemented the program for four years and is identified in this study as *CMP* Group I (also called "early implementer"). Twenty schools had used the program between two and three years and are identified as *CMP* Group II (also as "later implementers").

Beyond a simple measure of length of use, we did not seek to distinguish the level of implementation (e.g., percentage of teachers using the curriculum, units or chapters covered per grade, professional development provided for teachers) for either *Connected Mathematics* or *Everyday Mathematics* among the schools in the target curriculum group. Results from previous research studies (e.g., Boaler, 1998; Stein, Lane, & Silver, 1996) suggest that curriculum materials and instructional practices can be effective even when the curriculum is not optimally implemented, and our intent in this study was to determine the impact of the two curriculum programs as they were actually (but perhaps not optimally) used in Massachusetts schools.

Using the data collected by the Massachusetts Department of Education, we examined the characteristics of the target curriculum schools (e.g., previous achievement on statewide tests, the percent of students eligible for free and reduced price lunch, racial and ethnic makeup). These characteristics then became the benchmarks for choosing the comparison group.

Selection of the Comparison Group

Having identified target curriculum schools according to criteria for inclusion in this study and then placed them into four groups, we sought to identify for each group a comparison set of schools that had not implemented either *Everyday Mathematics* or *Connected Mathematics* but that were similar in how they would be expected to perform on the statewide test. To do this, we conducted a statewide analysis to determine the strongest predictors of student achievement on the 1999 MCAS in mathematics.

A stepwise multiple regression analysis provided information about the strongest predictors of school score on the 1999 mathematics MCAS. We considered the following variables: baseline mean school performance on the previous statewide test (MEAP), percent of students receiving free and reduced price lunch, percent of students in various ethnic groups (Asian, Black, Hispanic, White), percent of students who had limited English proficiency, and percent of students who required special education services. Consistently, two variables—baseline mean MEAP score and percent of students receiving free or reduced price lunch—accounted for between 66% and 81% of the total variance in 1999 average school-level mathematics scores. Including more than these two variables in the regression equations led to problems of multicollinearity without explaining more than another 5% of variance.

Therefore, in our study, schools within each of the four groups were *first* matched on the basis of mean MEAP score in the last year prior to the introduction of the standards-based curriculum in the target curriculum schools. After matching by MEAP at the school level, we further matched schools on the basis of the percentage of students in each school eligible for free and reduced price lunch. However, because the schools in the target curriculum and comparison groups differed in size (i.e., number of students attending the school), the aggregate student populations attending these schools still differed in the overall percentage of students receiving free and reduced price lunch. To adjust for this discrepancy, we then added more schools to the comparison group—again attending to the degree of match on past MEAP scores—until the student population of each comparison group was closely matched to the student population of the corresponding target curriculum group.

Table 1 contains data for each target curriculum group and its comparison group in terms of prior school achievement, free lunch status, and the percentage of students who are White. For each curriculum, Group I designates early implementers—those schools where the target curriculum had been in use for at least four years. Group II designates later implementers, that is, schools where the target curriculum had been in use for two to three years. The data in the table suggest that in the aggregate, schools included in this study had a relatively low percentage of students receiving free or reduced price lunch, and overall, about 89% of the students were White. On average, both target curriculum and comparison schools had performed at a level above the state mean on previous statewide tests. The table reveals only small differences—most of them not statistically significant—between target and comparison groups in the characteristics listed. There were no significant differences in baseline MEAP score, the variable most predictive of school MCAS performance. Where small differences in demographic variables (race/ethnicity or free lunch status) occurred, they would not be predicted to make a material difference in the major study findings.[2]

Table 1
Characteristics of Target Curriculum and Comparison Groups

	N Schools	*N* Students	Weighted Mean School MEAP[a]	% Free/Reduced Price Lunch	% White
EM Group I	48	2914	1443 (92)	6.0*	89.0
Comparison	51	3095	1445 (92)	8.1*	90.9
EM Group II	19	867	1333 (96)	9.0	88.9*
Comparison	27	1917	1332 (96)	10.6	91.9*
State	1047	63461	1322 (92) 1325 (96)	32.2	75.1
CMP Group I	1	73	1300 (94)	0.0	95.5
Comparison	4	609	1300 (94)	3.4	93.1
CMP Group II	20	1879	1370 (96)	10.3	80.8*
Comparison	30	4978	1370 (96)	11.0	88.2*
State	408	59623	1299 (94) 1320 (96)	28.3	76.1

Note. Early *Everyday Mathematics* implementers (*EM* Group I) began to use the curriculum in 1994 and later implementers (*EM* Group II) began to use the curriculum after 1996. The early *Connected Mathematics* implementer (*CMP* Group I) started to use the curriculum in 1995 and later implementers (*CMP* Group II) after 1996. Number of students (*N* Students) indicates the number of regular education grade 4 or grade 8 students who took the 1999 MCAS.

[a] Baseline MEAP year is specified in parentheses. The mean school MEAP score was weighted by the number of students who took the MEAP mathematics test in the year specified. The comparison of MEAP scores represents achievement *prior* to implementation of either *Everyday Mathematics* or *Connected Mathematics*.

*Difference between target curriculum and comparison schools for this variable is statistically significant at $p < .05$.

Information on curricula used by the matched comparison schools was obtained from the 1999 Mathematics, Science and Technology Survey administered by the Massachusetts Department of Education. At the elementary level, 78 comparison schools reported using 15 different textbook programs, with the most commonly

[2] Based on the multiple regression model we used, the finding that the comparison group for early implementers of *Everyday Mathematics* had 2.1% more students eligible for free and reduced price lunch translates to an expected score advantage for the target curriculum group of about 0.3 points on the 1999 MCAS. Where there were significant differences in the percentage of White students, the expected impact of these differences on school scores cannot be accurately assessed because of multicollinearity.

used programs being those published by Addison-Wesley, Houghton-Mifflin, and Scott Foresman. These three textbook programs were used in 56% of the schools in the comparison group and in 55% of schools statewide. Three schools reported using district-designed programs, and several more schools reported using more than one textbook or curriculum program. The range and relative proportions of curricula used in the comparison group were similar to the range of curricula (excluding *Everyday Mathematics*) used by all elementary schools in the state.

At the middle school level, 34 comparison schools reported using 15 different textbook programs, with the most commonly used programs being those published by Heath, Addison-Wesley, Prentice Hall, and Houghton-Mifflin. These four textbook series were used in 53% of the comparison schools and in 50% of all middle schools in Massachusetts. Four schools in the comparison group reported using district-designed curriculum materials. The range and relative proportions of curricula used in the comparison schools was similar to the range of curricula (excluding *Connected Mathematics*) used by all middle schools in the state.

It should be noted that this study was not designed to provide a head-to-head comparison of *Everyday Mathematics* or *Connected Mathematics* with individual traditional curriculum programs. Rather, we compared the two standards-based programs against a range of curricula that, in the aggregate, represent the instructional norm in Massachusetts.

Teacher Characteristics

After completing the matching process for schools, we examined the self-reported characteristics of the teachers and information about teaching practices in the target curriculum and comparison groups. It could be argued that the performance of students is related more to the qualifications of their teachers than to the impact of the curriculum. Therefore, a comparison was made between teachers in the target curriculum and comparison groups on the basis of responses to the questionnaire administered during the MCAS to all fourth-, eighth- and tenth-grade principals and teachers in 1998.

Questionnaire items that related to teacher characteristics included questions about the number of hours professional development received in the last two years; years of teaching experience; type of certification (Grade 8 only); the number of college mathematics courses completed in mathematics; membership in a professional organization; and the frequency of attendance at mathematics conferences. All differences in these characteristics were small and not statistically significant at $p < .05$ except for a difference for late implementers of *Connected Mathematics* (*CMP* Group II), where the comparison group of teachers was slightly more likely to be certified in mathematics than the teachers in *CMP* Group II, a difference that would seem to advantage that group.

Questionnaire items that related to teaching practice (shown in Appendix 1) were derived from a national RAND study (Klein et al., 2000) examining the impact of reform and traditional teaching practices on student achievement in mathematics and science at elementary and middle school levels. Massachusetts was one of several sites that participated in the study in its second year. Using the same methodology as RAND, we conducted a factor analysis of the 1998 teacher questionnaire. This analysis confirmed that the questions fell into the two components designated by RAND—one cluster of reform-oriented practices (e.g., having students explore alternate solutions to problems or using open response questions) and one cluster of traditional practices (e.g., lecturing to students or asking students to memorize facts, rules or formulas).

Again, following the RAND model, we constructed two scales, one of reform practice and one of traditional practice. We then compared the responses of teachers in our four target curriculum groups to those of their matched comparison groups and to all the mathematics teachers statewide at that grade level. Surprisingly, we found that the responses of teachers in three of our target curriculum groups did not differ significantly from the responses either of their matched comparison groups or the average responses of all teachers at their grade level statewide. The exception was *CMP* Group I, for which only one teacher completed the questionnaire. This teacher scored significantly higher on the reform scale and lower on the traditional scale than the matched comparison teachers and other eighth-grade teachers in the state, although is difficult to make generalizations based on the responses of only one teacher.

RESULTS

The study reported in this article employed a post-treatment study using matched comparison groups in a quasi-experimental design.[3] As indicated earlier, the goal of this study was to examine the impact of curriculum on student achievement. To that end, we examined Massachusetts standardized test score data at the student level and the school level. At the student level, we examined the difference in 1999 test scores between the target curriculum and comparison groups. We then disaggregated student results by race, gender, and free lunch status.

At the school level, we compared gains in performance made by the target curriculum and comparison schools between the year prior to implementation of the standards-based curricula and 1999. Also at the school level, we compared the performance of target curriculum schools to comparison schools on each mathematical strand and question type on the MCAS. A detailed explanation of each analysis follows.

---

[3] Although applying analysis of covariance does not always make groups equal (see Lofton & Madison, 1991), we did consider ANCOVA as an experimental design. However, in this study it was necessary to exclude schools in which only a few classrooms were using *Everyday Mathematics* or *Connected Mathematics*. Second, the covariates—in this case, test scores, income, and race—are so interrelated that multicollinearity becomes a problem.

Comparison of Individual Student Results and Effect Sizes

In a first analysis, we compared average 1999 individual mathematics scores between the target curriculum and comparison groups. Only regular education students who had attended school in the district for three or more years were included. As shown in Table 2, students using the *Everyday Mathematics* or *Connected Mathematics* curricula outscored their counterparts, with score differences ranging from 2.5 points to 5.7 points on an 80-point scale that ranges from 200 to 280. All differences were statistically significant. For comparison purposes, the statewide mean for the 1999 MCAS for mathematics is also presented in this table for each grade. The results also show that students in target curriculum schools that had been using *Everyday Mathematics* or *Connected Mathematics* longer—Group I schools—outscored their matched counterparts by more points than did Group II schools. That is, a longer implementation in the school was associated with a greater score advantage for students. For instance, *Connected Mathematics* students in the first two or three years of school implementation performed 4.0 points better than their counterparts, whereas *Connected Mathematics* students in the fourth year of implementation performed 5.5 points better. Similarly, *Everyday Mathematics* students in the first two or three years of school implementation scored 2.5 points better than their counterparts, whereas *Everyday Mathematics* students in their fourth year of school implementation scored 5.7 points better.

Table 2
Comparison of 1999 MCAS Mathematics Scores for Target Curriculum and Comparison Groups (with Effect Sizes in Parentheses)

	Group I *M*	Group II *M*	Statewide *M*
EM Students	249.5	241.8	236.9
Comparison Students	243.8	239.3	
Difference	5.7** pts	2.5** pts	
	(.34 *SD*)	(.15 *SD*)	
CMP Students	239.0	239.3	229.2
Comparison Students	233.5	235.3	
Difference	5.5 pts*	4.0 pts**	
	(.33 *SD*)	(.22 *SD*)	

Note. For comparison, 1999 statewide mean scores are shown for regular education students who had been in their districts for at least three years.
* $p < .05$
** $p < .001$

We calculated effect sizes to determine the magnitude of the effect of the type of curriculum on student mathematics test scores.[4] Effect sizes less than 0.1 are

[4]Although the mean difference between target curriculum and comparison groups can be divided by the comparison group standard deviation, this has been shown to bias results, and therefore it is recommended to use the average or "pooled" standard deviation (Hedges, 1981), as we have done in this study.

considered trivial; effect sizes between 0.1 and 0.3 are small; effect sizes of 0.3–0.5 are considered moderate (Rosenthal & Rosnow, 1984; Courtina & Nouri, 2000). All effect sizes we found were positive, and they ranged from small to moderate. The smallest effect size (0.15) occurred for late implementers of *Everyday Mathematics*. For early implementers of both programs, the effect size is about 0.34. Given the close match between the target curriculum and comparison schools, the moderate effect sizes suggest that the relationship between curriculum and achievement is unlikely to be spurious.

Performance of Different Student Subpopulations

In a second analysis, we examined whether the advantage provided by exposure to *Everyday Mathematics* or *Connected Mathematics* was consistent across different student populations. Once again, we included only regular education students who reported having been in the same school district for three years or more. This information was reported by students on a questionnaire given during the 1999 administration of the MCAS. The number of regular education students who reported having been in the district for three years or more represented approximately 80% of the students who took the test in 1999. We considered the following student groups: Asian, Black, Hispanic, and White; free and reduced price lunch and full price lunch; female and male. Table 3 shows the comparison of 1999 individual mathematics scores for these subpopulations between target curriculum and comparison groups. Scores for *CMP* Group I and its comparison groups were not computed because the sample size was too small for Blacks, Asians, Hispanics, and students eligible for free and reduced price lunch (i.e., less than 10 students in a school).

There was no student group for which exposure to a traditional curriculum resulted in a significantly higher score than exposure to a standards-based curriculum. Generally, *Everyday Mathematics* and *Connected Mathematics* students outperformed the comparison students. For *Everyday Mathematics*, these differences were greater among students in schools that had used the curriculum longer (Group I—early implementers). For all *Connected Mathematics* students and for early implementers of *Everyday Mathematics*, the positive score differences for Black and Hispanic students were greater than for White students.

Similar patterns were seen when we compared the performance of target curriculum and comparison group students eligible for free and reduced price lunch. Free and reduced price lunch students in schools using *Connected Mathematics* or in early implementing *Everyday Mathematics* schools outperformed their counterparts, and these differences were greater than they were for students paying full price for lunch.

Both males and females in the standards-based programs outperformed their counterparts in traditional programs, although for boys using *Connected Mathematics*, these differences were not statistically significant. Among students

Table 3

Comparison of 1999 MCAS Mathematics Scores for Target Curriculum and Comparison Groups for Student Subpopluations (with n *in parentheses)*

	Asian Students *M*	Black Students *M*	Hispanic Students *M*	White Students *M*	Free/Reduced Lunch Students *M*	Non-Free/Reduced Lunch Students *M*	Female Students *M*	Male Students *M*
EM Group I	254.1	234.1	240.2	250.0	234.9	250.4	249.6	249.4
Comparison	248.5	225.1	227.2	244.6	228.9	245.1	243.2	244.5
Difference	5.6*	9.0**	13*	5.4**	6.0**	5.3**	6.4**	4.9**
	(210)	(152)	(59)	(4279)	(351)	(4581)	(2592)	(2345)
EM Group II	249.6	223.0	225.3	243.1	230.0	243.0	241.0	242.7
Comparison	248.1	224.7	225.1	240.1	229.8	240.4	238.7	240.0
Difference	1.5	–1.7	.2	3.0**	.2	2.6**	2.3*	2.7*
	(54)	(73)	(27)	(1839)	(227)	(2017)	(1145)	(1099)
CMP Group I	—	—	—	239.1	—	239.0	240.2	237.6
Comparison	—	—	—	233.1	—	233.8	233.2	234.0
Difference	—	—	—	6.0	—	5.2*	7*	3.6
	—	—	—	(505)	—	(525)	(297)	(244)
CMP Group II	248.5	225.5	225.9	240.3	227.0	240.7	238.4	240.2
Comparison	244.2	221.2	214.2	236.1	221.2	237.0	234.4	236.3
Difference	4.3	4.3	11.7**	4.2**	5.8**	3.7**	4.0**	3.9
	(201)	(169)	(159)	(4856)	(617)	(5092)	(3000)	(2711)

Note. Results are reported only for students who had attended school in the same school district for three years or more. Scores by race are not reported for approximately 5% of students whose race was recorded as “other” or was not reported. The number of students who had attended school in the district for less than three years, or for whom length of attendance was unknown was as follows: *EM* Group I and match – 1083; *EM* Group II and match – 557; *CMP* Group I and match – 142; *CMP* Group II and match – 1147.

— Scores for subpopulations with fewer than 10 students are omitted, following the convention used by the Massachusetts Department of Education to protect student identities.

*$p < .05$

**$p < .001$

in schools that had been using the target curriculum for at least four years (for both *Everyday Mathematics* and *Connected Mathematics*), the score advantage for girls was greater than the score advantage for boys.

Performance of Students Across the Achievement Spectrum

Despite consistently better MCAS test scores for the target curriculum group seen across student groups in this analysis, we considered the possibility that the standards-based curriculum programs might not adequately address the needs of students at the upper or lower ends of the achievement spectrum. In a third analysis, we compared the full distribution of student performance for each program. Within the score distribution for each group of students (early and later implementers of *Everyday Mathematics* or *Connected Mathematics*), we compared the mean scores of students within quartiles to the mean score of students at the equivalent quartile in the score distribution of comparison schools. Differences in scores by quartile among groups are presented in Table 4. Within each quartile, for both early and later implementers, students in the target curriculum programs had higher scores. Furthermore, these differences were significant. These results suggest that *Everyday Mathematics* or *Connected Mathematics* is effective for all students, not just those at the bottom, middle, or top of the achievement spectrum.

Table 4.
*Comparison of 1999 MCAS Mathematics Scores for Target Curriculum and Comparison Groups within Quartiles (*n *in parentheses)*

	Quartile I ($\leq 25^{th}$ percentile) *M*	Quartile II (26^{th}–50^{th} percentile) *M*	Quartile III (51^{st}–75^{th} percentile) *M*	Quartile IV ($\geq 76^{th}$ percentile) *M*
EM Group I	228	244	257	270
Comparison	224	238	251	266
Difference	4*	6*	6*	4*
	(1355)	(1167)	(1239)	(1176)
EM Group II	223	236	248	265
Comparison	220	233	245	262
Difference	3*	3*	3*	3*
	(578)	(483)	(542)	(486)
CMP Group I	217	235	245	257
Comparison	209	229	241	255
Difference	8*	6*	4*	2[a]
	(296)	(185)	(149)	(159)
CMP Group II	215	235	250	263
Comparison	211	230	244	259
Difference	4*	5*	6*	4*
	(1449)	(1615)	(1302)	(1344)

*$p < .001$
[a] $p > .05$

School-Level Analyses

Two further comparisons were carried out using the school as the unit of analysis. The analysis of school achievement gains over time required the use of baseline MEAP scores. Since the matrix-sampled MEAP provided no individual student results, analysis of gains over time was carried out at the school level. Similarly, because calculation of subscores according to question type or mathematical strand depends on matrix-sampled items on MCAS (and therefore subscores are not provided as part of individual student scores), differences among these scores were calculated at the school level. In all school-level analyses, school results were weighted by the number of regular education students who took the mathematics MCAS in 1999. It should be noted that, unlike the student-level analyses reported above, the school-level results include students who had been in the district for fewer than three years because results for these students are not disaggregated in school-level data.

Comparison of Gain Scores

In order to better understand the course of change over time in average school performance after adoption of a standards-based program, we compared the overall gains or losses in mathematics scores among target curriculum and comparison schools from the time the new curricula were first introduced until 1999. Because the MEAP tests, in use until 1996, and MCAS, first administered in 1998, are not directly comparable and have different scales, school scores for the entire state were normalized for each year of interest to a mean school score of zero with a standard deviation of one. The resulting normalized score for each school type (target curriculum or comparison), or z-score, represents the number of standard deviations by which the target curriculum or comparison school's mean mathematics score surpassed or trailed behind the statewide mean school score for that year. The change in z-score as shown in Table 5 represents the *difference* between target curriculum and comparison schools in their relative improvement compared to the rest of the state over the period of interest. Positive values indicate a difference that favors the target curriculum group.

Examination of the data in Table 5 reveals that schools that began using *Everyday Mathematics* by 1994 had a mean test score gain at the fourth grade relative to the comparison schools of 0.57 standard deviations from 1992–1999. These early implementers (*EM* Group I) experienced small gains (.19) in the first few years, followed by moderate gains (.37) between 1996 and 1999. Later implementers of the program (*EM* Group II) made moderate gains (.31) in the first two to three years of implementation. These findings suggest a progressive achievement gain for *Everyday Mathematics*, that is, a positive longitudinal effect of the program on achievement. This is consistent with findings by Carroll (1997) in comparing statewide test scores of third-grade students using *Everyday Mathematics* since kindergarten to those students who had been using the program since first or second grade only (see also Ben-Chaim, 1997 for similar longitudinal effect with

Connected Mathematics). Another possible interpretation of the positive longitudinal effect is that teachers' effectiveness in teaching the programs increases with experience.

Table 5
Gain Score Differences Between Target Curriuclum and Comparison Schools Expressed in Standard Deviation Units

EM	Group I	Group II
1992–1996	.19	—
1996–1999	.37	.31
Total gain since implementation	.57	.31
CMP	Group I	Group II
1994–1996	.53	—
1996–1999	–.01	.39
Total gain since implementation	.52	.39

Note. Early *Everyday Mathematics* implementers (Group I) began to use the curriculum in 1994 and later implementers (Group II) began to use the curriculum after 1996. The early *Connected Mathematics* implementer (Group I) started to use the curriculum in 1995 and late implementers (Group II) after 1996. This table presents the difference in gain from the baseline year school mean score between the target curriculum and comparison schools, measured in standard deviation units. Scores were normalized to compare test performance between the MEAP and MCAS.

— Not applicable. Group II schools were not implementing the standards-based programs until 1996.

A similar gain of 0.52 is shown for eighth graders in the school using *Connected Mathematics* for four years between 1994 and 1999. For *Connected Mathematics*, the sole early implementing school made substantial early gains and essentially maintained them between 1996 and 1999, whereas later implementers (*CMP* Group II) made more moderate gains in the first three years of using the curriculum.[5]

COMPARISON OF RESULTS ON DIFFERENT TYPES OF QUESTIONS AND MATHEMATICS STRANDS

In the last analysis, we compared subscores on different mathematical strands and question types between the target curriculum and comparison schools, using aggregate school results from the 1999 MCAS test. The first strand, Number Sense, includes concepts of whole number operations, fractions and decimals, estimation, whole number computation and numeration in fourth grade; number relationships, number theory and systems, computation and estimation, ratio, propor-

[5] The reader may notice that these relative gains reported in standard deviation units appear larger than the effect sizes for individual score differences shown in Table 2. Gain scores are calculated relative to the standard deviation in school mean scores statewide rather than the standard deviation of individual student scores in only the schools included in this study. Because mean school scores vary less than do student scores, the school score standard deviation is smaller, which gives the appearance of a larger difference.

tion, and percent in eighth grade. The second strand, Patterns, Relations and Functions, includes algebra and mathematical structure in the fourth grade and algebra in eighth grade. The third strand, Geometry and Measurement, comprises many concepts including spatial sense and measurement at fourth grade and geometric measurement in eight grade. The fourth strand, Statistics and Probability, covers both areas in both grade levels.

The three types of MCAS questions are multiple choice, short answer, and open response. All four mathematical strands are represented in each question type. Sample released items from the 1999 MCAS at Grades 4 and 8 are shown in Appendix 2.

In this analysis we examined the percentage of possible points attained for each mathematical strand or question type, for target curriculum and comparison groups. Because the magnitude of differences in each category was consistently small (less than 5%) even when statistically significant, we represent these differences in Table 6 merely as a (+) when schools using the standards-based curriculum performed significantly better, (–) when comparison schools performed significantly better, and (0) when there was no significant difference.

Table 6
Significant Differences Between the Target Curriculum and Comparison Schools Across Mathematical Topics and Question Types on 1999 MCAS

	EM Group I (n = 2898) vs. Comparison (n = 3054)	*EM* Group II (n = 753) vs. Comparison (n = 1891)	*CMP* Group I (n = 73) vs. Comparison (n = 609)	*CMP* Group II (n = 1829) vs. Comparison (n = 4978)
Number Sense	+**	+**	+*	+**
Patterns & Functions	+**	+**	+*	+**
Geometry	+**	0	+**	+**
Statistics	0	+**	+*	+**
Multiple Choice	+**	+**	+**	+**
Short Answer	+**	+**	–	+*
Open Response	+**	+**	+*	+**

Note. This table is derived from school-level data, weighted by the number of regular education students who took the test in each school, and represents differences in the percentage of possible points received in different mathematical topics and question types. A (+) indicates that students in the reform curriculum group received a significantly higher percentage of possible points than students in the comparison group. A zero (0) indicates no significant differences and a minus (–) indicates that students in the comparison group received more points.

* $p < .05$

** $p < .001$

The results in Table 6 indicate that *Everyday Mathematics* schools outperformed comparison schools in all question types and all reporting categories, except that there was no difference in statistics for early implementers and in geometry for later implementers. *Connected Mathematics* students outperformed comparison students in all reporting categories and all question types except for short answer, where comparison schools performed better than did the single early implementing *Connected Mathematics* school. These results suggest that the differential gains made by target curriculum schools did not rest on markedly better performance on any single kind of mathematical task, but rather on small improvements in almost all areas.

Furthermore, these results suggest that students using either of the programs are still capable of performing procedural arithmetic items that are tested on the MCAS and doing so in a traditional, multiple choice format while also demonstrating an ability to solve higher order mathematics problems and to generate a response, rather than just recognize one. This confirms the research of Carroll (1997) whose work suggests that students can learn more advanced mathematics at earlier grades but not at the expense of traditional skills.

DISCUSSION

This study demonstrates that fourth-grade students in Massachusetts schools implementing *Everyday Mathematics* as their elementary mathematics program or eighth- grade students in schools using *Connected Mathematics* as their middle school mathematics program significantly outperformed matched peers from schools using a mix of traditional programs and curricula. The positive impact of the standards-based programs on student performance was remarkably consistent across students of different gender, race, and economic status. Students at the top, bottom, and middle of their classes all did better with the standards-based programs than did their counterparts using traditional programs. With minor exceptions, students in the target curriculum groups performed better in all four areas of mathematics and on all three types of test questions. For schools that had adopted these programs at least four years ago, early gains were sustained or increased further over time. Gains seen could not be attributed to differences in teacher qualifications nor differences in self-reported teacher instructional practice.

It is important to note that the study reported here includes *all* schools within the state that met criteria for implementing the program of interest as the primary curriculum within the school. Results attest to the effect of these curriculum programs as actually implemented under ordinary prevailing conditions in unselected schools, without regard to whether the programs were implemented optimally. It is possible that in some of these schools the programs are not well or fully implemented, or that programs may be supplemented with additional materials. Indeed, it is likely. Teachers seldom adhere perfectly to an external curriculum, but rather adapt it to their own classrooms. Use of supplementary materials and inconsistency between classrooms is also likely to occur in comparison schools using traditional curriculum. All of these considerations could be expected to blunt performance differences seen between target curriculum and comparison schools. Nevertheless, it seems clear that school use of *Everyday*

Mathematics as the core mathematics program at the elementary level or *Connected Mathematics* as the core mathematics program at the middle school level is associated with significant improvement in both school and student performance.

LIMITATIONS

One limitation of this study is that it addressed a population of target curriculum and comparison schools that are relatively advantaged: they are schools that have a small percentage of students eligible for free and reduced price lunch and that are predominantly White, at least when compared to schools in the rest of the state. However, when student performance was examined for different student populations, non-White and low-income students in the target curriculum schools clearly outperformed their counterparts in the comparison schools. Nevertheless, further study should be done on the impact of these standards-based programs in schools with a larger representation of low income and minority students.

This study did not provide detailed information about teacher instructional practices, nor did it include classroom observation. Limited information about teacher practices was obtained from a statewide teacher questionnaire, and this information revealed little difference in practice between the target curriculum and comparison teacher groups. But this information was self-reported. Teachers using different kinds of curriculum may have interpreted the questions differently or their responses to questions may have reflected their conception of ideal rather than actual practice. Observational studies may be required to better explicate the complementary contributions of reform-oriented curriculum materials and instructional practice to student achievement.

A further objection may be raised that implementing any new program predictably leads to short-term gains (i.e., the Hawthorne effect). But the results presented here show improvements persisting for *Connected Mathematics* or increasing further for *Everyday Mathematics* as the program moves beyond its third year of use.

It might be argued that the improvements seen reflect not actual improvements in student performance but merely a close match between test and curriculum. And indeed, one definition of a standards-based program is that it is a program that aligns teaching and learning to enable students to meet performance standards. Both *Connected Mathematics* and *Everyday Mathematics* were written to meet the *NCTM Curriculum and Evaluation Standards* (1989). The 1995 Massachusetts mathematics curriculum framework was based on the NCTM *Standards*, and the MCAS test is based directly on the state curriculum framework. But the MCAS was not written specifically to highlight the strengths of either *Everyday Mathematics* or *Connected Mathematics*. Moreover, by 1996, *Everyday Mathematics* students already showed greater gains than the comparison group on the old MEAP test, which preceded state frameworks and was not particularly aligned to the NCTM *Standards*. Furthermore, the consistent positive impact of the standards-based programs on performance of students across different areas of mathematics and different types of test questions argues against the differential improvement seen being an artifact of some imbalance in the test.

SUGGESTIONS FOR FURTHER STUDY

This article raises several questions worthy of further study. It would be particularly helpful to examine the progress of individual students' mathematics performance over time. In 2001, a new sixth-grade mathematics MCAS test was administered, and it will become feasible to look at students' growth over every two-year period from Grades 4 to 10. Continued study is also warranted to determine whether the kinds of improvement reported here persist or increase as teacher experience with the curriculum continues to grow. Further follow-up would be particularly interesting in view of the recent revision of the Massachusetts mathematics curriculum framework (MA DOE, 2000). The revision reflects some shift toward more emphasis on computational facility; it is likely that changes in test emphasis will follow. It will be important to determine whether improvements in student performance persist (as they did for *Everyday Mathematics* during the shift from MEAP to MCAS) in a once-again shifting assessment environment.

Another study of interest could examine the impact of these and similar standards-based programs on special education populations. Such a study would need to determine carefully which special education students have full access to the curriculum, and compare them to students with a similar level of disability using traditional or remedial curriculum.

Also of interest would be an examination of the effect of an elementary-middle sequence of standards-based curriculum programs in the same school systems over time. The purpose of the study would be to determine whether consistent exposure to such programs over the elementary and middle school years leads to compounded improvement or perhaps instead reveals weaknesses and gaps in the curriculum that will lead to a leveling-off or loss of the improvement seen.

Much work remains to be done in closer examination and qualitative analysis of implementation in individual schools. Such analysis could provide information about what determines successful implementation of standards-based curriculum programs. Further study should elucidate how to implement these programs for maximum effect.

CONCLUSION

The results of the study reported in this article add to the accumulating body of evidence that standards-based mathematics programs have a positive impact on student achievement. Fourth-grade students using *Everyday Mathematics* and eighth-grade students using *Connected Mathematics* outperformed matched comparison groups who were using a range of textbooks commonly used in Massachusetts. The gain in student performance was greater in schools that were farther along in their implementation of the standards-based programs. These performance gains, which were moderate in size, remained consistent for different groups of students, across mathematical topics and different types of questions on the state test. This study supports the notion held by proponents of standards-based curriculum, that curriculum itself can make a significant contribution to improving student learning.

REFERENCES

American Association for the Advancement of Science (2000). *Middle Grades Mathematics Textbooks: A Benchmarks-Based Evaluation.* [On-line]. Available: http://www.project2061.org.

Askey, R. (1999, October). *Good intentions are not enough*. Paper presented at the Conference on Curriculum Wars: Alternative Approaches to Reading and Mathematics. Harvard University, Cambridge, MA.

Baker, D. P., & Schaub, M. (1991). Solving the math problem: Exploring mathematics achievement in Japanese and American middle grades. *American Journal of Education, 99,* 623–642.

Bay, J. M., Reys, B. J., & Reys, R. E. (1999). The top 10 elements that must be in place to implement standards-based mathematics curricula. *Phi Delta Kappan*, *80*, 503–506.

Bell, M. (1988–1996). *Everyday Mathematics*. Chicago, IL: Everyday Learning Corporation.

Ben-Chaim, D., Fey, J., Fitzgerald. W., Benedetto, C., & Miller, J. (1997, April). A study of proportional reasoning among seventh and eighth grade students. Paper presented at the meeting of American Education Research Association, Chicago, IL.

Boaler, J. (1998). Open and closed mathematics: Student experiences and understanding. *Journal for Research in Mathematical Education*, *29*, 41–62.

Briars, D. J., & Resnick, L. B (2000). *Standards, assessments—and what else? The essential elements of standards-based school improvement*. Manuscript submitted for publication.

Burrill, G. (1997). The NCTM *Standards*: Eight years later. *School Science and Mathematics*, *97*, 335–339.

Carroll, W. M. (1995). *Third grade* Everyday Mathematics *students' performance on the 1993 and 1994 Illinois state test.* Chicago: University of Chicago School Mathematics Project.

Carroll, W. M. (1996). Mental computation of students in a reform-based mathematics curriculum. *School Science and Mathematics*, *96*, 305–11.

Carroll, W. M. (1997). Results of third grade students in a reform curriculum on the Illinois state mathematics test. *Journal for Research in Mathematics Education*, *28*, 237–242.

Carroll, W. M., & Porter, D. (1994). *A field test of fourth grade* Everyday Mathematics. Chicago: University of Chicago School Mathematics Project.

Clopton, P., McKeown, E., McKeown, M., & Clopton, J. (1999). *Mathematically Correct mathematics program reviews for grades 2, 5, and 7* [On-line]. Available: www.mathematicallycorrect.com /books.htm

Courtina, J. M., & Nouri, H. (2000). *Effect sizes for ANOVA designs*. (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-129). Thousand Oaks, CA: Sage.

Goldsmith, L. T., Mark, J., & Kantrov, I. (1998). *Choosing a standards-based mathematics curriculum*. Newton, MA: Education Development Center.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*, 107–128.

Hiebert, J. (1999). Relationships between research and the NCTM *Standards. Journal for Research in Mathematics Education*, *30*, 3–19.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.) (2001). *Adding it up: Helping children learn mathematics*. Washington, D.C.: National Academy Press.

Klein, S., Hamilton, L., McCaffrey, D., Stecher, B., Robyn, A., & Burroughs, D. (2000). *Teaching practices and student achievement: Report of first-year findings from the 'Mosaic' study of systemic initiatives in mathematics and science*. Santa Monica, CA: RAND.

Lapan, R. T., Reys, B. J., Barnes, D. E., & Reys, R. E. (1998, April). *Standards-based middle grade mathematics curricula: Impact on student achievement*. Paper presented at the meeting of American Education Research Association, San Diego, CA.

Lappan, G., Fey, J. T., Fitzgerald, W. M., Freil, S. N., and Phillips, E. D. (1991–1997). *Connected Mathematics.* White Plains, NY: Dale Seymour Publications.

Lofton, L. B., & Madison, S. Q. (1991). The extreme dangers of covariance corrections. In B. Thompson (Ed.), *Advances in educational research: Substantive findings, methodological developments* (Vol. 1, pp. 133–147). Greenwich, CT: JAI Press.

Loveless, T. (2000). *The Brown Center report on American education 2000: How well are American students learning?* [On-line]. Available: http://brookings.org/gs/brown/bc_report/2000/intro.htm

Massachusetts Department of Education. (1996*). Mathematics curriculum framework: Achieving mathematical power*. [On-line]. Available: www.doe.mass.edu/frameworks/archive/math96/CFMath.html

Massachusetts Department of Education (1999). *MCAS 1998 technical report*. [On-line]. Available: www.doe.mass.edu/mcas/archives/99/98tech_report_full.pdf

Massachusetts Department of Education (2000). *Massachusetts mathematics curriculum framework*. [On-line]. Available: http://www.doe.mass.edu/frameworks

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

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

National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics*. Reston, VA: NCTM.

National Council of Teachers of Mathematics. (2000). *Principals and standards for school mathematics*. Reston, VA: NCTM.

O'Neal, S. W., & Robinson-Singer, C. (1998). *The Arkansas Statewide Systemic Initiative pilot of the Connected Mathematics: An evaluation report*. [On-line]. Available: http://www.math.msu .edu/cmp/ArkRep.html

An open letter to United States Secretary of Education, Richard Riley. (1999, November 18). *The Washington Post*, p. A5.

Richardson, V. (1990). Significant and worthwhile change in teaching practice. *Educational Researcher*, *19*, 10-18.

Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.

Schoenfeld, A. H. (1985). *Mathematical problem-solving*. Orlando, FL: Academic Press.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem-solving, meta-cognition, and sense making in mathematics. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334–371). New York: Macmillan.

Steen, L. A. (1990). Making math matter. *Instructor*, *99*(8) 21–23.

Stein, M. K., Lane, S., & Silver, E. A. (1996, April). *Classrooms in which students successfully acquire mathematical proficiency: What are the critical features of teachers' instructional practice?* Paper presented at the annual meeting of the American Educational Research Association, New York.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: an analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, *32*, 455–488.

Stigler, J. W., & Hiebert, J. (1997). Understanding and improving classroom mathematics instruction: An overview of the TIMSS video study. *Phi Delta Kappan*, *79*(1), 14–21.

U.S. Department of Education. (1999). *Mathematics and Science Expert Panel of Exemplary and Promising Programs*. [On-line]. Available: http://www.enc.org

Winking, D., Bartel, A., & Ford, B. (1998). *The Connected Mathematics Project - helping Minneapolis middle schools students "beat the odds": Year one evaluation report.* East Lansing, MI: Unpublished technical report.

Wood, T., & Sellers, P. (1997). Deepening the analysis: A longitudinal assessment of a problem-centered mathematics program. *Journal for Research in Mathematics Education*, *28*, 163–186.

Wu, H. (1999). Basic skills versus conceptual understanding. *American Educator/American Federation of Teachers*, *23*(3), 14–22.

Authors

Julie E. Riordan, The Noyce Foundation, 201 Newton Street, Weston, MA 02493; riordanjulie @hotmail.com.

Pendred E. Noyce, The Noyce Foundation, 201 Newton Street, Weston, MA 02493; pnoyce @mediaone.net

APPENDIX 1

Teacher Questionnaire

During the 1999 administration of MCAS, a questionnaire was administered to each fourth-, eighth-, and tenth-grade teacher of mathematics and science. The questionnaires included several items (under two main question stems) relating to instructional practices or the frequency in which students are engaged in various mathematics and science activities as part of their instruction. These questions were derived from a national RAND study in which Massachusetts participated.

To obtain a sample of instructional practices in different courses and for students of different ability levels, respondents were asked to refer to their first regularly scheduled mathematics or science class of the week when answering the questions. The respondents answered on a scale of 0–4 with response choices ranging from "never" using the practice to using it in "all or almost all mathematics or science lessons." A reform and a traditional score was computed for each teacher by summing individual teacher responses to reform or traditional items. Teacher pedagogy scores were aggregated to the school level for analysis. Across all grade levels statewide, a school's reform pedagogy score was positively correlated with MCAS score; that is, schools whose teachers reported more reform-oriented pedagogy tended to have higher student scores.

Questionnaire Items

For the purposes of the study reported in this article, items were identified as "reform-oriented" or "traditional practice." We followed RAND's practice in designating items as either reform-oriented or traditional. They were not so identified on the questionnaire.

Question: "… About how often do you typically do the following in your mathematics instruction in this class?"

Reform-Oriented

Use open-response questions.
Require students to explain their reasoning when giving answers.
Provide an opportunity for students to discuss mathematics with one another.
Provide an opportunity for students to explore alternative methods for solutions.

Read and comment on the reflections written in student notebooks or journals.
Arrange seating to facilitate student discussions.

Traditional Practice
Lecture/introduce content through formal presentations.

Question: "…About how often do students in this class typically take part in each of the following types of mathematics activities as part of their math instruction?"

Reform
Use calculators as a tool in problem solving.
Collect, analyze, and organize data including graphs and charts.
Take tests requiring open-ended responses (e.g., descriptions, justifications of solutions).
Share ideas or solve problems with each other in small groups.
Write a description of a plan, procedure, or problem-solving process.
Use rubrics to evaluate student work.

Traditional Practice
Read from a mathematics textbook in class.
Practice computational skills (e.g., worksheets)
Take tests using short-answer questions (e.g., multiple choice, true/false, fill-in-blank)
Memorize mathematics facts, rules, or formulas.
Observe a teacher demonstrate how to do a procedure or solve a problem.
Read and comment on the reflections written in student notebooks or journals.

APPENDIX 2

Sample MCAS Questions for Grades 4 and 8.

Grade 4

Multiple Choice (Statistics and probability)

What is the GREATEST number of different outfits you can make with 2 pairs of pants and 5 shirts? (Each outfit must have exactly one pair of pants and one shirt.)

a. 5
b. 7
c. 10
d. 25

Short Answer (Number sense)

Compute:

$$\begin{array}{r} 3{,}972 \\ 44{,}826 \\ +8{,}321 \\ \hline \end{array}$$

Open Response (Patterns, relations and functions)

These are input-output tables. Each table has a different rule. When a number n is put in, it is changed by the rule so that a different number comes out. Table 1 has been completed for you.

Complete Tables 2 and 3 in your Student Answer Booklet

Table 1

Input	n	8	1	5	9	21
Output	$n + 5$	13	6	10	14	26

Input-Output Rule: $n + 5$

Table 2

Input	n	2		9	7
Output	$n \times 9$	18	54		63

Input-Output Rule: $n \times 9$

Table 3

Input	n	36	16	8	
Output		9		2	7

Input-Output Rule: ___________

Write an input-output rule for Table 3 using the letter *n*.

Use a new rule to make up your own input-output table. Complete Your Table in your Student Answer Booklet. Be sure to include your rule using the letter *n*. (You may NOT use the rules from Tables 1, 2, or 3.)

Your Table

Input	n
Output	

Input-Output Rule: ___________

Grade 8

Multiple Choice (Number sense)

The regular price of a computer game is $49.95. Which of the following sale prices would result in the lowest price for the computer game?

a. $8 less than the regular price
b. 20% discount on the regular price
c. $^{1}/_{4}$ reduction on the regular price
d. 85% of the regular price

Short Answer (Patterns, relations, and functions)

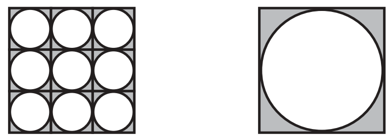
What does x equal in this equation?

$^{x}/4 + 8 = 32$

Open Response (Geometry and measurement)

Use the figures below to answer question 38. Both Figures *A* and *B* measure 6 inches by 6 inches.

Figure *A* Figure *B*



Micah thinks that the shaded part of Figure *A* has a greater area than the shaded part of Figure *B*. Tonya thinks that the shaded part of Figure *B* has a greater area than the shaded part of Figure *A*.

Write a note to Micah and Tonya telling whether either of them is right. Explain in detail the calculations you made to compare the areas of the shaded parts of Figures *A* and *B*.