

SCHOOL-RELATED INFLUENCES ON GRADE 8 MATHEMATICS PERFORMANCE IN MASSACHUSETTS

Sandra Stotsky
Northeastern University

Rafael Bradley and Eugene Warren
Thomas, Warren + Associates

ABSTRACT

Less than one third of American eighth graders score in the two highest performance levels on the grade 8 mathematics test given by the National Assessment of Educational Progress. Only a little over one third of Massachusetts eighth graders score at the two highest performance levels on the state's own grade 8 mathematics test. In 2002, the Massachusetts Department of Education funded research to explore why there had been no significant growth in the percent of grade 8 students performing at the two highest levels on the state's grade 8 mathematics tests. An analysis of quantitative data obtained from administrators and teachers in a representative sample of 60 schools throughout the state in 2003 identified school-based factors that were significantly associated with the 20 of the 60 schools that both *increased above the state average increase* the percent of grade 8 students performing at the two highest performance levels on the state's grade 8 mathematics test and simultaneously *decreased above the state average decrease* the percent of grade 8 students performing at the lowest performance level. A significantly higher percent of teachers in these 20 schools reported spending a great deal of time reviewing and using test results, having a voice in the choice of their instructional materials, using accelerated and leveled algebra I classes to address the needs of above grade students, and less frequent use of calculators in non-algebra classes. At a time when teachers in all states are being held accountable for increasing the achievement of all their students, these findings warrant exploration on a nationwide scale.

I. INTRODUCTION

The Massachusetts Education Reform Act of 1993 (Chapter 71 of the Acts of 1993, Statutes of the Commonwealth of Massachusetts) changed almost every aspect of elementary and secondary education in the state in order to improve student learning in all subjects. With the support of industry leaders, teacher unions, and the public at large, the Massachusetts legislature mandated the development of a comprehensive and far-reaching system of standards and accountability measures that would affect all students, teachers, and school districts. For students, this system took the form of pre-kindergarten to grade 12 standards (called curriculum frameworks) and accountability measures (annual state tests that are part of the Massachusetts Comprehensive Assessment System, or MCAS). For teachers, this system took the form of five-year cycles for license renewal and the requirement of individual professional development plans approved by the teacher's principal or supervisor. For school districts, this system took the form of school and district standards with accountability measures applied through an established schedule of inspections, and ratings based on the inspections and student test results.

Over the past ten years, Massachusetts has dedicated significant resources to improving the academic performance of all its students, its lowest achieving students in particular—those whose performance on the MCAS tests is at the Warning level. One major effort to address this goal by the Massachusetts Department of Education (Department) was the Middle School Mathematics Initiative (MSMI), a two-year intervention and research project begun in 2000 to help mathematics teachers in under-performing middle schools, as identified by MCAS scores, to improve student achievement in mathematics. We provide a short description of the 2000-2002 study because this study and its results served as the point of departure for the study reported here.

A. Methodology and Results of the MSMI

For the MSMI, the Department employed six highly experienced mathematics teachers as mathematics specialists, or coaches, and a highly recommended pedagogical strategy for strengthening teachers' effectiveness in their classrooms. The Department was especially interested in assessing the value of coaching in improving student learning in mathematics because it is an expensive strategy for school systems to use, with no body of scientifically based research evidence yet available to attest to its efficacy (Russo, 2004). The basic task of the six specialists was to train over 50 teachers in grades 6, 7, and 8 in eight school districts in lesson planning and implementation over the course of more than one year (24 teachers in the first year of the study continued into the second year of the study). The emphasis was on lesson planning and implementation because the principles guiding them are generic and can be applied to any mathematics curriculum (Panasuk & Sullivan, 1998).

All students in the intervention and comparison classes (volunteered by their principals and teachers, with over 1000 students in each group each year) were given pre-post tests consisting of items similar to released MCAS grade 6 mathematics items addressing basic arithmetic operations. The Department sought to determine learning gains during the academic year and to pinpoint students' achievement level in arithmetical skills and understanding more precisely than can be learned from MCAS tests, which have been given only at the end of two-year grade spans. Because most low-performing schools today receive targeted assistance of varying kinds (whether for the whole school and the regular classroom teacher or for the low-performing, Limited English Proficient, or English as a Second Language student through a Title I or bilingual education teacher), the intervention and comparison groups as a whole in this initiative could be considered matched mixed models; the only clear difference between them was the Department's own carefully defined model of coaching.

As part of the first year of the project, 15 teachers voluntarily took a middle school mathematics course at the University of Massachusetts-Lowell. As part of the second year, 36 more teachers took a Department-sponsored middle school mathematics course taught in four locations by

three mathematics professors using both a common syllabus and a pre-post test that they had developed. Mathematical knowledge only was taught in these courses to help the Department explore the relationship between teacher knowledge in mathematics and gains in student learning.

This project found that students in the MSMI classrooms had change scores that were significantly higher than similar students in classrooms with no intervention, even though a much higher percentage of students identified as LEP were in the MSMI classrooms. Additionally, teachers' lesson planning ability was related to change scores, that is, students of teachers with higher score planning made significantly more improvement than students of teachers with lower lesson planning scores. The study also found that students of teachers with more teaching experience achieved higher gains than students of teachers with less teaching experience (University of Massachusetts Donahue Institute, 2002).

Although the differences in outcomes between the two groups were statistically significant, the practical significance of these differences was questionable. The grades 6, 7, and 8 students in the intervention classes could achieve a maximum of 20 points on a test of basic arithmetical operations that included word problems all pitched to a grade 6 level. On average, the students got about 9 points at the beginning of the year and about 12 points at the end of the year. This is a modest gain, even if the differences between the two groups were statistically significant, thus providing only modest support for the efficacy of mathematics coaches, as defined in this project, in improving mathematical learning in low-achieving students. Although the participating teachers in this project all found their work with the mathematics specialists beneficial to their teaching, these benefits did not translate directly into meaningful increases in mathematics achievement for the low-performing students in their classrooms.

Nor did the benefits of the coursework in mathematics translate into increased student achievement. Students whose teachers took the mathematics course in the second year of the study, showed gains on the teacher pre-post test, and found the coursework beneficial showed no greater gains overall than other students.

During the course of the study and in discussions of its results with specialists and teachers at a Title I conference in 2002, several factors affecting the learning of all low-performing students, whether or not in the intervention group, were identified by the field as needing further exploration. One factor was student reading level; the students in both the intervention and comparison classes in the MSMI study were below average in reading as well as in mathematics.

A second factor was the use of grade level textbooks in a standards-based environment. In standards-oriented schools, it is understandable why administrators purchase grade level textbooks for the middle school; the grade 8 MCAS mathematics test is based on grade 8 standards and if they are to prepare students for the grade 8 MCAS they feel obligated to address the standards on which

the grade 8 test is based. However, unlike the widespread availability of developmentally appropriate below grade-level reading materials (often called high interest/low vocabulary), there seem to be few if any below grade-level mathematics materials available to teach skills that students have not yet acquired but which are needed for problem solving in the grade-level textbooks.

A third factor was student grouping. In the relatively large body of research on the effects on achievement of grouping students with varying skill levels in different ways, the evidence suggests that students learn more mathematics when they are in more homogeneous groups with a curriculum and materials geared to their needs (Loveless, 1999; Loveless, 2000; Slavin, 1987; Slavin, 1990). In classes with a wide range of student achievement, it is not clear how well classroom teachers address the specific weaknesses of low-performing students, especially if they are using grade-level materials.

The fourth factor mentioned was student absenteeism, a factor that directly affects student learning. Student absentee rates for 2001-2002 were not available at the time the final report for the MSMI was completed by its external evaluators (University of Massachusetts Donahue Institute, 2002), but they were available for the first year of the project. In grade 8 for the first year of the project, in the MSMI schools, 598 out of 2,654 students (23%) were absent 11 to 20 days for the year, while 20% (an additional 525 students) were absent more than 20 days. Absentee rates in the comparison schools were slightly higher. While attendance rates may be lower for the lowest-performing students than for the school as a whole, it was not possible to obtain attendance data for individual students or specific groups of students. We could only assume that the rates were similar across both groups of schools.

B. Immediate Background for the Present Study

The Department learned from the MSMI that there was much more to explore than it had initially thought in order to determine how to spend public appropriations wisely for increasing middle school mathematics achievement. In addition, by 2002 the Department had become as concerned about higher achieving students in the state as about lower performing students. As Table 1 shows, Massachusetts students in grade 8 mathematics classes already at the Needs Improvement or Proficient level (the second and third highest performance levels on the state's tests) were not moving quickly as a group to the Proficient or Advanced level, or even as quickly as grade 8 mathematics students as a group were moving from the lowest level to the Needs Improvement level.

In 1998, 31% of grade 8 students scored at the two highest performance levels, and in 2002, 34% did, an increase of only 3% of the total number of students. On the other hand, the percent scoring at the Warning level decreased from 42% in 1998 to 33% in 2002, a decrease of 9%. The

concern here was equity. Why weren't grade 8 students moving into the two highest performance levels at least at the same rate as students moving out of the lowest level? Were schools in Massachusetts expending less educational effort on the top 60% to 70% of their students in grade 8 than on the bottom 30% to 40% because of the sanctions attending continuous low school performance, thus turning state tests into de facto minimum competency tests? The Department decided to find out what school-based factors might differentiate schools that had increased the percent scoring at the two highest levels as much as they had decreased the percent scoring at the lowest level from schools that had decreased the percent at the lowest level more than they had increased the percent scoring at the highest levels (if in fact they had increased the percent at the two highest levels at all).

Table 1. Grade 8 MCAS Results in Mathematics from 1998 to 2003: Percentage of Students at Each Performance Level

	<i>WARNING</i>	<i>NEEDS IMPROVEMENT</i>	<i>PROFICIENT</i>	<i>ADVANCED</i>
1998	42	26	23	8
1999	40	31	22	6
2000	39	27	24	10
2001	31	34	23	11
2002	33	33	23	11
2003	33	30	25	12
2004	29	32	26	13

Source: Massachusetts Department of Education

The research question was: What school-based factors might be related to the lack of significant growth in the percent of students in grade 8 performing at the two highest levels since the inception of state tests in 1998? To explore this question, the Department chose a research design that might be more informative and much less expensive than the one used in the MSMI (see Carnine & Gersten, 2000, for a discussion of the debates about the types of research that might best inform policy and practice). Using funds from its National Science Foundation-supported State Systemic Initiative, the Department retained Thomas, Warren + Associates to gather quantitative data from a stratified random sampling of schools across the state, focusing just on grade 8—a pivotal grade in mathematics education in K-12—and to explore, among other probable influences on student achievement, the second and third factors described above (grade level and choice of textbooks, and grouping arrangements) because specialists and teachers had stressed their relevance in discussions with Department staff. The Department sought a stratified random sampling of schools across the state in order to avoid the complex problems inherent in matching large numbers of schools to produce valid comparison groups, such as the problems encountered by Riordan &

Noyce (2001) in a study comparing mathematics achievement in grades 4 and 8 in selected Massachusetts schools. The Department also sought a stratified random sampling of schools across the state in order to allow identification of the schools selected for the study: this would enable other researchers to confirm or further explore its results (see www.csun.edu/~vcmt00m/noyce.htm for an exchange of communications on this topic).

II. THE PRESENT STUDY¹

The present study was designed to be exploratory in nature. Its purpose was to identify school-based factors that were significantly associated with schools that had both increased *above the state average increase* the percent of grade 8 students performing at the two highest levels on the state's grade 8 mathematics test and at the same time had decreased *above the state average decrease* the percent of grade 8 students testing at the lowest level on the state's mathematics tests between school year 1998-99 and school year 2001-02 (henceforth to be referred to as the study period). The contractors were to examine and compare curricula; instructional and grouping practices; extra support (*e.g.*, tutoring, parental assistance); teacher qualifications; textbook use; and instructional organization (*e.g.*, block scheduling, team-teaching) across the state's schools. As Stigler and Hiebert suggest in *The Teaching Gap* (1999), it may not be the teachers' instructional choices that are retarding student achievement in this country but a "system" that tells them what they should or should not do in their classrooms.

The Thomas, Warren + Associates research design was developed in three parts. First, a methodological approach for analysis was identified. Second, a sampling strategy was prepared (Lohr, 1999). Finally, two survey instruments were written and administered to collect the school-specific information required for the study. These instruments consisted of questions to be asked of a representative sample of grade 8 administrators and mathematics teachers and were based on the suggestions of Department staff (reflecting their communications with the field) and the content of existing questionnaires (Massachusetts Education Reform Review Commission, 2000a; 2000b; 2001). A detailed account of the methodology used and copies of the survey instruments are available in the final report that Thomas, Warren + Associates submitted to the Department in June 2003.²

A. Research Methodology

The sampling strategy for choosing schools for inclusion in the study required partitioning the universe of Massachusetts schools. First, schools were considered only if they administered the state's grade 8 mathematics test every year of the study period and administered it to a minimum of 50 students. All public (and public charter) schools are required to administer the state tests, with no

¹ The present study was conducted by Diane Bishop, Rafael Bradley, Alice Christie, Dalila Gomes, and Linda Warren under the overall supervision of Eugene Warren.

² A copy of this report is available at www.twaiconsulting.com.

exceptions. Altogether 308 schools in Massachusetts met these criteria as of 2001-02 (Massachusetts Department of Education, 1999a; 1999b; 1999c; 2001a; 2001b; 2001c; 2002). Next, in order to capture the effects of a school being part of a large or a small district, districts were classified according to their size. Districts with fewer than four schools giving the mathematics test in 2001-02 were classified as small districts. All other school districts were classified as large.

Inclusion in the sample was based on performance on the state's mathematics test. Schools were first partitioned into two groups based on whether their observed change was above or below the state average increase in the percent testing at the two highest levels. The state average change was calculated as the mean of the changes in all 308 schools in the sampling frame. A second partition was based on a greater or less than average decrease in the percent of a school's students at the lowest level on the state's mathematics test, creating four groups in all. The group of interest in the study represented schools that had both increased the percent of students testing at the two highest levels by more than the state average and simultaneously decreased the percent of students at the lowest level by more than the state average over the study period. These schools will be referred to as Improving Proficient, Advanced, and Warning (IPAW) schools. The study was based on the assumption that they were doing something better. All the schools in the other three groups were analyzed as a single group, hereafter referred to as non-IPAW schools. Table 2 provides a count of the schools in each of these groups and a description of the overall sample development.

Table 2. Development of the Sample Used in the Study

Schools	IPAW Schools		Non-IPAW Schools		Total Schools
	In Large Districts	In Small Districts	In Large Districts	In Small Districts	
In sampling frame	21	71	69	147	308
In sample	13	13	25	24	75
Eligible and agreed to participate	10	11	23	21	65
Administrators and Teachers					
Eligible and agreed to participate	35	36	75	67	213

Algorithmic random sampling was performed to select schools in small districts. Schools in large districts were selected for participation in the study by Thomas, Warren + Associates in a different way. Schools in large districts were selected based on a committee rating process rather than algorithmic sampling. It was agreed that algorithmic random sampling from a small population of large districts (90 schools) could potentially lead to a very biased sample and that there were no significant implications from using two different methods of sampling. The goal of the committee was to develop a sample that was representative of the population in terms of MCAS results but also exhibited the diversity of socioeconomic status found in the population (Massachusetts Department

of Employment and Training, 2002; Boston Plan for Excellence, 2002a; 2002b). The committee was composed of three senior staff from Thomas, Warren + Associates, two education specialists, and one statistician, all of whom were familiar with the Massachusetts school system.³ School selection in the large districts was made independently by the members of the committee. The kappa statistic for rater agreement among the members was 0.60 ($p=0.00$).⁴ Additionally, schools in large districts were over-sampled because of a concern that within district variability in test results and socioeconomic status needed to be adequately represented in the final sample. A preliminary list of 37 schools in small districts and 38 schools in large districts was selected.

Following notification of selection for participation in the study by Thomas, Warren + Associates, the 75 selected school principals were each contacted by telephone in order to obtain their agreement to participate in the study. Part of this agreement was that the principal, the school's mathematics coordinator or department chair (if there was one), and at least one teacher (or as many as two teachers) who had taught at the school and administered the state's grade 8 mathematics test in the 2001-02 school year would participate in the study. Teachers were selected for participation by their principals from the (usual) pool of two or three eligible grade 8 math teachers in their school. Sixty-five schools met all criteria (32 in small districts and 33 in large districts) and agreed to participate in the project.

Table 3 identifies similarities and differences between the IPAW and non-IPAW schools for the 60 schools from which complete survey data were collected and which constituted the final sample. Although the two groups of schools were similar in many important areas, two areas of difference warrant comment. The percentage of LEP students in the non-IPAW schools was almost twice that of LEP students in the IPAW schools (23% to 12%). Although in theory this could be an important difference between the two groups of schools, administrators and teachers in both the non-IPAW and IPAW schools rarely commented on second language problems, in focus groups or on an instrument designed to gather qualitative data (not reported here). It is also the case that the non-IPAW schools began with higher scores than the IPAW schools and thus might find it more difficult to raise achievement, especially since they enrolled more LEP students. Even if this did make a difference in their capacity to raise scores, what the IPAW schools did to increase the percentage of students in the two highest performance levels is still of interest, especially since the overall percentage of students in the state in these two levels is puzzlingly low in a state with an

³ To the extent that only a certain percent of schools from each district needed to be IPAW schools, the committee was influenced by knowing which were IPAW schools. However, the selection was made **before** results of the surveys were known. If the selection affected the results of the survey, it was not done intentionally.

⁴ Essentially, we measured the number of schools that were common to each selection committee member's list of schools to be included in the sample of large schools.

overall high level of parent education. It should be noted that the final sample size for Massachusetts is close to the final sample size of 77 schools participating in the study conducted by Hiebert and others (2003) to develop a general picture of mathematics instruction in the United States.

Table 3 Similarities and Differences in Schools

Demographics in 2001-2002	IPAW Schools (N=20)	Non-IPAW Schools (N=40)
Number of schools from large districts	10	20
Number of schools serving only grades 6-8	11	16
Number of magnet or special focus schools	3	7
Average school enrollment	728	866
Percent of students receiving free or reduced lunches	36%	36%
LEP students as a percent of enrollment	12%	23%
MCAS Performance		
Average percent of students at Proficient or Advanced (1998-99)	18%	34%
Average percent increase in Proficient or Advanced (from 1998-99 to 2001-02)	12.9%	0.5%
Average percent of students at Warning (1998-1999)	49%	38%
Average percent decrease in Warning (from 1998-99 to 2001-02) ⁵	-17.2%	-2.1%
Teachers in 2002		
Percent of teachers licensed to teach mathematics	65%	73%
Percent of teachers with over five years of experience	42%	30%
Average number of sections taught by mathematics teachers	3.7	3.6
Average number of students per section	21	21
Classroom Practices in 2002		
Percent of sections where homework was assigned	32%	49%
Percent of grade 8 students enrolled in algebra I	21%	39%

Note: In general, schools that used homogeneous groupings had various levels of mathematics courses such as algebra, pre-algebra, and general mathematics.

Primary and supplemental data collection instruments were developed for administrators (principals and math chairs) and teachers. The primary data collection instrument contained four types of questions: multiple-choice, open-ended, choose all that apply, and Likert scale ranking. The supplemental data collection instrument had multiple choice and open-ended questions. All questions applied to 2001-02. Additionally, to provide comparative data, some questions asked about 1998-99 but only from personnel who had been at the school since 1998-99.

Senior staff of Thomas, Warren + Associates visited the 65 schools in the original sample between January 5, 2003 and March 17, 2003 to collect data. In total, educators in 60 (30 in small districts and 30 in large districts) of the original 65 schools that had been selected for the study completed the survey. (The fact that 10 of the IPAW schools were in large districts and 10 were in

⁵ In the IPAW schools, the increase could reflect “regression to the mean” in the sense that, in this context, they were catching up to the performance of the average school in their socioeconomic strata.

small districts was not planned but simply due to chance.) These 60 schools are shown in the Appendix.

The survey data collection instruments were administered online at 53 schools and paper surveys were used at the remaining schools. The average participant return rate for surveys was 95.9%. Table 4 presents the counts of respondents to the survey.

Table 4. Overall Survey Response Statistics

	Participants	Surveys Completed	Percent Response
Principals	65	61	94%
Teachers	113	108	96%
Math Coordinators or Department Chairs	29	27	93%
TOTAL	207	196	95.9%

Although 108 teachers, 61 principals, and 27 math coordinators or department chairs completed the survey, one principal's response, one teacher's response, and two department chairs' responses had to be excluded because it was subsequently determined that they did not meet the eligibility criteria on the date the survey was administered. In total, the analysis used data from only 107 teachers, 60 principals, and 25 coordinators or department chairs. An analysis of the participant response rates indicated no significant differences in non-response across the sampling strata.

The analysis of the data collected from the surveys was undertaken in several parts, each using a different approach to identify factors affecting test performance of the IPAW schools. Categorical data from the surveys were analyzed using statistics to identify associations between the responses and test results in groups of schools. A contingency table analysis was performed for the two groups of schools. In each case, a Pearson test of independence between a specific response to a survey question (e.g., a school factor) and membership in the two groups of schools was performed. The test was performed for each response separately. All tests were therefore univariate tests of association; all tests incorporated appropriate sample weights.

Responses to questions were treated as separate school-based factors for the analysis. Each school had a single response from its principal and was treated as if it had a single response from its teachers. Rejection of the null hypothesis from the Pearson test was taken to indicate that a given factor (response) was associated with observed differences between IPAW and non-IPAW schools in test performance; in other words, the null hypothesis was that a factor was not related to the MCAS results. Odds ratios were used to identify the direction and strength of the association.

A. The State's Grade 8 Mathematics MCAS Test

A state test is expected to have a strong influence on what teachers do in their mathematics classes because teachers generally teach to what is on a test for which there is accountability. The Massachusetts tests are no exception. Thus a brief description of the test is warranted.

The Grade 8 Math MCAS test is based on the content standards for grades 7 and 8 in the 2000 Massachusetts Mathematics Curriculum Framework (Massachusetts Department of Education, 2000). These standards are grouped into five strands. The grade 8 test covers these five strands, requires application of three different types of thinking skills, and includes multiple-choice, short-answer, and open-response items. Table 5 shows the approximate score points and percentage of total score points for each of the content strands before and since 2001. As Table 6 shows, an adjustment for the 2001 tests reduced the percentage of total score points for Number Sense and Operations by 7% and increased the percentage of total points for Data Analysis, Statistics, and Probability by 5%. Those were the only changes in the weights of the strands in the test blueprint during these years.

Table 5. Approximate Percentage of Total Score Points for Common Items on the Massachusetts Grade 8 Mathematics Test by Framework Strand

<i>Framework Strand</i>	<i>Total Score Points 1998 to 2003</i>	<i>Percent of Total Score Points</i>	
		<i>Before 2001</i>	<i>From 2001</i>
Number Sense and Operations	14	33	26
Patterns, Relations, and Algebra	15	26	28
Geometry	7	13	13
Measurement	7	13	13
Data Analysis, Statistics, and Probability	11	15	20
TOTALS	54	100	100

Source: Massachusetts Department of Education

Note: Geometry and Measurement were in one strand in the test based on the original 1995 Framework. They were separated in the 2002 test, with each of the two new strands worth half of the combined points of the original strand.

Table 6 shows the distribution of score points for the common items by mathematical thinking skill.

Table 6. Approximate Percentage of Total Score Points for Common Items on the Massachusetts Grade 8 Mathematics Test by Mathematical Thinking Skill

<i>Thinking Skill</i>	<i>Total Score Points 1998-2003</i>	<i>Percent of Total Score Points</i>	
		<i>Before 2001</i>	<i>From 2001</i>
Procedural	14	30	26
Conceptual	16	25	30
Application/Problem Solving	24	45	44
TOTALS	54	100	100

Source: Massachusetts Department of Education

Table 7 shows the approximate distribution of items by type on each test form. Table 7 also shows how items are distributed between the common and matrix-sampled portions of the test. Note that the five Open-Response items account for almost two-fifths of the total score (37%), and that

Short-Answer and Open-Response items together account for almost half of the total score (46%).

Table 7. Approximate Number of Test Items on the Massachusetts Grade 8 Mathematics Test Per Form by Type, 1998-2003

	<i>Multiple-Choice</i>		<i>Short-Answer</i>		<i>Open-Response</i>		<i>Total Items Per Test Form</i>	
	<i># of Items</i>	<i>% of Total Score</i>	<i># of Items</i>	<i>% of Total Score</i>	<i># of Items</i>	<i>% of Total Score</i>	<i># of Items</i>	<i>% of Total Score</i>
Common	29	54	5	9	5	37	39	100
Matrix-Sampled	7		1		1		9	
Total per Form	36	54	6	9	6	37	48	100

Source: Massachusetts Department of Education.

It should be noted that the common items on each MCAS test are released to the public each year after that year's test results are released. The items that have been released since 1998 are all available on the Department's website (www.doe.mass.edu) and constitute a growing pool of practice items for teacher and tutor use.

B. Statistically Significant Factors Associated with the Principals

Statistical analysis of the principals' responses identified factors that were significantly associated with IPAW schools.⁶ Factors that were not significantly associated with IPAW schools are not reported here. Table 8 shows the factors significantly associated with the principals in the IPAW schools.⁷

Table 8. Factors Reported by Principals

Description of Factor Reported by Principal	IPAW Schools
Superintendent was involved in hiring new grade 8 mathematics teachers	No
"Planning and Delivering Lessons" identified as a professional development topic that would help teachers at school	No
Number of hours of professional development in mathematics content offered last year	< 15
Number of hours of professional development in mathematics pedagogy offered last year	< 10
"More remediation" was identified as a solution that was used to address general skill weaknesses in students	Yes

⁶ The analysis of administrator responses drew on only the responses provided by the principals of the 60 schools in which surveys were completed in order to have only one administrator response per school. The validity of this approach was checked by examining the agreement between the responses of the principal and the mathematics chair. There was virtually uniform agreement between principals and mathematics chairs. Teacher responses were weighted so that the collective teacher response from a given school was treated as a single response. This prevented responses from schools with two teachers taking the survey from being given twice the weight of schools with only one teacher taking the survey.

⁷ All tests of statistical significance of responses were performed at the $\alpha = 0.05$ level.

Table 8 indicates that district superintendents of IPAW schools were less likely to report being involved in hiring decisions than were the superintendent of non-IPAW schools. It also indicates that principals at IPAW schools were less likely to report that professional development on planning and delivering lessons would help the teachers at their schools than were principals at non-IPAW schools. Few of the principals (17%) that indicated that such training would help their teachers came from IPAW schools. Moreover, on average IPAW principals were less than half as likely as non-IPAW principals to report the need for such professional development.

Table 8 further indicates that IPAW schools differed from non-IPAW schools in the hours of professional development offered at the school. IPAW schools were more likely to offer fewer hours of professional development than non-IPAW schools. Specifically, in 2001-02, IPAW schools were more likely to offer 15 or less hours of development in math content, and 10 or less hours in math pedagogy than the non-IPAW schools. In the case of pedagogy, fewer than one out of five principals (17%) who indicated that their school provided over 10 hours of pedagogy development were from IPAW schools. Similarly, fewer than one out of five principals (12%) who indicated that their school provided over 15 hours of content development were from IPAW schools.

Finally, Table 8 indicates that principals at IPAW schools were more likely than principals at non-IPAW schools to use remediation as a solution to address students' general skills weaknesses. Principals in over two thirds of IPAW schools (70%) indicated that they used remediation, whereas only about one third of the principals from non-IPAW schools (35%) indicated the use of remediation.

The findings in Table 8 suggest that principals in IPAW schools were both more able to, and more likely to, choose teachers who they believed would be more effective with students in grade 8 (and possibly other middle school grades) than non-IPAW principals and who, as a possible consequence, did not need as much professional development in math content or pedagogy as those in non-IPAW schools. That IPAW schools also tended to use remediation much more heavily than non-IPAW schools may reflect the fact that they had proportionally more students at the lowest level in 1999 than did non-IPAW schools (almost half of their students), despite having a smaller proportion of LEP students.

C. Statistically Significant Factors Associated with the Teachers

Table 9 shows factors significantly associated with the teachers in the IPAW schools. As the Qualifications and Teamwork section of Table 9 indicates, IPAW schools were less likely to have teachers with either a middle school (MS) mathematics license or a secondary (SEC) mathematics license. Instructors from non-IPAW schools were nearly 1.5 times more likely to indicate that they held either a MS Math or a SEC Math license. Only 1 out of 4 instructors (26%) who indicated that they held either of the two mathematics licenses were from IPAW schools. IPAW schools were also

more likely than non-IPAW schools to have teachers who reported that they were well prepared to design assessments for lessons and units (3.5 times more likely, or 78%), less likely to have teachers who reported that they spend time planning in a group, and more likely to identify “new math teaching methods” as a professional development topic that would help teachers in the school.

Table 9. Factors Reported by Teachers

Qualifications and Teamwork	Associated with IPAW Schools
Instructor had an MS Mathematics <u>or</u> an SEC Mathematics license	No
Instructor indicated s/he was “well prepared” to design lessons and unit assessments	Yes
Instructor spent time planning instruction with other mathematics teachers	No
“New math teaching methods” was identified as a professional development topic that would help instructors at school	No
Student Placement	
Principal was a major influence on decisions about placement	Yes
Other individuals such as guidance counselors influenced placement decisions	Yes
“Math achievement in grades” identified as an important factor in placement decisions	Yes
“Parental selection” identified as an important factor in placement decisions	No
Class Time and Activities	
Students were assessed with tests or quizzes at least once per week in 2001-02	Yes
Calculators were used more than once per week in 2001-02	No
Instructor supplemented mathematics textbook with computers	Yes
Instructor supplemented mathematics textbook with calculators	No
Addressing Student Needs	
“Pedagogical change” was identified as an important means to address strand weaknesses	Yes
“Accelerated classes” were used as an important means to address needs of above grade students	Yes
“More hands-on approaches” was suggested as a strategy to increase student math learning	No
“More practice and homework” was suggested as a strategy to increase student math learning	No
Uses of MCAS Data	
Hours spent by mathematics teachers reviewing MCAS mathematics test results	> 7
Assistance with, or analysis of, MCAS results influenced preparation for MCAS	Yes
Assistance with, or analysis of, MCAS results influenced mathematics assessments used	Yes
Assistance with, or analysis of, MCAS results influenced expectations for learning	Yes
Assistance with, or analysis of, MCAS results influenced subject matter emphasized	Yes
Assistance with, or analysis of, MCAS results influenced homework assignments	Yes

The section of Table 9 on student placement into math classes shows that IPAW schools tended to have principals who were involved in placement decisions, and that these schools considered mathematics achievement as measured by prior grades in making placement decisions. Sixteen teachers in 10 of the IPAW schools indicated that some other individual influenced placement

decisions. Most (11) reported a guidance counselor as the other major influence.⁸ Two of the remaining five teachers reported the assistant principal, and the other three, erroneously,⁹ reported placement tests as the other major influence. One factor, parental selection, was negatively associated with IPAW schools, indicating that parents were a much smaller influence on the placement decision in IPAW schools than in non-IPAW schools. Ten of the 11 schools in which teachers reported parents as a placement influence were non-IPAW schools.

The Class Time and Activities section of Table 9 shows that teachers at IPAW schools were more likely to assess students with tests or quizzes; teachers at 16 of the 20 IPAW schools reported assessing students with quizzes at least once a week in 2001-02. In addition, roughly one half of the teachers from IPAW schools reported calculator use more than once per week in 2001-02, whereas fully three-fourths of non-IPAW teachers reported using calculators more than once a week. Teachers at IPAW schools were also less likely than teachers at non-IPAW schools to report supplementing the class text with calculators; less than one teacher in ten at IPAW schools reported supplementing the class text with calculators. On the other hand, teachers at IPAW schools were more likely to report supplementing the class textbook with computers. Nearly two thirds (65%) of teachers at IPAW schools indicated that they supplemented the class text with computers.

The Addressing Student Needs section shows the four factors related to potential solutions to needs or weaknesses in students. A pedagogy change was more likely to be used to address observed strand weaknesses in IPAW schools than in non-IPAW schools; teachers at 16 of the 20 IPAW schools indicated that a pedagogy change had been used for that reason. Accelerated classes were also more likely to be used in the IPAW schools than in the non-IPAW schools; teachers at 14 of the 20 IPAW schools indicated that accelerated classes were used to address the needs of above grade level students.

Two methods to help increase math learning for their students were significantly associated with non-IPAW schools. Teachers in 24 of the 40 non-IPAW schools suggested more hands-on approaches or more practice and homework or both as recommended means to increase student learning. In contrast, only teachers at five IPAW schools recommended either solution.

Finally, for the Uses of MCAS Data section of Table 9, teachers in IPAW schools were more likely than teachers in non-IPAW schools to have spent more than seven hours reviewing test results. In addition, teachers in all 20 IPAW schools indicated that their review of MCAS results tended to

8. It should be noted that the use of guidance counselors for placement was studied in both IPAW and non-IPAW schools. Based on statistical tests, there appeared to be no significant difference in the use of counselors for placement decisions between the two school groups *after controlling for the probability that a school reported some influence other than those provided on the list of close-ended items on the survey.*

9. The survey question was designed to elicit information about individuals that influenced placement decisions, not tests. These three responses may indicate a misunderstanding on the part of the respondents or possibly unclear phrasing of the survey item.

influence at least one of the following instructional practices: preparation of students for the MCAS test itself, classroom assessments used, the expectations for learning, the subject matter emphasized in class, and homework assignments.

III. FURTHER ANALYSIS OF THE TEACHER DATA ¹⁰

Department staff further analyzed the teacher data to see if more light could be shed on the central concern driving the study—the lack of a significant increase in the percent of students at the two highest performance levels on the state’s grade 8 mathematics test since 1998—and on the three central findings that emerged from the statistical analysis: teachers at the IPAW schools were more likely to report spending significant amounts of time reviewing and using MCAS results, more likely to report the use of accelerated classes as a way to address the needs of above grade students, and *less* likely to use calculators or suggest “hands-on approaches,” practice, and homework as strategies to improve student learning in mathematics. We were especially interested in the influence of the tests themselves.

A. The Influence of the State’s Grade 8 Mathematics Test on Teachers’ Practices

The state’s grade 8 mathematics tests influenced the two groups of teachers in many ways, sometimes differently and sometimes similarly. We report the similarities we found between the two groups, as well their differences, when they seemed helpful in understanding the differences.

Differences between the Two Groups: When asked if analysis (or assistance in analysis) of MCAS results influenced various school-wide and classroom practices, Table 10 shows that a larger percent of IPAW (85%) than non-IPAW teachers (71%) consistently responded positively. Table 11 shows whether they thought they spent more, the same amount of, or less time teaching the content of the

Table 10. Responses to “Did analysis of MCAS results influence any of the following? (Check all that apply.)”

	IPAW (n=41)		Non-IPAW (n=63)	
	#	%	#	%
Curricular materials purchased	21	51.2	22	34.9
Course content	26	63.4	31	49.2

¹⁰ This extended analysis was done by Sandra Stotsky, Senior Associate Commissioner and Director of the Center for Teaching and Learning at the Massachusetts Department of Education at that time, and Kathe Kirkman, a program evaluator at the Department. For the extended analysis, we applied somewhat stricter criteria for including schools in the two groups of interest. As a result, two schools in the non-IPAW group were eliminated from this analysis and one was moved from the non-IPAW group to the IPAW group (see the note in the Appendix). Added to the IPAW group (where its percent of improvement placed it) was another school whose teacher data had not been used earlier because of a lack of data from the school’s principal. Thus, the extended analysis used data from 104 teachers in all, 41 in 22 IPAW schools and 63 in 37 non-IPAW schools. There were a few non-responses on some questions, roughly the same number in each group. We note when the non-response rate is high on a particular question.

Preparation for MCAS test taking	34	82.9	43	68.3
Mathematics assessments used	20	48.8	24	38.1
Classroom assignments	23	56.1	22	34.9
Professional development content	24	58.5	27	41.8
Amount of professional development time	12	29.3	12	19.0
Classroom practice	41	100	58	92.1
Your own classroom instructional approach	30	73.2	35	55.6
Your own classroom preparation of students for MCAS test taking	37	90.2	48	76.2
Your own classroom mathematics assessments used	27	65.9	26	41.3
Your own classroom expectations for student learning	26	63.4	30	47.6
Your own classroom subject matter emphasized	35	85.4	45	71.4
Your own classroom homework assignments given	21	51.2	24	38.1
Your own classroom use of curricular materials	22	53.7	22	34.9

Source: Data from files sent by Thomas, Warren + Associates to the Massachusetts Department of Education.

different strands in 2002 compared to 1999. A higher percent of IPAW teachers reported spending more teaching time on two of the content strands (number sense and operations, and measurement) in 2002, while non-IPAW teachers reported spending more teaching time on the other three (patterns, relations, and algebra, geometry, and data analysis, statistics, and probability).

Table 11: Responses to “Time spent teaching _____ strand as compared to 1999.”

		<i>IPAW</i>		<i>Non-IPAW</i>	
		<i>Frequency</i>	<i>Percent</i>	<i>Frequency</i>	<i>Percent</i>
Number Sense & Operations	<i>More time</i>	8	19.5	10	15.9
	<i>About the same time</i>	14	34.1	33	52.4
	<i>Less Time</i>	4	9.8	3	4.8
Patterns, Relations & Algebra	<i>More time</i>	10	24.4	21	33.3
	<i>About the same time</i>	14	34.1	25	39.7
	<i>Less time</i>	2	4.9		
Geometry	<i>More time</i>	11	26.8	15	23.8
	<i>About the same time</i>	13	31.7	28	44.4
	<i>Less time</i>	1	2.4	3	4.8
Measurement	<i>More time</i>	9	22.0	7	11.1
	<i>About the same time</i>	11	26.8	29	46.0
	<i>Less time</i>	5	12.2	8	12.7
Analysis, Statistics & Probability	<i>More time</i>	13	31.7	27	42.9
	<i>About the same time</i>	8	19.5	18	28.6
	<i>Less time</i>	4	9.8	1	1.6

Source: Data from files sent by Thomas, Warren + Associates to the Massachusetts Department of Education.

Note: About one-third of the teachers in each group did not respond to this question in 2002 because they were not teaching mathematics in 1999 or teaching at all. This is about the rate of teacher turnover today.

Similarities between the Two Groups: The similarities on this issue were informative. For both groups of teachers, a much smaller percent report spending more time in 2002 teaching the content of each strand than the percent reporting having spent the same amount of time (or less). The only exception is for data analysis, statistics, and probability, the strand for which the percentage of total points on the grade 8 test increased in 1991 from 15% to 20%. For this strand, almost one third of

the IPAW teachers and over 40% of the non-IPAW teachers reported spending more teaching time in 2002, suggesting how strong an influence the change in points for that strand was. What is surprising is that only 23% of the IPAW teachers and 21% of the non-IPAW teachers reported teaching more content in the other four strands in 2002 than they did in 1999. The question that arises is why the majority of teachers in both groups did not spend more time teaching content in 2002 than in 1999, given the pressure on the schools to show improvement over time (from media reports on school and district performance and from performance ratings by Department staff) as well as the slight increase in the overall percent of students at the two highest performance levels over these four years.

Responses to several other questions throw some light on this question. When asked to rank nine options on how they spent their teaching time, the two options receiving the highest percent of first and second choice responses by both groups of teachers were “explaining concepts/procedures” (76% of the IPAW teachers and 73% of the non-IPAW teachers) and “demonstrating problem-solving” (41% of the IPAW teachers and 64% of the non-IPAW teachers). When asked to rank eight options on how their students spent classroom time in 2002, the option receiving the highest percent of first and second choice responses by both groups of teachers by far was “understanding and solving problems” (63% of the IPAW teachers and 76% of the non-IPAW teachers, followed by “reviewing homework” (51% of the IPAW teachers, and 38% of non-IPAW teachers) and “learning to use algorithms” (34% of the IPAW teachers and 38% of the non-IPAW teachers). (The percents for “reviewing homework” are a little puzzling because only 36% of the IPAW teachers and 42% of the non-IPAW teachers reported assigning homework daily.)

Given the current emphasis on problem solving in mathematics education and the weight attached to open-response items (which include the writing out of explanations) on the state tests, these rankings may well translate into large chunks of class time for both groups of teachers (especially if homework also consists of problem solving—which is likely), and perhaps an excessive amount for non-IPAW teachers. Although a higher percent of IPAW teachers reported spending seven or more hours analyzing MCAS results and changing one or more instructional practices, nevertheless, a high percent of the non-IPAW teachers (56%) did report spending five or more hours assisting students with test strategies in 2002. Test preparation was also the most frequent recommendation by both groups of teachers for getting more students to the two highest levels.

B. More Time for Teaching Wanted

Differences between Both Groups: Although a majority of teachers reported spending at least two periods to complete a lesson, more IPAW teachers reporting doing so than non-IPAW teachers.

Only 24% of the IPAW teachers in contrast to 41% of the non-IPAW teachers answered “in one period” to the question of how much time “it typically takes to complete a math lesson.”

Similarities between Both Groups: Both groups of teachers not only wanted more time for teaching mathematics and giving individualized instruction but also ranked their choice of strategies similarly. Teachers were asked to rank 23 different strategies that “increase math learning.” The two options receiving the highest percent of first or second choice responses were “more class time (20% for IPAW teachers and 25% for non-IPAW teachers) and “decreased class size” (42% for IPAW teachers and 46% for non-IPAW teachers). And this is despite the fact that average class size across all the schools in the study was 21, with the vast majority 25 or under.

Part of the explanation for this somewhat puzzling phenomenon (an already reasonable class size but a need for more teaching time) may lie not only with an increase in the instructional and practice time needed for problem solving and test preparation for MCAS but also in how teachers organize class time for problem solving activities. It is useful to note that 25% of the IPAW teachers and 22% of the non-IPAW teachers ranked “supervising small group work” first or second as a way time was spent teaching. Small group work takes up much more class time than whole class instruction or individual work with respect to covering content, even though there is little if any evidence to support its efficacy in mathematics, and it may reduce teaching time for above grade students. No studies provide information on how much of a trade-off between content and process the use of this teaching strategy may amount to, especially in science and math classes (see Gross & Stotsky, 2000, for a discussion of this issue).

Part of the explanation for this phenomenon may also lie with an increase in the number of students with limited English or learning or behavior problems in general classrooms in grade 8. More than 20% of the teachers overall reported an increase in both these groups of students since 1998. Yet a relatively low percent feels well prepared to teach “students with limited English skills” (10% of the IPAW teachers and 16% of the non-IPAW teachers) or students with severe discipline problems (33% of the IPAW teachers and 23% of the non-IPAW teachers). This may account in part for why 33% of the IPAW teachers and 41% of the non-IPAW teachers ranked “pull-out for math” first or second as a means to address below grade students.

Mathematics teachers in general grade 8 classrooms today may be finding it increasingly difficult to teach classes exhibiting a wide and increasingly widening diversity of instructional needs and levels. In some school districts, close to 20% of the students have Individual Educational Plans and most of them may now be taught mathematics in the general classroom. When asked how many instructional levels were in their heterogeneously grouped classes in 2002, a large number did not respond at all; of those who responded, most IPAW teachers indicated one or two, while non-IPAW teachers in roughly equal numbers indicated 1, 2, or 3 or more levels. This context suggests why, in

ranking ten options for remediating below grade students, 26% of the teachers overall ranked “use of same skill level materials” first or second, this option receiving the third highest percent after tutoring and teacher-led after-school sessions.

This does not mean that middle school teachers are paying less attention to their low-performing students. Indeed, as Table 1 showed, there has been a greater decrease in the percent at the lowest performance level since 1998 than an increase in the percent at the two highest levels. This phenomenon appeared on the grade 8 tests in mathematics conducted by the National Assessment for Educational Progress from 1992 to 2000, as Table 12 shows, with a slightly higher

**Table 12. Results in Grade 8 Mathematics
on the National Assessment of Educational Progress for Massachusetts**

	<i>Scaled Score for the USA</i>	<i>Scaled Score for Massachusetts</i>	<i>Percent Below Basic</i>	<i>Percent at Basic</i>	<i>Percent at Proficient</i>	<i>Percent at Advanced</i>
1992	267	273	37	40	20	3
1996	271	278	32	40	23	5
2000	274	283	24	44	26	6
2003	278	287	24	38	30	8

Source: Massachusetts Department of Education

increase in the percent of Massachusetts grade 8 students at the two highest levels on the NAEP mathematics test by 2003 (15%) than the decrease in the percent at Below Basic (13%).

C. Addressing the Needs of Above and Below Grade Students

Because time is needed to teach problem solving in the way in which it is assessed on the state’s mathematics test and to teach it to a wider range of achievement levels in their classes, and because accountability compels teachers to spend a great deal of time teaching their low-performing students, we sought to find out more about the differences between the two groups of teachers in what they use or recommend for meeting the needs of above grade students.

Differences between the Two Groups: When asked to rank seven strategies they have used to address the needs of above grade students; 66% of the IPAW teachers and 41% of the non-IPAW teacher ranked “accelerated classes” first or second. In addition, 63% of the IPAW teachers and 49% of the non-IPAW teachers ranked “levels of algebra 1” (homogenous classes for students at different levels of achievement in mathematics) first or second, as shown in Table 13. Almost none of the teachers reported using “pull-out for math” as a means to address needs of above grade students. Interestingly, the spread between the two groups was similar with respect to below grade students; 67% of the IPAW teachers and 46% of the non-IPAW teachers ranked “special education placement” first or second as a means to address below grade students.

Table 13. Responses to “Rank order the three most important procedures that were used to

address the needs of grade eight students with above grade level skills.”

		<i>IPAW</i>		<i>Non-IPAW</i>	
		<i>Frequency</i>	<i>Percent</i>	<i>Frequency</i>	<i>Percent</i>
Levels of Algebra 1	<i>1st Strategy</i>	19	46.3	27	42.9
	<i>2nd Strategy</i>	7	17.1	4	6.3
Pull-out for Mathematics	<i>1st Strategy</i>	1	2.4	2	3.2
	<i>2nd Strategy</i>				
Special Education Placement	<i>1st Strategy</i>			1	1.6
	<i>2nd Strategy</i>			2	3.2
Accelerated Classes	<i>1st Strategy</i>	14	34.1	15	23.8
	<i>2nd Strategy</i>	13	31.7	11	17.5
Alternative Curriculum	<i>1st Strategy</i>			4	6.3
	<i>2nd Strategy</i>	3	7.3	5	7.9
Math Club	<i>1st Strategy</i>	3	7.3	4	6.3
	<i>2nd Strategy</i>	6	14.6	6	9.5

Source: Data from files sent by Thomas, Warren + Associates to the Massachusetts Department of Education.

Similarities between the Two Groups: Although IPAW teachers ranked the use of “algebra 1 levels” for above grade students much higher than did non-IPAW teachers, about the same percent of teachers in both groups (28% of the IPAW teachers and 23% of non-IPAW teachers) ranked it first or second as a strategy they have used to address the needs of on grade students. The two groups addressed several other aspects of grouping similarly. When asked to rank 12 “helpful professional development topics,” 24% of the IPAW teachers and 30% of the non-IPAW teachers ranked “homogeneous groups for instruction” first or second. This topic received the fourth highest weight, after “using new instructional material,” “new math teaching methods,” and “needs of below grade students.” And when asked to rank 23 strategies to “increase math learning,” the strategy receiving the third highest weight was “achievement/skill grouping,” with 22% of the teachers overall ranking it first or second.

D. Differences in Calculator Use between the Two Groups

As noted earlier, teachers in the IPAW schools reported using calculators less frequently than did teachers in non-IPAW schools. But did the type of class the teacher taught make a difference? As Table 14 shows, the type of mathematics class they taught did not seem to make a difference. The percent of self-described algebra teachers was roughly equivalent in both groups: 14 of 34 IPAW teachers (40%), and 25 of 59 non-IPAW teachers (42%) (nine teachers did not report what they taught), with 57% of the algebra teachers in the IPAW schools and 64% of the algebra teachers in the non-IPAW schools reporting use of calculators two to five times a week. The difference showed up in the non-algebra classes: while 43% of the teachers not teaching algebra in the IPAW group reported using calculators two to five times a week, 71% in the non-IPAW group did. In other words, the significant increases in the percent of students at higher performance levels in the IPAW

schools may be related to less frequent use of calculators in the non-algebra classes, a possibility that is consistent with the correlation between high calculator use and low NAEP scores, especially for low-income and minority students, reported by Loveless (2000).

Table 14. Frequency of Use of Calculators in the Classroom

	Algebra		Not-Algebra	
	Number	Percent	Number	Percent
IPAW Teachers				
Two to Five Times a Week	8	57.1%	9	42.8%
Once a Week or Less	6	42.9%	12	57.1%
TOTAL	14		21	
Non-IPAW Teachers				
Two to Five Times a Week	16	64.0%	24	70.6%
Once a Week or Less	9	36.0%	10	29.4%
TOTAL	25		34	

Source: Data from files sent by Thomas, Warren + Associates to the Massachusetts Department of Education.

E. The Extent to Which Teachers' Professional Judgments Matter

Differences between the Two Groups: That many teachers report using new instructional materials, some of which do not seem to be at the instructional level for many below grade students, raised the question in our minds of who is choosing these materials. When asked to rank seven options for influence on curricular materials, "central office personnel" was rated first or second by 22% of the IPAW teachers and 61% of the non-IPAW teachers, "math curriculum specialist" was rated first or second by 48% of the IPAW teachers and 65% of the non-IPAW teachers, and "middle school math teachers" was rated first or second by 62% of the IPAW teachers and 44% of the non-IPAW teachers. Teachers in the IPAW group clearly seem to have a stronger voice in what they use for mathematics instruction, a not unimportant factor in promoting higher achievement for **all** students but especially above or on grade students. Teachers as a group tend to have a better sense that central office personnel of the difficulty level of mathematics curriculum materials and what would appropriately challenge their above grade students

To explore this factor further, we listed the titles, publishers, and dates of the textbooks the grade 8 teachers in both groups said they were using in the questionnaire responses and asked the Department's former state coordinator of mathematics and a colleague (each is a former high school mathematics teacher and at present a mathematics coordinator in an urban school district) to classify them as "traditional" or "reform," based on their understanding of how these two terms applied to mathematics textbooks. Based on their judgments, the type most frequently used in the IPAW schools is "traditional" (12 schools). (We had received no information from teachers in two IPAW schools, mixed information from one IPAW school, and no or no clear information from teachers in five non-IPAW schools.) What is more interesting, as Table 15 shows, is that the increase in the

percent of students at the two highest levels, and in the percent of students moving up from the lowest level, is greater in the IPAW schools using “traditional” textbooks than in the IPAW schools using “reform” textbooks (7). In other words, although teachers in seven IPAW schools use “reform” textbooks, their schools did not show as much improvement for both high and low students as did the 12 IPAW schools whose teachers use “traditional” textbooks. In the non-IPAW schools, “traditional” is far more frequently used than “reform,” but the overall profile is less clear because of a higher non-response rate. It is possible that some of the differences in outcomes between the IPAW and non-IPAW schools that use “traditional” textbooks result from their use with “accelerated” classes. A much higher percent of IPAW teachers reported using “accelerated” classes to address the needs of above grade students than did non-IPAW teachers.

Table 15. Increases in Percent of Students at Proficient and Advanced Levels and Decreases in Percent of Students at the Warning Level on the Massachusetts Grade 8 Mathematics Test from 1999-2002

	<i>Increases in Proficient/Advanced</i>	<i>Decreases in Warning</i>
Statewide	4.29	-5.95
IPAW schools using “reform” textbooks (N = 7)	7.71	-11.57
IPAW schools using “traditional” textbooks (N = 12)	15.42	-18.75

Similarities between the Two Groups: On the other hand, when asked who influences decisions about placement, the ranking of the two groups is similar. As Table 16 shows, 61% of teachers in the IPAW group and 70% of teachers in the non-IPAW group ranked teachers as either the first or second influence. (As reported earlier, principals in the IPAW schools reported having a significantly greater voice in placement decision than principals in non-IPAW schools—46% to 29%, respectively, as the first or second influence.)

Correspondingly, as Table 17 shows, in their ranking of factors considered for placement, 44% of IPAW teachers and 52% of non-IPAW teachers ranked “demonstrated achievement” as the first or second factor considered. Approximately 44% of both groups ranked math achievement in grades as first or second and 44% ranked teacher recommendation as first or second.

Table 16. Responses to “Rank of _____ as influence in decisions about placement.”

		<i>IPAW Teachers</i>		<i>Non-IPAW Teachers</i>	
		<i>Frequency</i>	<i>Percent</i>	<i>Frequency</i>	<i>Percent</i>
Principal	1 st	14	34.1	12	19.0
	2 nd	5	12.2	6	9.5
Teachers	1 st	10	24.4	33	52.4
	2 nd	15	36.6	11	17.5
Department Head	1 st			2	3.2
	2 nd	4	9.8	5	7.9
Mathematics Specialist	1 st	2	4.9	1	1.6

	2 nd	2	4.9	2	3.2
Team Leader	1 st			2	3.2
	2 nd			2	3.2
Parents	1 st	2	4.9	5	7.9
	2 nd	5	12.2	14	22.2
Students	1 st	1	2.4	1	1.6
	2 nd	1	2.4	2	3.2

Source: Data from files sent by Thomas, Warren + Associates to the Massachusetts Department of Education

There is clearly a disconnect between two very important sets of instructional decisions. Decisions about the appropriate instructional placement of students tend to be made by one group of educators (teachers), while decisions about the curriculum materials these teachers must use with their students tend to be made by another group of educators (administrators), especially in the non-IPAW schools. Nor does it seem to be the case that teachers have much voice in deciding whether

Table 17. Responses to “Rank order of _____ as one of the four most important factors considered when placing students in math courses in your school.”

		<i>IPAW Teachers</i>		<i>Non-IPAW Teachers</i>	
		<i>Frequency</i>	<i>Percent</i>	<i>Frequency</i>	<i>Percent</i>
Demonstrated Achievement	1 st	12	29.3	26	41.3
	2 nd	6	14.6	7	11.1
Math Achievement based on MCAS	1 st				
	2 nd	1	2.4	4	6.3
Math Achievement based on Grades	1 st	9	22.0	14	22.2
	2 nd	9	22.0	14	22.2
Math Achievement based on Courses Taken	1 st	2	4.9	2	3.2
	2 nd	1	2.4	3	4.8
Teacher Recommendation	1 st	8	19.5	13	20.6
	2 nd	10	24.4	15	23.8
Student Selection	1 st			1	1.6
	2 nd	1	2.4	3	4.8
Parent Selection	1 st			1	1.6
	2 nd			3	4.8
Limited English Proficiency	1 st			1	1.6
	2 nd			3	4.8
Individual Education Plan (IEP)	1 st	1	2.4	5	7.9
	2 nd	3	7.3	6	9.5
Random Grouping	1 st	3	7.3	5	7.9
	2 nd	3	7.3	2	3.2

Source: Data from files sent by Thomas, Warren + Associates to the Massachusetts Department of Education.

accelerated classes can be offered to meet the needs of above grade students. This disconnect should be addressed if teachers are to be held accountable for the achievement of their students, especially if the needs of above grade students in grade 8 are to be addressed more satisfactorily than they now appear to be.

That this disconnect is not at all uncommon was further suggested by two newspaper articles highlighting **teachers'** dissatisfaction with the choice of mathematics programs made by their

administrators that appeared during the time this report was being written. One appeared in the October 9, 2003, *Andover Townsman* (Massachusetts) on parent complaints about a “perceived lack of challenging material in a new math program for seventh-grade students.” According to the article, parents charged that “math teachers were being forced to act supportive of a program they have concerns about.” Administrators were quoted as saying that teachers were fully aware of the new program and helped to select it themselves. However, the vice president of the local teachers’ union confirmed to the reporter that teachers had made their concerns about the new curriculum known, but administrators had chosen to implement it any way.

The second article appeared in the form of a very long editorial in the *Boston Globe* on November 8, 2004. It noted the “collective groans” of over 100 Boston teachers attending a weekend retreat when the name of their K-5 math program was brought up. The program was introduced into the school system by a top-down administrative decision, with little teacher input, and, according to the mathematics director for the Boston schools, leaves Boston’s students without the strong computational skills needed for higher level mathematics courses. Not only must the teachers figure out how to supplement the program’s many deficiencies, they must also take massive, never-ending professional development and—to rub more salt into the wounds—have coaches. This kind of administrative action has been and continues to be systemic in the Boston public schools at all grade levels.

IV. Discussion

There has been a greater decrease in the percent of students at the lowest level than an increase in the percent of students at the two highest levels in grade 8 on the mathematics MCAS tests from 1998 to 2003. The situation is similar on the NAEP mathematics assessments for Massachusetts from 1992 to 2000, with a greater increase at the two highest NAEP levels than a decrease at its Below Basic level apparent only by 2003. The purpose of this study was to identify the school-based factors that are significantly associated with schools whose higher and lower achieving students had improved their scores on the state’s grade 8 mathematics tests between 1999 and 2002. The following factors were significantly associated with the IPAW schools:

- First, the time teachers spent reviewing and using test results. Teachers at the IPAW schools were more likely to report spending significant amounts of time reviewing and using MCAS results. The way in which mathematics content is categorized and weighted in the grade 8 mathematics test, as well as the weight given open-ended responses, have an enormous influence

on what teachers do in their classrooms. Teachers in the IPAW schools reported making changes in every aspect of instruction in response to the tests' format, demands, and results.¹¹

- Second, restrained use of calculators in non-algebra classes. Teachers at the IPAW schools were *less* likely to report using calculators or to suggest hands-on approaches, practice, or homework as strategies to improve student learning in mathematics. Less frequent use of calculators in the IPAW schools took place in the non-algebra classes. The differences in calculator use between algebra teachers in the two groups are small.
- Third, accelerated or leveled (homogeneous) algebra I classes for above grade students. A higher percent of IPAW teachers report using accelerated or leveled algebra I classes to address the needs of above grade students.
- Fourth, teachers' role in choosing their mathematics program. A higher percent of teachers in the IPAW group reported having a voice in choosing their curriculum materials. Non-IPAW teachers tended to see central office administrators or curriculum specialists making the choice.
- Fifth, use of textbooks classified by two independent raters as "traditional." When the textbooks used by the teachers in the IPAW schools were rated as "traditional" or "reform," the type most frequently used in the IPAW schools is "traditional." Further, the increase in the percent of above and below grade students moving into higher performance levels is greater in the IPAW schools using "traditional" textbooks than in the IPAW schools using "reform" textbooks.

The fact that the IPAW schools used accelerated and leveled (homogenous) classes to increase the percent of students at the two highest levels and at the same time decreased the percent of students at the lowest performance level deserves more comment. It contradicts the claim that achievement grouping (whether within or across classes) retards the progress of below grade students. Instead, it provides further support for the studies cited by Loveless (2000) showing that "ability grouping's effect is consistently positive, especially in math," and that students in tracked classes at both grades 4 and 8 "registered higher math scores than the untracked students." The IPAW teachers were clearly not using or recommending use of the same mathematics program for all

¹¹ Although one might view these teachers as "teaching to the test" and thus view their activities negatively, this is exactly what those who developed the original mathematics standards and assessments in the mid-1990s wanted. The mathematics assessments were designed as "state of the art" tests by educators, technical consultants, and others involved in the State Systemic Initiative who strongly favored the "reform" mathematics programs recommended by the National Science Foundation. They explicitly wanted to use state assessments in mathematics (as well as the newly redesigned mathematics tests given by the National Assessment of Educational Progress) as the means for changing the content of the K-12 mathematics curriculum and teachers' instructional strategies in the state. It is ironic that while the mathematics MCAS tests have led to changes in many teachers' instructional strategies and influenced many school systems to adopt "reform" mathematics programs, especially in K-8, the results of these tests in K-8 in Massachusetts, and at the national level on NAEP tests, do not show much improvement in student achievement since their inception.

students (those below, on, and above grade), the condition under which there seems to be little advantage in achievement grouping for above or below grade students (Whitehurst, 2003; Benbow & Stanley, 1996). The findings of this study, therefore, however tentative, tend to undermine the notion that grouping students by achievement level within or across mathematics classes is likely to widen achievement gaps rather than diminish them. Indeed, they reflect the practices and views of principals and experienced grade 8 teachers in schools that are being held accountable for their students' achievement on state tests.

A possible reason why IPAW teachers highly preferred achievement grouping as a way to improve mathematics learning may relate to the needs of on grade students—traditionally the most neglected students in our schools—the average student, or the student who may be close to or barely at the Proficient level. Whether or not grade 8 teachers feel confident about their ability to address the instructional needs of all their students in heterogeneous classes with a very wide range in achievement (wanting smaller class sizes, more teaching time, more pull-outs, tutoring, and special education placements for below grade students), the recommendation for leveled algebra I classes for on grade students and for curriculum materials targeted to below grade students make sense. It makes especially good sense in the context of accountability.

There are clear sanctions for schools that have large numbers of students at the lowest level in grade 8 who are not moving at a regular pace to the next higher level. Incentives are built into No Child Left Behind requirements to move all grade 8 students currently below the Proficient level to the Proficient level by 2014, and thus the major focus of the schools is apt to be those students at the lowest level. Unfortunately, there is no incentive in NCLB requirements or the state's own accountability system in grade 8 to move those already (or barely) at the Proficient level to the Advanced level, and no incentive to focus on those students at the Advanced level at all.

IV. CONCLUSIONS AND SUGGESTIONS FOR FURTHER STUDY

In Massachusetts and in other states, further exploration of the major findings of this study should be carried out with teachers in schools with a **high** percent above the state average both in increasing the percent of students at the two highest levels and in simultaneously decreasing the percent of students at the lowest level. Some of the schools randomly chosen for this study were barely above the state average in one or both of these ways. Results might be clearer and thus more informative than the findings of this study if inquiry were concentrated more consistently on schools with a high percent above the state average in both respects.

The following questions seem especially worthwhile to pursue. Do students taught by teachers who have a strong voice in selecting their school's mathematics curriculum materials and organizing the mathematics classes in their school have a higher level of achievement than students in schools

where decisions about mathematics curriculum materials and classroom organization are made by top-level administrators? If all grade 8 teachers were allowed to choose textbooks for their students with above average achievement by grade 8, what kinds of textbooks would they choose? As Schmidt and others remark in Why Schools Matter (2001), "...textbooks exert a strong influence on what teachers teach....Textbook coverage is important both for what topics are taught and for the levels of performances and accomplishments expected of students" (p. 357). Finally, why do many grade 8 teachers recommend special education placements and pull-outs as ways to increase the achievement of below grade students?

Questions about classroom practices could easily be pursued using the basic research design of this study. Is use of a calculator most effective after students have learned enough mathematics to understand what they are doing, as suggested by the finding on less frequent calculator use in non-algebra IPAW classes? And, what is the frequency with which various "traditional" and "reform" practices are used in schools moving higher percents of on, above, and below grade students into higher performance levels compared to schools that do not? Of particular interest is the balance between small group work, whole class instruction, and individual work in both groups of schools, as well as differences, if any, in the extent to which students are expected to "discover" mathematical concepts and in the amount of classroom time that such discovery takes.

A consistent body of research suggests that teachers' knowledge of the subject matter they teach is the determining factor in student achievement in mathematics. While the knowledge of mathematics that teachers of the subject bring to their teaching will always be a relevant influence on student learning at any grade level, this study suggests the importance of other school-based factors at the middle school level, especially since there were no dramatic differences in teacher qualifications between the two groups (as indicated in Table 3). This implication is further supported by the fact that only 27% of American students in grade 8 scored in the two highest categories on the grade 8 mathematics test given in 2003 by the National Assessment of Educational Progress, a pitifully small percent.

If a much larger random sample of schools could be used (basically a question of cost), it might be possible to explore the effects of particular curricula on grade 8 achievement. Given the large number of different textbooks or mathematics programs now used in grades 6 to 8 (as we discovered from the titles of the textbooks the teachers indicated they were using, encompassing different editions or revisions of a textbook as well as different publishers), it is not possible with a small random sample to find enough schools using the exact same program to draw conclusions about the effects of any one program.

The design of this study might also be useful in any state to evaluate the effectiveness of professional development programs for middle school mathematics teachers. At present, state departments of education are now funding a number of middle school mathematics initiatives through Title II B Mathematics and Science Partnership grants. Annual changes in student scores in participating school districts in each state expressed as a percent of above grade, on grade, and below grade students moving into higher performance levels on statewide tests above or below a state average could serve as one objective index with which to evaluate the usefulness of these grants.

It is important to keep in mind that the schools in this study are for the most part typical public schools, whether urban, suburban, or rural, and the teachers in them must work within their constraints. Moreover, the research design deliberately over-sampled urban schools. The nationally acclaimed KIPP Academies—two charter-like schools for grades 5 to 8, one in South Bronx and the other in South Houston—have demonstrated that it is possible to increase the mathematics achievement of very large classes of low-income (and mostly below-grade) students dramatically and without grouping. In fact, by grade 8, all of the students in these schools are taking Algebra 1, by common consensus the “gateway” course for more advanced mathematics and science courses in all four years of high school (U.S. Department of Education, 1997). Abigail and Stephan Thernstrom describe the KIPP schools and other schools like them in their book on the achievement gap, *No Excuses: Closing the Racial Gap in Learning* (2003), stressing that high-achieving schools for low-income students tend to have dedicated principals who can choose their teachers, their curriculum materials, and the kinds of classroom organizations they want, as well as maintain a disciplined and structured learning environment within and outside the classroom. More studies are needed that give middle school teachers and principals the opportunity to voice their views on what they think would increase mathematics learning in **all** their students, especially for the **above grade** mathematics students in low-income urban schools.

REFERENCES

- Benbow, C.P., & Stanley, J. (1996). INEQUALITY IN EQUITY: How equity can lead to inequity for high-potential students." Psychology, Public Policy, and Law, Vol. 2, No. 2. Pp. 249-292.
- Boston Plan for Excellence. (2002a) Fact sheet: Major accomplishments of the Boston Plan • 1995-2002, Boston Plan for Excellence in the Public Schools Foundation. Boston, MA. ([http://www.bpe.org/betProficient and Advancedubs/fs/FS_Acc_%200102.pdf](http://www.bpe.org/betProficient%20and%20Advancedubs/fs/FS_Acc_%200102.pdf))
- Boston Plan for Excellence. (2002b) Fact sheet: What the Boston plan does • 2001-02, Boston Plan for Excellence in the Public Schools Foundation. Boston, MA. ([http://www.bpe.org/betProficient and Advancedubs/fs/FS_Activ_0102.pdf](http://www.bpe.org/betProficient%20and%20Advancedubs/fs/FS_Activ_0102.pdf))
- Carnine, D., & Gersten, R. The nature and roles of research in improving achievement in mathematics. Journal for Research in Mathematics Education, 2000, Vol. 31, No. 2, 138-143.
- Gross, P.R., & Stotsky, S. How children learn science: Do we now know? In S.Stotsky (Ed.), What's at stake in the K-12 standards wars: A primer for educational policy makers. NY: Peter Lang, 2000.
- Hiebert, J., and others. Teaching mathematics in seven countries: Results from the TIMSS 1999 Video Study NCEES (2003-013), Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Loveless, T. (1999). The tracking wars. Washington, D.C.: The Brookings Institution.
- Loveless, T. (2000). How well are American students learning? Focus on math achievement. Vol. 1, No. 1. The Brown Center Report on American Education. Washington, D.C.: The Brookings Institution.
- Lohr, S. (1999). Sampling: Design and analysis. Pacific Grove, CA: Brooks/Cole Publishing.
- Massachusetts Department of Education. (1999a) Download page for statewide MCAS school and district results – 1999. Malden, MA. (<http://www.doe.mass.edu/mcas/1999/results/data/district/G8dis99.txt>) and (<http://www.doe.mass.edu/mcas/1999/results/data/district/G8sch99.txt>)
- Massachusetts Department of Education. (1999b) Layout of 1999 MCAS data files. Malden, MA. (http://www.doe.mass.edu/mcas/1999/results/data/dbf_layout.html)
- Massachusetts Department of Education. (1999c) The Massachusetts Comprehensive Assessment System: Release of spring 1999 test items. Malden, MA.
- Massachusetts Department of Education. (2000) Massachusetts mathematics curriculum framework. Malden, MA. [On-line]. Available: <http://www.doe.mass.edu/frameworks>
- Massachusetts Department of Education. (2001a) Download page for 2001 MCAS results. Malden, MA. <http://www.doe.mass.edu/mcas/2001/results/data/G8DIS01.TXT> and <http://www.doe.mass.edu/mcas/2001/results/data/G8SCH01.TXT>
- Massachusetts Department of Education. (2001b) Massachusetts Board of Education 2001 annual report. Malden, MA.
- Massachusetts Department of Education. (2001c) The Massachusetts Comprehensive Assessment System: Release of spring 2001 test items. Malden, MA.
- Massachusetts Department of Education. (2002). Reports issued on 2002 school panel reviews to identify under-performing schools. Malden, MA. [On-line]. Available: [http://www.doe.mass.edu/atProficient and Advancedanel/02/](http://www.doe.mass.edu/atProficient%20and%20Advancedanel/02/)
- Massachusetts Department of Employment and Training. (2001) MassStats: Your connection to the Massachusetts economy. Boston, MA. [On-line]. Available: http://www.detma.org/forms/pdf/2137_701.pdf
- Massachusetts Education Reform Review Commission. (2000a) How Massachusetts schools are using MCAS to change curriculum, instruction, assessment, and resource allocation. Boston, MA.

- Massachusetts Education Reform Review Commission. (2000b) Districts' perspectives on the Massachusetts Department of Education's capacity to support professional development in the implementation of the Massachusetts Education Reform Act. Boston, MA.
- Massachusetts Education Reform Review Commission. (2001) Annual report 2001. Boston, MA.
- Panasuk, R., & Sullivan, M. (1998). Need for lesson analysis in effective lesson planning. *EDUCATION* (an international journal). Vol. 118, pp. 330-345.
- Riordan, J., & Noyce, P. The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education*. 2001, Vol. 32, No. 4, 368-398.
- Russo, A. (July/August 2004). School-based coaching: A revolution in professional development—or just the latest fad. *Harvard Education Letter*.
- Schmidt, W.H., and others. *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco: Jossey-Bass, 2001.
- Slavin, R. (1987). Ability grouping and student achievement in elementary schools: A best evidence synthesis. *Review of Educational Research*. Vol. 57, No. 3, pp. 293-336.
- Slavin, R. (1990). *Achievement effects of ability grouping in secondary schools: A best-evidence synthesis*. Madison, WI: National Center on Effective Secondary Schools. ED 322 565.
- Stigler, J.W., & Hiebert, J. *The teaching gap*. NY: Free Press, 1999.
- Thernstrom, A., & Thernstrom, S. (2003). *No excuses: Closing the racial learning gap*. NY: Simon & Schuster.
- University of Massachusetts Donahue Institute. (2002). *Analysis of student outcomes: Evaluation of the DOE Middle School Mathematics Initiative*. Report prepared for the Massachusetts Department of Education. Malden. MA.
- U.S. Department of Education. (October, 1997). Mathematics equals opportunity (white paper prepared for U.S. Secretary of Education Richard W. Riley). [On-line]. Available: ed.gov/pubs/math
- Whitehurst, G. (2003). Paper on research presented at the Mathematics Summit in Washington, DC on February 6. [On-line]. Available: <http://www.ed.gov/rschstat/research/progs/mathscience/whitehurst.html>

APPENDIX: Schools in the Study***Small District Schools**

Paul R. Baird Middle School
 Boston Renaissance Charter School
 Cyril K. Brennan Middle School
 Carlisle School
 Joseph Case Junior High School **
 Jonas Clarke Middle School
 Clinton Middle School
 Silvio O. Conte Middle School
 Great Falls Middle School
 Hanover Middle School **
 Hastings Middle School **
 Lincoln School
 Locke Middle School **
 Marston Mills Middle School
 Mashpee High School **
 Medway Middle School
 Milford Middle School East
 Nauset Regional Middle School **
 Nessacus Regional Middle School **
 Rupert A. Nock Middle School
 North Brookfield High School
 Oak Ridge School
 O'Donnell Middle School
 Plymouth Community Intermediate School **
 Samoset School **
 Wamsutta Middle School **
 Wellesley Middle School
 West Springfield Middle School
 Laura A. White Middle School
 Whitman Middle School

Large District Schools

Boston Latin School
 Boston Latin Academy
 Charles E. Brown Middle School
 Central Middle School
 Chestnut Street Middle School
 F. A. Day Middle School
 East Somerville Community School **
 Edward Devotion School
 Robert Frost School
 Forest Grove Middle School
 John F. Kennedy School
 M. Marcus Kiley School
 Henry Lord Middle School
 Morton Middle School **
 Mountview Middle School **
 James L. Mulcahey School **
 North Junior High School **
 O'Bryant Math-Science School
 Henry K. Oliver School
 John F. Parker Middle School
 Dr. William R. Peck Middle School
 Pickering Middle School **
 Thomas Prince School
 E. N. Rogers School **
 Roosevelt Junior High School **
 South Lawrence East School
 Sullivan Middle School
 James P. Timilty Middle School **
 Umana-Barnes Middle School
 Phyllis Wheatley Middle School **

*For the follow-up analysis, Boston Latin School and Wellesley Middle School were removed from the non-IPAW group because both had such low percentages of students at the lowest level in 1998 that they could not have reduced this category more than the state average by 2002. Cyril K Brennan Middle School was moved from the non-IPAW group to the IPAW group because its percent of improvement at the Proficient, Advanced, and Warning levels was above the state average, however slight. Luther Burbank Middle School in the Nashoba Regional School District was added to the IPAW group because its teacher data were complete; it had not been used in the original study because administrator data were missing.

** IPAW schools are indicated by a double asterisk after the school name.