Relationships Between Student Performance on the MCAS (Massachusetts Comprehensive Assessment System) and Other Tests—Collaborating District A, Grades 4 and 10

Prepared for the Massachusetts Department of Education
March, 1999

By
Brian Gong
The National Center for the Improvement of Educational Assessment, Inc.
P.O. Box 491
Dover, NH 03821-0491

Acknowledgements

The Massachusetts Department of Education funded this study by the National Center for the Improvement of Educational Assessment, Inc. through a contract with Advanced Systems in Measurement and Evaluation, Inc. Data were generously provided by the collaborating district, Worcester Public Schools, and Advanced Systems. This study would not have been possible without the very helpful support of Gerri Williamson, Patricia Mostue, and Jim Caradonio of the district offices of the Worcester Public Schools. Kevin Sweeney of Advanced Systems provided information about the MCAS data files. Most notably, this work was done in collaboration with Gene Hoffman and his colleagues at HumRRO. The design of the analyses and the text of the report drew heavily on validity work previously conducted by them[1]. Their thoughtful advice made this work possible. Of course, any remaining shortcomings are the responsibility of the author. The positions expressed in this paper are those of the author, and do not necessarily represent the views of Advanced Systems, the collaborating district, or the Massachusetts Department of Education.

[1] C.f. Hoffman, G. R. & Tannen, M. B. (Aug. 1998). *Relationships between Kentucky's open-response scores for eighth-grade students and their CTBS-5 scores as ninth-grade students*. (HumRRO Report No. FR-WATSD-98-30 and OCAA Occasional Paper 98-5) Frankfort, KY: Kentucky Department of Education. (Available from the Kentucky Department of Education, Office of Assessment and Accountability, 500 Mero St., Frankfort, KY 40601, or from HumRRO, 295 W. Lincoln Trail Blvd., Radcliff, KY 40160.)

Relationships Between Student Performance on the MCAS (Massachusetts Comprehensive Assessment System) and Other Tests—Collaborating District A, Grades 4 and 10

Table of Contents

List of Tables and Figures

Executive Summary

This summary was written to encompass important results from the companion reports by Gong (1999)[1] and Thacker and Hoffman (1999)[2]. While the two projects were completed for different Massachusetts' school districts, the similarity of methodology and purpose made combining the results in common summaries prudent. The executive summary is arranged around five distinct topics addressed by both reports. They include;

- MCAS standards and expectations for student performance,
- The relationship between MCAS and commercial standardized tests,
- MCAS and gender and ethnicity issues,
- MCAS and student transience, and
- The relationship between course-taking patterns and MCAS scores.

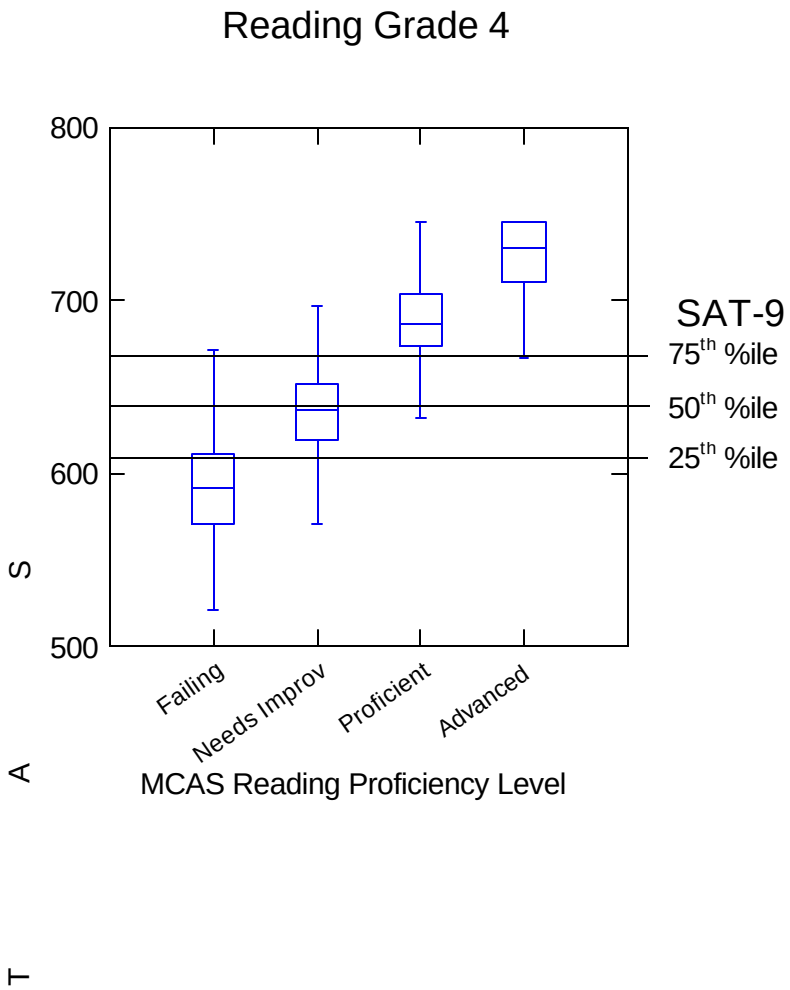
Each section of this executive summary addresses one of these concerns. The executive summary also appears as a separate document entitled *Relationships Between MCAS (Massachusetts Comprehensive Assessment System) and Commercial Standardized Tests for Two Collaborating Districts: A Summary of Five Important Issues*. The executive summary was designed for use in presenting sections of the findings included in the technical reports to interested parties. As such, each section of the executive summary follows a "stand-alone" format and contains its own bibliography.

[1] Gong, B. (1999). Relationships between student performance on the MCAS (Massachusetts Comprehensive Assessment System) and other tests – collaborating District A, grades 4 and 10. Report submitted to the Massachusetts Department of Education and Advanced Systems in Measurement and Evaluation. Dover, NH: The National Center for the Improvement of Educational Assessment, Inc.

[2] Thacker, A. A. & Hoffman, R. G. (1999). Relationships between MCAS and SAT-9 for one district in Massachusetts. Report submitted to Advanced Systems in Measurement and Evaluation and the Massachusetts Department of Education. Alexandria, VA: HumRRO.

The Challenge of MCAS

The Massachusetts Comprehensive Assessment System (MCAS) is a challenging assessment that demonstrates high standards for student achievement. Performance to these high standards was validated by strongly related performance on other tests. Two studies comparing MCAS with commercial standardized tests were conducted in two Massachusetts school districts. Results from those studies indicate that students in each of the MCAS proficiency levels (*Failing*, *Needs Improvement*, *Proficient*, and *Advanced*) generally performed similarly on a commercial standardized test. Students who scored *Proficient* or *Advanced* on MCAS tended to score above the 75th percentile on the standardized tests. Students who scored lower on MCAS also scored lower on the other tests. The following graph and table present some typical results. (Note: MCAS English language arts is called "reading" in this graph.)



Standardized Test Scores for Each MCAS Proficiency Level, Grade 4 Reading/English Language Arts, District B

Grade 4 Cohort, SAT-9 Scores by MCAS Proficiency Level						
MCAS Proficiency Level	SAT-9 Reading Scale Scores				Percentile of the Mean	N
	Mean	S.D.	1st Quartile	3rd Quartile		
English language arts						
Failing	591.3	26.5	571	611	12	1,402
Needs Improv.	636.9	28.1	619	652	49	2,272
Proficient	690.4	28.4	674	704	88	160
Advanced	719.2	31.0	711	745	97	9

The above graph represents the distribution of SAT-9 (Stanford Achievement Test, one of the commercial tests used in these studies) scores associated with performance at the *Failing, Needs Improvement, Proficient,* and *Advanced* MCAS proficiency levels. The boxes represent the middle 50% of students at each proficiency level. The whiskers represent the dispersion of students at the indicated proficiency level. The stair-step nature of the proficiency levels and the separation of them in relation to those same students' scores on the commercial standardized tests indicate that MCAS differentiated well between students of varying performance levels on the commercial test.

SAT-9 percentile scores are also overlaid on the above graph. These percentile scores show the challenge of MCAS. Students in the *Failing* category on MCAS averaged consistently below the 25th percentile on SAT-9. Students in the *Needs Improvement* category were clustered around the 50th percentile. Students in the *Proficient* and *Advanced* MCAS categories were typically above the 75th percentile on SAT-9.

The graph depicts fourth-grade reading scores (English Language Arts on MCAS). Eighth- and tenth-grade scores were also included in the original studies as well as scores in mathematics and science and technology. In addition to SAT-9, the MAT-7 (Metropolitan Achievement Test), another standardized commercial test, was used for comparison. The data from all tested grades and subjects were strikingly similar to the graph presented here, irrespective of the commercial test used for comparison. The one notable exception was found in comparisons of MCAS eighth-grade English language arts scores to the standardized tests. The grade 8 ELA boxplots showed a similar pattern to the others; however the numbers of students who scored in the higher categories on MCAS were greater than for the other tested grades and subjects. These data indicate that MCAS standards for eighth-grade reading, while challenging, might not be as high as for other grades and subjects.

The table presents similar data for the same students as are represented in the graph. N represents the number of students in each MCAS proficiency category. The mean SAT-9 score for students in each MCAS proficiency category is given along with

the percentile ranking of that mean score. The table also contains standard deviations for each category (S.D.) and the minimum and maximum SAT-9 scores posted by students within each category. The average percentile ranking for *Failing* students was 12. *Needs Improvement* students average score was at the 49th percentile. *Proficient* students were at the 88th percentile, and *Advanced* students were at the 97th percentile.

Most students, typically about 85% from the two districts participating in these studies, were in the *Failing* and *Needs Improvement* categories. Only nine were in the *Advanced* category in the sample data presented here. As schools become more adept at meeting the instructional challenges represented by the Massachusetts Curriculum Frameworks, those scores should improve. These initial studies are a strong indication of the high standards of student performance represented by MCAS.

While performance on the commercial tests was related to performance on MCAS, MCAS is specifically designed to measure the Massachusetts Curriculum Frameworks, yield reliable information for school and student accountability, and be useful as an indication of school improvement over time. It is important to remember, however, that these studies constitute only an initial step in determining and monitoring the validity of the MCAS testing system. These studies represent only two districts in Massachusetts and only a single point in time. In order to ensure the validity, reliability, and utility of MCAS, now and in the future, further research should be conducted. Possible next steps include:

- Extend the scope of these studies to include a statewide sample (possibly by comparing MCAS with the ITBS (Iowa Test of Basic Skills) in the elementary grades).
- Examine the data at the school level in addition to the district level.
- Perform consequential validity studies to determine the degree to which setting standards and testing students impacts classroom instruction.
- Examine the relationships between classroom instruction and MCAS test scores.

This document summarizes one aspect of the two research projects. Other summaries are available regarding analyses of correlations between MCAS scores and commercial tests, gender and ethnicity/race effects, student transience, and influence of course-taking patterns. Please refer to the full reports for clarification of any technical issues or for a more thorough version of the findings presented here (Gong, 1999; Thacker & Hoffman, 1999).

Bibliography

Gong, B. (1999). *Relationships between student performance on the MCAS (Massachusetts Comprehensive Assessment System) and other tests—collaborating district A.* Prepared for the Massachusetts Department of Education. The National Center for the Improvement of Educational Assessment, Inc., Dover, NH.

Thacker, A. A. & Hoffman, R. G. (1999). *Relationships between MCAS and SAT-9 for one district in Massachusetts.* Prepared for the Massachusetts Department of Education. Human Resources Research Organization, Alexandria, VA.

Performance on MCAS and Commercial Standardized Tests

Recent studies in two school districts found that student performance on MCAS (Massachusetts Comprehensive Assessment System) was appropriately related to student performance on commercial standardized tests. These studies are a good initial indication that MCAS is a strong measure of student performance in English language arts, mathematics, and science and technology. The strength of the comparisons is also a good indication that students took their performance on MCAS seriously.

The studies used correlations between students' MCAS scores and those same students' scores on either the SAT-9 (Stanford Achievement Test) or the MAT-7 (Metropolitan Achievement Test) to draw their conclusions. A sample table of these correlations is presented below.

Correlations between MCAS and MAT-7 scores, District A, Grade 10

MCAS Subject Test	**MAT-7 Subject Test**				
	Reading	**Language**	**Composition**	**Math**	**Science**
ELA	0.72	0.68	0.61	0.67	0.61
Math	0.66	0.66	0.59	0.81	0.65
Sci. & Tech.	0.72	0.64	0.59	0.72	0.71

The table presents correlations between MCAS subject tests and subject tests from the MAT-7. A correlation of 0.00 would indicate that the two measures were not measurably related, while a correlation of 1.00 would indicate that the two measures were perfectly related. The underlined correlations represent correlations between tests of similar subject material, e.g., MCAS English language arts and MAT-7 reading. Other correlation tables comparing MCAS to the SAT-9 for the other district, as well as correlation tables for other tested grades (4 and 8), were all very similar to this example.

It is important to remember that any single correlation presented in the table is less revealing than the pattern of correlations across the table. What stands out most strikingly is that the correlations between the two tests in all tested subjects are relatively strong. This indicates that there was a tendency for students who performed well on one section of one test to perform well on all sections of both tests. Strong mathematics students tended to also be strong language arts and science students. This pattern was repeated irrespective of the test compared with MCAS in both districts studied.

The stronger correlations tended to be between like subject areas. The strongest correlations were between MCAS mathematics and commercial standardized measures of mathematics. The implication of the pattern is clear. MCAS, MAT-7, and SAT-9 all show a good deal of similarity in assessing student's academic achievements. The strength of these correlations is a reassuring indication that the test is strong and that students tended to take their performance on it seriously.

When comparing MCAS with other tests, how high should the correlations be? The answer to that question is relatively ambiguous and requires an examination of the purpose of MCAS. Traditional explorations of validity involving correlations between two similar content tests sought high correlations as assurance that the tests were measuring the same thing. MCAS was not designed to measure exactly the same thing in

the same way that the commercial tests were designed to measure and the test administration and stakes for the students differ somewhat. On the other hand, MCAS was designed to measure student achievement in English language arts, mathematics, and science and technology, so we can't expect the tests to be unrelated. We are left with a criterion that the correlations should not be either too high or too low, or what Hoffman (1998) refers to as a "Goldilocks" criterion. Exactly where the "too high" or "too low" mark is depends on the degree of difference between the purposes of the MCAS and the purposes of the commercial standardized test with which it is being compared. The correlations reported seem to be within this "Goldilocks" range given the stated purpose of the MCAS.

It is important to remember, however, that these studies represent only an initial step in determining and monitoring the validity of the MCAS testing system. These studies represent only two districts in Massachusetts and only a single point in time. In order to ensure the validity, reliability, and utility of MCAS now and in the future, further research should be conducted. Possible next steps include:

- Extend the range of these studies to a statewide sample.
- Repeat these and similar studies in each subsequent year of testing to monitor changes in the correlations associated with tailored instruction or other factors.
- Link school-level factors with MCAS test scores.
- Link teacher practice and teacher professional development with student performance on the MCAS.
- Examine student factors that contribute to achievement in relation to MCAS scores.
- Examine the differences between the multiple choice, short answer, and written response sections of MCAS more closely in reference to learning and teaching.

This document summarizes one aspect of the two research projects. Other summaries are available regarding analysis of MCAS standards, gender and ethnicity issues, student transience, and influence of course taking-patterns. Please refer to the full reports for clarification of any technical issues or for a more thorough version of the findings presented here (Gong, 1999; Thacker & Hoffman, 1999).

Bibliography

Gong, B. (1999). *Relationships between student performance on the MCAS (Massachusetts Comprehensive Assessment System) and other tests—collaborating district A.* Prepared for the Massachusetts Department of Education. The National Center for the Improvement of Educational Assessment, Inc., Dover, NH.

Hoffman, R. G. (1998). *Relationships among KIRIS open-response assessments, ACT scores, and students' self-reported high school grades* (HumRRO Report, FR-WATSD-98-27). Alexandria, VA: Human Resources Research Organization.

Thacker, A. A. & Hoffman, R. G. (1999). *Relationships between MCAS and SAT-9 for one district in Massachusetts.* Prepared for the Massachusetts Department of Education. Human Resources Research Organization, Alexandria, VA.

Gender, Ethnicity, and MCAS

The differences in academic performance of students of different genders and ethnic groups have long been a concern of educators and policymakers. Differential performance on large-scale assessments has been well documented. Students in the 4th, 8th, and 10th grades took the MCAS (Massachusetts Comprehensive Assessment System) test for the first time during the 1997-8 academic year. Students in two Massachusetts school districts also took commercial standardized tests, either the SAT-9 (Stanford Achievement Test) or the MAT-7 (Metropolitan Achievement Test). It is possible to examine the results of those tests for differences in student scores associated with gender and ethnicity.

First it should be stated that differences in test performance tend to exist for a variety of reasons. The important question with regard to MCAS is not if male and female students or Hispanic and African-American students had the same average score. Instead, the important question is whether some aspect of the test itself increases those differences. A key indicator, studied in two recent research projects, is whether differences between gender or racial subgroups is greater on MCAS than would be expected based on the results of other tests.

As might be expected, both studies showed differences between male and female performance on MCAS as well as SAT-9 and MAT-7. The differences between males' and females' MCAS scores tended to be minor in both studies and followed the same pattern as scores on the commercial standardized tests. They followed stereotypical patterns, with males tending to perform slightly better than females on the mathematics and science portions of all tests and females performing slightly better than males on the reading and writing portions of the tests. Statistical analysis of the results showed that MCAS is essentially equivalent to the commercial standardized tests in terms of gender differences.

Differences in scores for various ethnic groups were also examined. Unlike gender, large differences in mean scores exist between the various ethnic groups on all three studied tests. Typically, White/Caucasian students posted the highest scores, followed by Asian/Pacific Islanders, African Americans, and lastly Hispanic/Latino students. A wide variety of reasons exist that may help explain these results, including socioeconomic issues and students' opportunities to learn. However, they follow a pattern similar to students in other states (Hoffman, 1998). The larger technical reports from which this summary was written elaborate on some of the possible factors influencing these results.

The issue examined during this research was not primarily an evaluation of the differences that exist between mean MCAS scores from various ethnic groups, however. The primary concern was determining if the differences between the ethnic groups was larger than would be expected given those students scores on the commercial standardized tests. Statistical analysis was used to calculate the expected differences between ethnic groups on MCAS from those same students' scores on the commercial standardized tests. In both districts studied, statistical results indicate that MCAS is similar to the other tests with regard to differences between ethnic groups, but not exactly the same.

The existence of differences is not necessarily an indication of bias in MCAS. Differences on the test may actually reflect differences learning due to different opportunities to learn. One of the studies (Gong, 1999) referenced suggests that there is a considerable amount of difference in the course taking patterns and success rates of the various ethnic groups studied. Those types of student factors may help account for these results.

It is also important to remember that these studies represent only an initial step in determining and monitoring the validity of the MCAS testing system. These studies represent only two districts in Massachusetts and only a single point in time. In order to ensure the validity, reliability, utility, and fairness of MCAS, now and in the future, further research should be conducted. Possible next steps include:

- Extend the range of these studies to a statewide sample.
- Evaluate these studies at the school level. Studying only district-level reports may mask important school-level differences in gender and ethnicity.
- Examine student factors that contribute to MCAS scores that may help explain differences in achievement for various ethnic groups.
- Continue to evaluate gender and ethnic differences on MCAS as the program continues over the years.
- Monitor the progress of historically lower-scoring ethnic groups as the program continues.

This document summarizes one aspect of the two research projects. Other summaries are available regarding analyses of correlations between MCAS scores and commercial tests, MCAS standards, student transience, and influence of course-taking patterns. Please refer to the full reports for clarification of any technical issues or for a more thorough version of the findings presented here (Gong, 1999; Thacker & Hoffman, 1999).

Bibliography

Gong, B. (1999). *Relationships between student performance on the MCAS (Massachusetts Comprehensive Assessment System) and other tests—collaborating district A.* Prepared for the Massachusetts Department of Education. The National Center for the Improvement of Educational Assessment, Inc., Dover, NH.

Hoffman, R. G. (1998). *Relationships among KIRIS open-response assessments, ACT scores, and students' self-reported high school grades* (HumRRO Report, FR-WATSD-98-27). Alexandria, VA: Human Resources Research Organization.

Thacker, A. A. & Hoffman, R. G. (1999). *Relationships between MCAS and SAT-9 for one district in Massachusetts.* Prepared for the Massachusetts Department of Education. Human Resources Research Organization, Alexandria, VA.

Student Transience and MCAS

The MCAS (Massachusetts Comprehensive Assessment System) first tested students in the 4th, 8th, and 10th grades during the 1997-98 academic year. A questionnaire was included with the test that asked students about the number of years they had attended their current school and district. While this information was not sufficiently precise to allow the construction of student transience rates, it does serve as an indicator (Medsker, 1998). Typically, students who change schools often do not perform as well academically as students who regularly attend a single school or school system.

Research conducted in two Massachusetts school districts examined student transience in relation to MCAS test scores as well as to scores on other commercial standardized tests. Students from those districts took the MCAS and either the SAT-9 (Stanford Achievement Test) or the MAT-7 (Metropolitan Achievement Test). This research evaluated how time in a single school or district related to test scores and, perhaps more importantly, it evaluated whether student transience related to MCAS scores differently than to commercial standardized test scores.

Not surprisingly, researchers found that students who spent more time in a single school or district tended to have higher test scores on all three tests. Transience shows considerable congruence with socioeconomic status, which is a well-publicized predictor of test scores. What was surprising was that this relationship was not linear in nature and was somewhat different for the two school districts.

In one district, fourth grade students who reported having attended the school or district less than one year scored significantly higher than their peers who had been in the school or district for longer (Thacker & Hoffman, 1999). This district was among the lowest scoring districts in the state, which might help to account for this anomaly. This trend was also noted in the other district included in the study, but to a lesser degree (Gong, 1999). That district's mean MCAS scores were close to the state average. However, if the fourth-grade students who reported attending the school less than one year are omitted from either district, those who reported coming into the school in the first, second, or third grade show the relationship that would be expected. The more time spent in the school or district, the higher the test scores.

Students from the low-scoring district who reported entering the school or district in the seventh grade had significantly higher scores than their peers on all three tests. Also, more students reported moving into the school or district in the seventh grade than in the fifth, sixth, eighth, or ninth grade. This could indicate that there is a considerable influx of students into the public school system at the seventh grade from private and parochial schools.

An important issue with regard to student transience from both studies extends beyond the question of whether transience was related to test scores. These studies were conducted to evaluate the validity of MCAS, and as such, they examine the extent to which MCAS is similar to commercial standardized tests in its relationship to transience. Statistical analysis showed little difference in the relationship of transience and MCAS scores versus SAT-9 and MAT-7 scores. This indicates that while transience is related to MCAS test scores, the relationship is very similar to other tests.

There are several possible explanations for the anomalous results from the two districts and the somewhat curious non-linear nature of the relationship between

transience and student test scores. These results may reflect problems with the student self-reports, the wording of the questionnaire, patterns of movement between private/parochial schools and the public schools, immigration, or other factors. Both districts studied showed reasonably high levels of mobility, which might not be typical throughout the state. Any combination of these factors might help account for the curious anomalies found in the data.

One of the most important aspects of this initial set of transience studies to remember is that it is an initial examination of the issue. These studies come at a single point in time at the beginning of the implementation of MCAS. MCAS and the commercial standardized tests are very different in form and purpose. MCAS is a standards-based assessment, and as such, it is potentially subject to an increasing influence from student transience. As schools and districts become more focused on helping students achieve the specific goals outlined in the Massachusetts Curriculum Frameworks, students coming into the system from states with either dissimilar or unspecified standards might be at a considerable disadvantage. Clearly, studies similar to these should be repeated as schools and districts become more adept at helping students meet the standards.

It is also important to remember that these studies represent only two districts from the state and that all of the analyses related to transience relied on student self-reports. In order to ensure the validity, reliability, utility, and fairness of MCAS, now and in the future, further research should be conducted. Possible next steps include:

- Re-evaluate these results using district enrollment records in order to eliminate any doubts about the accuracy of the student self-reports.
- Extend the range of these studies to a statewide sample.
- Examine the effects of transience on individual schools within districts.
- Examine the effects of transience in subsequent years as schools and school systems become more familiar with MCAS and the standards tested by MCAS.
- Identify exemplary schools with high or improving MCAS scores and high transience rates and perform case studies at them in order to assist similar schools.

This document summarizes one aspect of the two research projects. Other summaries are available regarding analyses of correlations between MCAS scores and commercial tests, MCAS standards, gender and race/ethnicity effects, and influence of course-taking patterns. Please refer to the full reports for clarification of any technical issues or for a more thorough version of the findings presented here (Gong, 1999; Thacker & Hoffman, 1999).

Bibliography

Gong, B. (1999). *Relationships between student performance on the MCAS (Massachusetts Comprehensive Assessment System) and other tests—collaborating district A.* Prepared for the Massachusetts Department of Education. The National Center for the Improvement of Educational Assessment, Inc., Dover, NH.

Medsker, G. J. (1998). *Determining the relationship between student transience and KIRIS school results: are schools with transient students unfairly impacted?* (HumRRO Report, FR-WATSD-98-12). Alexandria, VA: Human Resources Research Organization.

Thacker, A. A. & Hoffman, R. G. (1999). *Relationships between MCAS and SAT-9 for one district in Massachusetts.* Prepared for the Massachusetts Department of Education. Human Resources Research Organization, Alexandria, VA.

MCAS and Student Course-Taking Patterns

The MCAS (Massachusetts Comprehensive Assessment System) first tested students in the 4th, 8th, and 10th grades during the 1997-98 school year. An important question for schools, districts, and policymakers is, "Do the classes students attend have any influence on their test scores?" This question is not quite as easy to answer as might be suspected upon initial consideration. Students choose to take classes for a variety of reasons, and to examine the scores of students who choose to take this or that class in relation those who do not provides only an incomplete and perhaps a misleading representation of the relationship between course-taking patterns and test scores. This was certainly the case for the MCAS test scores.

Two studies examined student MCAS scores in relation to those same students' scores on commercial standardized tests, either the SAT-9 (Stanford Achievement Test) or the MAT-7 (Metropolitan Achievement Test). Two districts participated in these studies. Students who took the MCAS also completed a questionnaire that contained questions about the subjects they studied and/or the classes they attended. From this data researchers were able to compare the test scores of students in relation to the classes they reported taking. Course-taking data was constructed entirely from student self-reports.

Before considering the data linked to specific subjects and courses, it might be helpful to keep in mind how closely related students' scores were for all subjects tested on each of the three assessments mentioned. Academically talented students tend to perform well on all subjects irrespective of testing format. Strong mathematics students who were not also strong in science and English language arts were rare. It might not be surprising to discover that these academically talented students attended similar courses. It is impossible to know from this study whether the courses caused the students to be academically talented, or whether otherwise academically talented students simply favor taking certain courses.

Both studies found that there was a fairly strong relationship between the courses students reported taking and their test scores. The most obvious trend was in mathematics. Students who reported taking higher level mathematics outperformed their peers on all sections of all tests. The trend in science was similar, although not as pronounced. Fourth-grade students did not report the courses they had attended, but were asked to estimate the amount of time spent on science and mathematics. The more classroom time spent studying mathematics and science the better the scores in all subjects on MCAS and the commercial standardized tests.

Students were also asked about class time spent in non-assessed subjects. Students who took courses in foreign language, technology, health, and the arts tended to score better than students who did not for all tests and subjects. This may reflect the nature of elective course requirements. Students who perform well in mathematics, English language arts, and science may be given more freedom to select elective courses. These results may also reflect varying opportunities to attend some of these elective courses because of curriculum variations in specific schools within the district.

Analysis of social studies course taking was more problematic. Students who took more social studies classes tended to have lower scores in all tested subjects than those who did not. MCAS does not currently assess social studies; however, a social studies test is expected to be added in the 1998-99 academic year. The relationship between test scores and social studies classes may be very different when social studies is

specifically assessed. The close relationship among scores on other subjects however, indicates that a shift in the relationship between social studies classes and scores is unlikely. The data suggests that academically talented students may be bypassing social studies classes in favor of mathematics and science. This phenomenon has been reported in other states as well (Hoffman, 1998).

Course-taking patterns allowed researchers to examine some issues surrounding students' opportunities to learn as well. In one district, there were large differences in test performance between racial/ethnic groups, even those who took similar courses (Gong, 1999). Several schools within the second district studied were small and the ethnic proportions varied greatly from school to school (Thacker & Hoffman, 1999). The relationships between course taking, ethnicity, and test scores may be more appropriately examined at the school level of analysis.

It is also important to remember that these studies represent only two districts from the state and that all of the analyses related to course taking relied on student self-reports. In order to ensure the validity, reliability, utility, and fairness of MCAS, now and in the future, further research should be conducted. Possible next steps include:

- Re-evaluate these results using district enrollment records in order to eliminate any doubts about the accuracy of the student self-reports.
- Extend the range of these studies to a statewide sample.
- Examine the effects of course-taking patterns on individual schools within districts.
- Examine the effects of course-taking patterns and teacher practice in subsequent years as schools and school systems become more familiar with MCAS and the standards tested by MCAS.
- Research the link between ethnicity and course taking, both in terms of classes chosen and MCAS scores of students attending similar courses.

This document summarizes one aspect of the two research projects. Other summaries are available regarding analyses of correlations between MCAS scores and commercial tests, MCAS standards, gender and race/ethnicity effects, and student transience. Please refer to the full reports for clarification of any technical issues or for a more thorough version of the findings presented here (Gong, 1999; Thacker & Hoffman, 1999).

Bibliography

Gong, B. (1999). *Relationships between student performance on the MCAS (Massachusetts Comprehensive Assessment System) and other tests—collaborating district A.* Prepared for the Massachusetts Department of Education. The National Center for the Improvement of Educational Assessment, Inc. Dover, NH.

Hoffman, R. G. (1998). *Relationships among KIRIS open-response assessments, ACT scores, and students' self-reported high school grades* (HumRRO Report, FR-WATSD-98-27). Alexandria, VA: Human Resources Research Organization.

Thacker, A. A. & Hoffman, R. G. (1999). *Relationships between MCAS and SAT-9 for one district in Massachusetts.* Prepared for the Massachusetts Department of Education. Human Resources Research Organization, Alexandria, VA.

Relationships Between Student Performance on the MCAS (Massachusetts Comprehensive Assessment System) and Other Tests—Collaborating District A, Grades 4 and 10

Summary

This study analyzed some important aspects of the validity of the MCAS (Massachusetts Comprehensive Assessment System) state-mandated test. MCAS was administered for the first time in spring 1998 to all Massachusetts public school students in grades 4, 8, and 10. English language arts, mathematics, and science and technology were assessed. This study examined the relationships between student performance on the MCAS and performance on commercial, standardized tests that had been administered to the same students in one district. The commercial tests were the SAT9 for the cohort of students that took the MCAS in grade 4, and the MAT7 for the MCAS grade 10 students. In addition, MCAS results were compared with performance on the commercial tests to check for undue differences related to gender, ethnicity, and transience. Student achievement on the MCAS was also examined for relation to students' reports about studying specific topics or frequency of addressing a subject area in school.

Overall, these initial analyses support the view that the MCAS is a valid assessment.

- Student performance on the MCAS was related to student performance on familiar, commercial, norm-referenced tests. An appropriate, moderate relationship was found between student performance on the MCAS and achievement of the same students on the other tests.
- The MCAS proficiency levels represented notably high standards in relation to the norm-referenced tests.
- No undue differential effects were found for the MCAS in terms of gender and ethnicity. Consideration of gender or ethnicity did not appreciably affect the relationships between performance on MCAS and performance on the commercial tests. The similarities and differences in student achievement could not be attributable to consideration of gender or ethnicity.
- Student transience was related to student achievement on the MCAS as well as the commercial tests in general. However, some anomalous results indicated that this issue needs to be researched more extensively.
- There were moderately strong relationships between performance on the MCAS and students' reported course-taking patterns in Grade 10. Higher student performance in mathematics and science on the MCAS was related to students taking more advanced courses. Results indicate that MCAS is sensitive to students' opportunities to learn, and suggest that student MCAS scores will likely rise as students learn the curricula aligned with the Massachusetts curriculum frameworks.

These results, although based on a limited sample of students and districts, will provide a useful backdrop for more extensive validity studies in the future.

Relationships Between Student Performance on the MCAS (Massachusetts Comprehensive Assessment System) and Other Tests—Collaborating District A, Grades 4 and 10

Purpose of Study

This research was conducted to examine the validity of the Massachusetts Comprehensive Assessment System (MCAS). Current views of *validity* hold that an assessment is valid if the evidence indicates it consistently assesses a *construct*, and if the *consequences* of its use are consistent with its purpose. A system as complex and new as MCAS will require extensive examination of data to establish how valid it is. An important question at the beginning of an assessment is how it related to other, more familiar assessments in terms of relative performance and potential adverse impact.

For this project, validity will be examined by determining how MCAS is related to other indicators of educational achievement, using students' scores from selected school districts. The analyses will focus on:

1. Relationships between MCAS scores and commercial test results by subject and grade level;
2. Possible undue differential effects in MCAS scores related to gender, ethnicity, and transience;
3. Relationships between MCAS and standardized assessment scores to courses taken or topics studied in school.

A companion study was done concurrently that analyzed results in another Massachusetts district.[1] The executive summary associated with this report reflects results from both studies.

This study represents a first step in gathering information about the validity of MCAS and answering questions regarding the merits of using the assessment for school level accountability and as a student graduation requirement. This project begins to answer questions about the fairness, reliability, and merit of using MCAS as a school and student level assessment tool and as an instrument to foster of educational improvement. The report concludes with recommendations about future actions to establish the validity of MCAS.

---

[1] Thacker, A. A. & Hoffman, R. G. (1999). Relationships between MCAS and SAT-9 for one district in Massachusetts. Report submitted to Advanced Systems in Measurement and Evaluation and the Massachusetts Department of Education. Alexandria, VA: HumRRO.

Background

Description of MCAS

According to documents published by the Massachusetts Department of Education[1], the MCAS (Massachusetts Comprehensive Assessment System) is intended to provide a statewide test based on the common curriculum standards of the state's Massachusetts Curriculum Frameworks that will provide student, school, and district scores. By 2001 students will be required to pass the 10th grade test to be eligible for a high school diploma. The first administration of MCAS was in spring 1998, with scores reported in the fall of 1998.

MCAS is administered to students in grades 4, 8, and 10. In 1998 the tests included English language arts, mathematics, and science and technology. The tests include both multiple choice questions and questions requiring a written response. English language arts combined scores from reading and writing. The reading component was assessed by multiple choice items and questions requiring written responses a few paragraphs in length. The writing was assessed through a longer essay written by the student.

This study used two types of scores associated with the MCAS assessment: scale scores and proficiency level designations. MCAS scale scores range from 200 to 280. Students were assigned one of four proficiency levels, depending on their scale scores. Students scoring below 220 were given a label of *Failing*; at or above 220 but below 240 received a label of *Needs Improvement*; 240-259 were designated *Proficient*, and 260 and above received an *Advanced* label. Each student had a scale score and a proficiency level for English/Language arts, mathematics, and science and technology.

Study Design

The MCAS was first administered in spring 1998. For this study, the state department of education and the assessment contractor sought cooperating districts that had required administration of commercial, standardized tests that could be matched with the same students who had taken the MCAS. For this initial study, two districts were chosen that were relatively large, had a range of student demographics, and were willing to collaborate in the study. Most importantly, these districts had available commercial test data for the students who had taken the MCAS, and had assigned the students who took the MCAS a student identification number that could be matched with an identification number in the district's testing files. This common student identification key made it feasible to match the MCAS results with the district results. It was difficult to find districts that met all these criteria, and could provide the district commercial test data on

[1] *Q & A: The NEW Massachusetts Test for Students* and other documents that provide information about MCAS are available on the Massachusetts Department of Education website at www.doe.mass.edu, or from the Massachusetts Department of Education, 350 Main St., Malden, MA 02148.

the required time schedule. For these reasons, the data from two districts were analyzed, even though they were somewhat different from each other.

One district had administered one commercial test to the MCAS students close to the same time that the MCAS was administered in spring 1998. This district has substantial overlap with the same students, close time proximity, but has an unknown factor of student motivation and possible fatigue. The other district administered commercial tests at a variety of times, but none at the same time as the MCAS. This district was able to provide matched data for students one year prior to taking the MCAS in one grade, and one year following the MCAS for a different grade. Changes in student enrollment, motivation, and learning over time would affect the interpretation of these data. Examining performance of the same students over multiple years was determined to be unfeasible for this study given the available time and the fact that both districts had recently changed commercial tests.

The two districts selected had the test data show in Table 1.

Table 1. Collaborating Districts - Data Available for Study

	School Year Tested		
Grade	**1996-97**	**1997-98**	**1998-99**
3			
4		*MCAS* **SAT9**	
5			SAT9F
6			
7			
8		*MCAS* **SAT9**	
9	MAT7F		
10		*MCAS* **SAT9**	
11			

Key = *MCAS:* Massachusetts Comprehensive Assessment System, Spring 1998
Regular: District A (last letter indicates Fall or Spring administration)
Bold: District B (Math and Reading, MC Survey, Spring)

There were different data sets available for each district. The following data were chosen to include in the analyses:

1. 1998 MCAS Grade 4 cohort: District A (SAT9 from subsequent year)
   District B (SAT9)
2. 1998 MCAS Grade 8 cohort: District A (no data available)
   District B (SAT9)
3. 1998 MCAS Grade 10 cohort: District A (MAT7 from previous year)
   District B (SAT9)

Note that for District A, the commercial test results are actually for Grade 5 students who had taken the MCAS as fourth grade students the previous spring, and for Grade 9 students who subsequently took the MCAS as sophomores. The commercial test results refer to these as "Grade 4" and "Grade 10" cohort results, but actually use the Grade 5 and Grade 9 norms and other appropriate information for interpreting the students' performance on the commercial tests.

Collaborating District

District A, an urban district, is one of the larger districts in the state. It serves a diverse racial/ethnic population. Many of the students are economically disadvantaged. The district has a higher than average proportion of students who had limited English proficiency. The district has historically scored at or below the state average on other tests in the past.

Standardized Test Information

For District A, the study involved data from two commercial standardized tests: the MAT7 and the SAT9.

The MAT7 (Metropolitan Achievement Tests, 7th edition, 1992 norms, published by Harcourt Brace Educational Measurement) was administered by the district for several years to all students enrolled. Those cohorts that took the MCAS included students who took the MAT7 in grade 7 in fall 1996, in grade 8 fall 1997, and in grade 9 in fall 1996. The students took the complete MAT7 battery, which included reading, mathematics, science, composition, and language. These students are referenced in the paper as "Grade 10" cohort, although they were in Grade 9 when they took the commercial test.

The SAT9 (Stanford Achievement Tests, 9th edition, form SA, 1996** norms, published by Harcourt Brace) was administered by the district to all students enrolled in selected grades in the 1998-99 school year. Those cohorts that took the MCAS included students who took the SAT9 in grade 5 in fall 1998. The students took the complete SAT9 battery, which included reading, language arts, mathematics, and science. These students are referenced in the paper as "Grade 4" cohort, although they were in Grade 5 when they took the commercial test.

Data Files

MCAS Data

Advanced Systems in Measurement and Evaluation, Inc., the assessment contractor to the Massachusetts Department of Education, provided files of the MCAS data for this study. Advanced Systems also provided file layout descriptions. Advanced Systems was responsible for scanning the data, generating the data files, and ensuring their accuracy

and quality. Advanced Systems used the data files in their reports to the state, districts, schools, and individual students. The MCAS data provided for this study included student name, identification number, responses to the student questionnaire, and MCAS score information for English language arts, mathematics, and science and technology. The score information included standard scores and proficiency levels for each student.

For this study, a number of composite variables were generated from the variables in the data files. For example, each student wrote an extended essay, which was scored on the two dimensions of idea development and organization, and usage and mechanics. The scores on these two aspects were added together to form a single score for the writing essay. Similarly, other composites were generated to capture the responses from the questionnaire, where a student may have responded affirmatively to questions whether he/she had studied three science topics; a composite score for studying broadly was generated.

MCAS Student Questionnaire

Advanced Systems also provided with the data file the students' responses to the Student Questionnaire. The Student Questionnaire was administered as a non-scored part of the MCAS test, and consisted of approximately 20 multiple-choice questions dealing with ethnicity, transience, and patterns of studying specific topics or taking specific courses. The Student Questionnaire was included as the last pages in the student test booklet. Responses to the Student Questionnaire were voluntary and not verified for accuracy. The questionnaires were designed with substantial overlap, but customized with grade-appropriate questions. Copies of the student questionnaire for each grade level are included in the Appendix.

Commercial Standardized Test Data

The collaborating district provided data files of the commercial standardized test results for the students in each of the cohort years shown in Table 1. The district also provided file layout descriptions.

Quality Assurance of Data Files

The MCAS data files were examined to ensure the data were within permissible range, and that errors and missing data were appropriately handled. The Student Questionnaire data were analyzed for random responses. Although the test booklets had approximately 20 questions for the Student Questionnaire, the response booklet had space for 50 responses. Students who bubbled in responses to the "non-questions" were judged as not haven taken the test seriously, and were dropped from the analysis. Only one student was dropped for this reason in grade 4, and none in grade 10 in the district.

The Birthdate in the MCAS file was unusable as a merge key with the district data, apparently because many students did not pay attention to the requirements that responses

be bubbled for day, month, and four-digit year. There were no other apparent problems with the MCAS data file.

The district data files had many student names and some birthdates that did not follow the expected conventions. However, the score data appeared accurate and usable, except for missing values.

Merging MCAS and Standardized Test Data Files

The MCAS and district data files were combined through matching the district-assigned student identification number, and then through matching names. A total of about 74% of the students were matched from the district to the MCAS files for Grades 9 and 10, and 80% between Grades 4 and 5. (Because District A did not administer commercial tests to students the same year they took the MCAS, it was necessary to match students' MCAS performance with their commercial test performance from years previous or subsequent to their taking the MCAS.) The successful match rates indicate some of the transience in the district. More students were matched between the spring MCAS in Grade 4 and the subsequent fall SAT9 administration in Grade 5, less than six months later. Fewer students were matched between the fall MAT7 in Grade 9 and the spring MCAS in Grade 10, about 18 months apart.

The student identification numbers were assigned by the district for use in a longitudinal tracking database by the district. The district had assigned the same identification numbers to students taking the MCAS. The students would not match on the identification number if the code were enter incorrectly on the MCAS form, or if the student had moved into or out of the district and had not taken both tests.

The MCAS and district commercial test files were first merged using the student identification number and last name. The large majority of students were confirmed as having the same last name and student number from the two data files in Grade 4, but a much smaller percentage matched in Grade 10. In about 10% of the cases the student name in the MCAS file differed from that in the district file. Visual inspection of these cases resolved almost all as probably being the same student. Common situations were transpositions (SMITH JOHN in one file, JOHN SMITH in the other), extra spaces (SM ITH vs. SMITH), misspellings (SMIHT vs. SMITH), miscodings (SMITHR OB vs. SMITH ROB) and different conventions (e.g., SMITHJR in one file, SMITH JR in the other). The student identification number proved to be more reliable than the name alone.

Those students who could not be matched by student number were then compared by last and first name. Students were included who matched on the last name (up to 11 characters), and on the first name (up to six characters).

Ambiguous cases were removed. These included cases where two students had the same student identification number but different first and last names, and where two students

had the same first and last names and no student identification number. Fewer than 10 cases were eliminated because they could not be positively resolved.

The total numbers of students included in the study are shown in Table 2.

Table 2: Numbers of Students Retained in Analyses

MCAS	District A
Total in MCAS file (state) Grade 4 74,539 Grade 10 61,956	[Not applicable at this level]
Total in MCAS file for district Grade 4 1949 (2.6% of state) Grade 10 928	Total in district file (Grade 5) 1772 (Grade 9) 1000
MCAS after missing student IDs Grade 4 1909 Grade 10 915	District after missing student IDs (Grade 5) 1763 (Grade 9) 900
Merge In both MCAS and District test files with matching student ID Grade 4 1178 Grade 10 496	
After adding students who matched first, last names Grade 4 1565 Grade 10 689	
After eliminating ambiguous cases Grade 4 1553 (79.7% of original MCAS file, 90.1% of original district file) Grade 10 686 (73.9% of original MCAS file, 68.9% of original district file)	

Procedures for Data Analysis and Results

Matched Sample from District A as a Representative Sample

The matched sample from District A appears to be sufficiently representative of the state and district to inform this study. The matched sample is proportionally lower than the state average on each MCAS content area test. The matched sample overall results are a few points higher than those of the district as a whole on both the MCAS and the commercial test. The differences are not large, and are consistent.

Table 3 presents Grade 4 means, standard deviations, and numbers of students for the MCAS and SAT9 test scores for three groups: the students that could be matched with both MCAS and district commercial (SAT9) data; the district as a whole; and the state as a whole. For the state as a whole, only the MCAS results are reported since districts were

not required to administer the same commercial test. Similar statistics are reported for Grade 10 in Table 4.

Table 3: Descriptive Statistics for MCAS and District SAT9, Grade 4

MCAS 1998	**Matched**			**District**			**State***
Grade 4	Mean	S.D.	N	Mean	S.D.	N	Mean
SAT9							
Reading	634.5	42.4	1493	631.9	42.7	1696	NA
Language	601.9	38.1	1548	599.9	38.2	1762	NA
Composition	607.2	44.0	1541	605.2	44.1	1754	NA
Mathematics	621.9	36.7	1545	620.3	37.0	1756	NA
Science	631.7	35.3	1541	629.9	35.9	1760	NA
MCAS							
English LA	227.7	10.2	1553	227.2	10.3	1948	230
Mathematics	231.6	15.7	1552	231.2	15.8	1947	234
Sci. & Tech.	236.0	13.5	1553	235.3	13.9	1947	238
Writing (raw)	10.2	3.3	1546	10.1	3.3	1931	-

*From the *Report of 1998 Statewide Results: The Massachusetts Comprehensive Assessment System (MCAS),* November 1998. Malden, MA: Massachusetts Department of Education. Available on the department website: www.doe.mass.edu.

The matched sample is quite similar to the district on the MCAS. On the SAT9, the matched sample is about 2-3 standard score points higher than the full district, indicating that proportionally more lower-scoring students were not matched. The differences, however, are relatively small and consistent across content areas.

The district is modestly below the state average on each of the MCAS areas. In English language arts, mathematics, and science and technology, the district was about three scale score points lower than the state average, or about one-third to one-fifth of a standard deviation for the district. The gap was less for the matched sample.

All results in this report are for the matched sample unless otherwise noted.

Table 4 : Descriptive Statistics for MCAS and District A MAT7, Grade 10

MCAS 1998 Grade 10	Matched			District			State
	Mean	S.D.	N	Mean	S.D.	N	Mean
MAT7							
Reading	667.5	41.4	682	659.0	43.7	994	NA
Language	643.3	36.2	671	636.9	38.2	942	NA
Composition	641.2	45.4	678	633.5	47.5	965	NA
Mathematics	654.7	36.0	682	649.0	36.5	977	NA
Science	645.5	32.2	667	640.2	34.0	936	NA
MCAS							
English LA	229.7	16.5	682	225.8	17.8	909	230
Mathematics	219.2	18.3	683	216.1	18.1	910	222
Sci. & Tech.	224.5	14.4	683	221.2	15.1	910	225
Writing (raw)	12.1	3.2	644	11.6	3.4	808	-

Another view of the district performance in relation to the state is provided by examining percentages of students in each of the four MCAS proficiency levels. Table 5 provides the results for the district A and the state. District A is generally lower than the state. The District A has 5-10% more of its students in the *Failing* and *Needs Improvement* levels than does the state, and the district also has proportionally fewer students at the *Advanced* level. The gaps, however, are not extreme.

Table 5: Percentages of Students by MCAS Performance Levels, Grades 4 and 10, District A and State

	Advanced		Proficient		Needs Imp.		Failing	
Grade 4	State	Dist.	State	Dist.	State	Dist.	State	Dist.
ELA	1	0	19	12	66	67	15	21
Math	11	7	23	20	44	49	23	23
Sci. & Tech.	6	4	42	39	40	45	12	12
Grade 10								
ELA	5	3	33	29	34	43	28	25
Math	7	3	17	15	24	26	52	57
Sci. & Tech.	1	1	21	16	42	48	36	35

The relative performances between the district and state are graphed in Figure 1 for Grade 4, and in Figure 2 for Grade 10. The most striking result is not any disparity between district and state. The graphs show how high the MCAS performance standards are. There were few students who met the state student proficiency goals of *Proficient* or *Advanced*—fewer than 50% in Science and Technology, fewer than 40% in mathematics, and fewer than 20% in English language arts in Grade 4. The patterns of relative

performance between the district and state are similar for Grade 10, although Grade 10 is strongest in English language arts.

Figure 1: MCAS Performance Level Results, State and District A, Grade 4

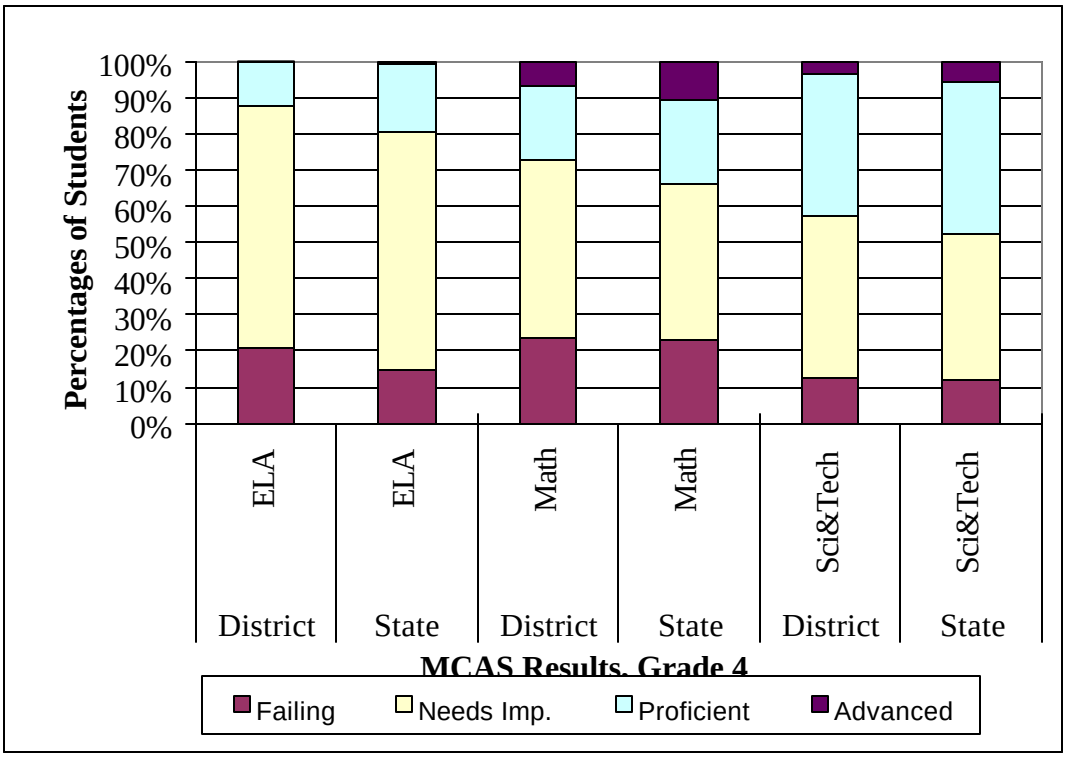
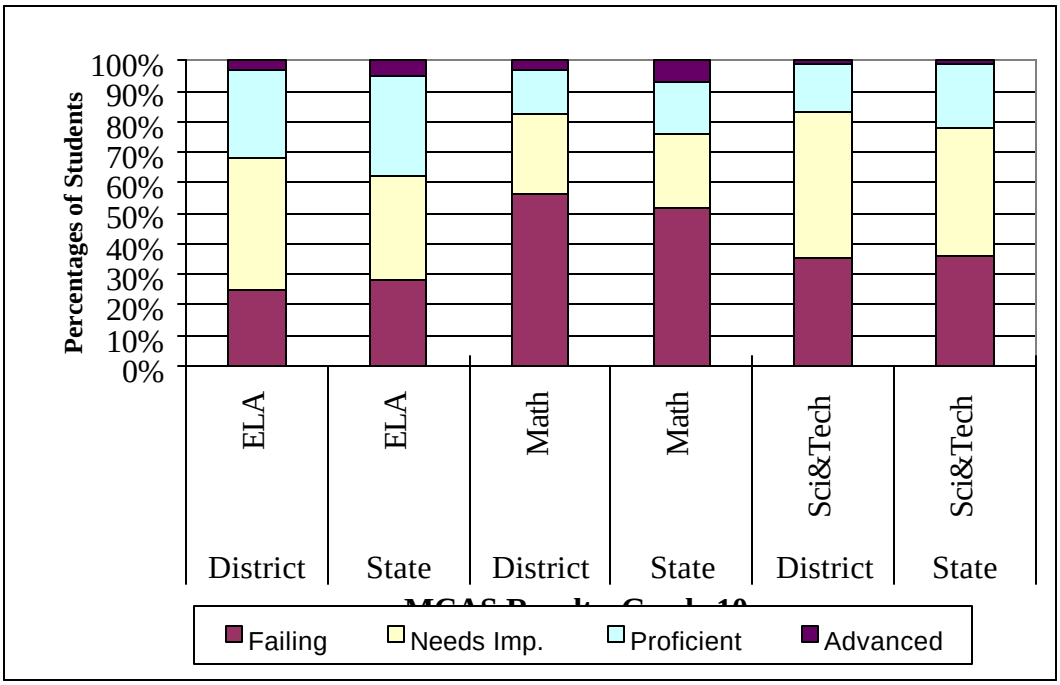


Figure 2: MCAS Performance Level Results, State and District A, Grade 10



Relationships Between Student MCAS Proficiency Levels and Commercial Test Scores

There are no comparable standards on the commercial tests to the proficiency categories set by Massachusetts educators for MCAS. However, it is possible to compare the relative performance to see whether higher performance on one measure is related to higher performance on the other test.

Student performance on the MCAS was related to student performance on familiar, commercial, norm-referenced tests in both Grade 4 and Grade 10. An appropriate, moderate relationship was found between student performance on the MCAS and achievement of the same students on the other tests. The MCAS proficiency levels represented notably high standards in relation to the norm-referenced tests.

The detailed results are presented first for Grade 4 and then more briefly for Grade 10.

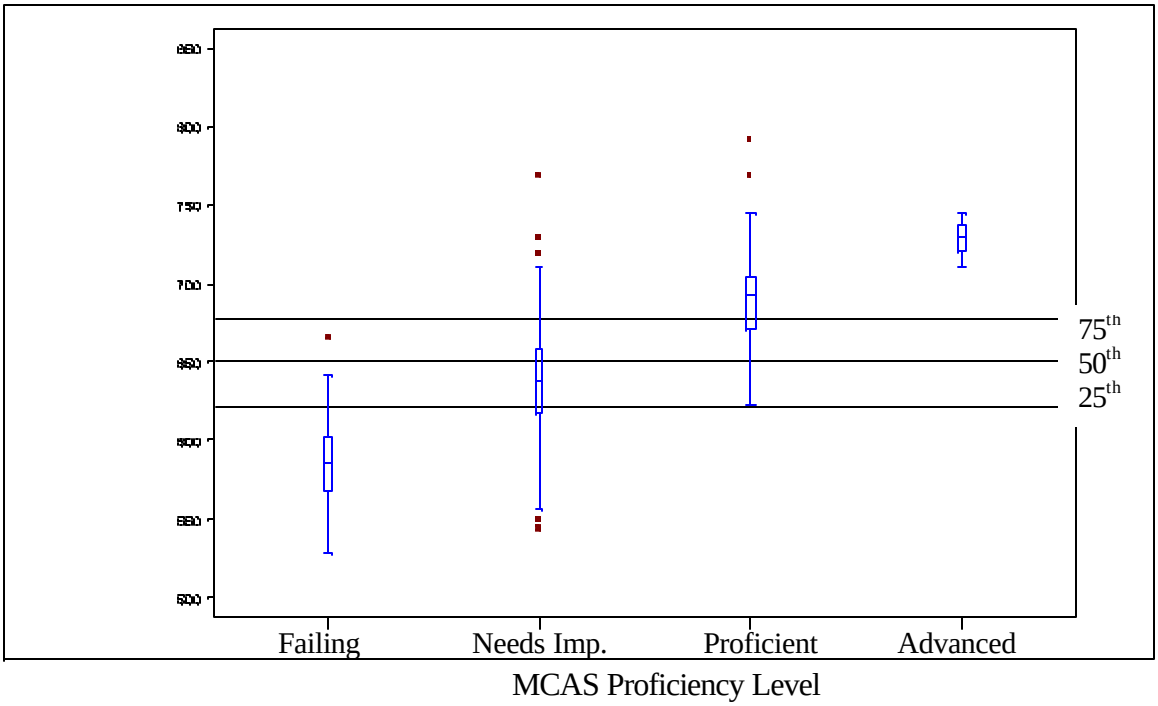
Grade 4

As would be expected for tests that assess substantially related abilities, students' MCAS proficiency levels in Grade 4 were consistently related to performance on the SAT9. Higher MCAS proficiency levels were consistently related to higher performance on the SAT9. This was true both for the MCAS proficiency categories, and for the MCAS scale scores.

Figure 3 through Figure 5 present box plots that graphically present the relationships between performance on the MCAS and the SAT9 for the three content areas of reading, mathematics, and science and technology. Each box represents the distribution of SAT9 scores associated with performance at the *Failing*, *Needs Improvement*, *Proficient*, and *Advanced* MCAS proficiency levels. The box represents the middle 50% of the students in the MCAS category, with the top and bottom lines at the 75th and 25th percentiles. A center line is drawn at the sample median value, and an "x" is printed at the mean. The vertical "whiskers" indicate the dispersion of students, and are at most 1.5 interquartile ranges long. (An interquartile range is the distance between the 25th and 75th sample percentiles.) More extreme values are marked with a dot. For reference, three horizontal lines are labeled on the graph, indicating the 25th, 50th, and 75th percentile scores for the commercial test.

English/Language arts, Grades 4

Figure 3: SAT9 Reading Scores by MCAS English LA Proficiency Level, Grade 4

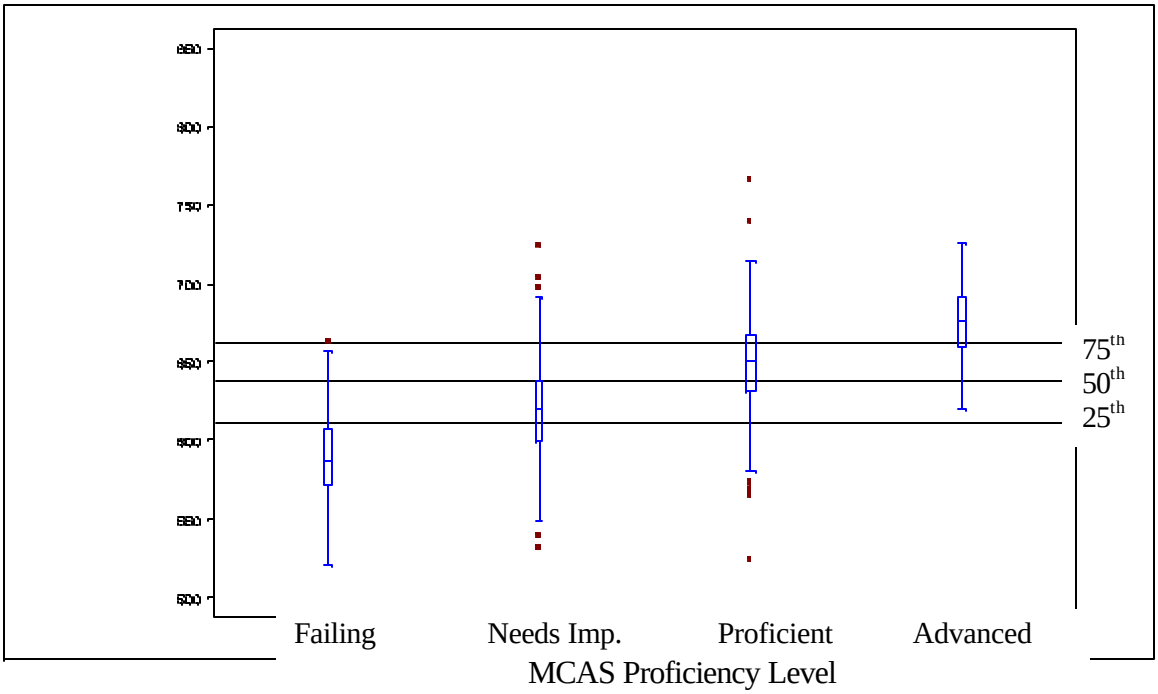


There is good separation between the large number of students who are *Failing*, *Needs Improvement*, *Proficient*, or *Advanced*, and their SAT9 scores in reading. In general, the SAT9 test scores of the middle 50% of the students (depicted by the "box") at each MCAS performance level overlap little or not at all with the middle 50% of students at another performance level. The "whiskers" or lines extending from the boxes are roughly symmetrical, indicating that the scores are roughly symmetrical, or as many scored below the mean as above. In particular, there were very few outliers, or students who scored

high on the SAT9, and scored *Needs Improvement* or *Proficient*. This indicates that students who were high-scoring on the SAT9 took the MCAS seriously. The high standards of MCAS are evident as well, with the average student who was *Failing* scoring about the 6th percentile on the commercial norm-referenced test, the average *Needs Improvement* at the 38th percentile, *Proficient* at the 86th percentile, and *Advanced* at the 98th percentile. Of course, these average percentile ranks would be expected to change with another sample of students or another year of testing, but it indicates that the MCAS proficiency levels represent real differences in performance on the SAT9, and standards well beyond the "average" 50th percentile score on the norm-referenced test.

Mathematics, Grades 4

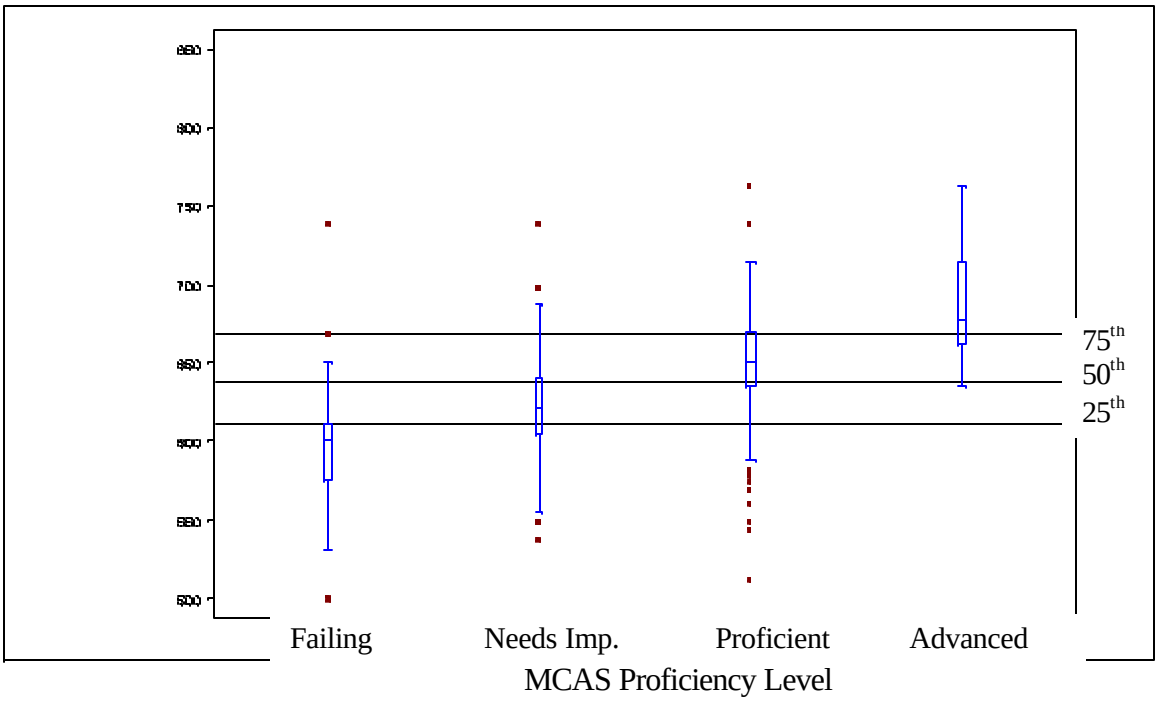
Figure 4: SAT9 Math Scores by MCAS Math Proficiency Level, Grade 4



There was less of a range on the SAT9 on mathematics than on reading, with resulting less separation between mean scores for the MCAS proficiency levels. Still, the means for each group are in separate quartile ranges on the SAT9, with *Failing* located about the 10th percentile, *Needs Improvement* about the 31st percentile, *Proficient* about the 62nd percentile, and *Advanced* about the 83rd percentile.

Science and Technology, Grades 4

Figure 5: SAT9 Science Scores by MCAS Science Proficiency Level, Grade 4



The pattern is similar for Science and Technology, with separation between the various MCAS performance levels. There are a few more outliers with extreme SAT9 values on Science and Technology. The average student who was *Failing* on the MCAS scored at the 12th percentile on the norm-referenced commercial test, *Needs Improvement* scored at the 32nd percentile, *Proficient* scored at the 64th percentile, and *Advanced* scored at the 93th percentile.

Table 6 provides numerical information similar to that portrayed in the box plots, including means, standard deviations, ranges, and the national percentile rank equivalent. Note that students took the MCAS in the spring of their Grade 4, and took the commercial test in the fall of their Grade 5. The norms reflect the Grade 5 fall administration.

Table 6: Commercial Test Scores for Each MCAS Proficiency Level, Grade 4

Grade 4 Cohort, SAT9 Scores by MCAS Proficiency Levels						
Proficiency Level	Mean	NPR	S.D.	Minimum	Maximum	N
English/Language arts (Reading)						
Failing	586.0	6	23.8	529	667	294
Needs Improve.	637.8	38	31.4	545	770	1013
Proficient	692.2	86	30.2	623	793	182
Advanced	729.0	98	13.9	711	745	4
Math (Math)						
Failing	590.0	10	25.6	521	664	356
Needs Improve.	618.6	31	28.7	533	725	765
Proficient	648.0	62	29.7	525	767	312
Advanced	673.7	83	22.8	620	725	111
Science & Techn. (Science)						
Failing	595.6	12	28.3	501	740	188
Needs Improve.	621.7	32	26.2	538	740	694
Proficient	648.9	64	30.4	513	763	605
Advanced	687.9	93	34.2	635	763	59

Grade 10

The patterns for Grade 10 for English/language arts, mathematics, and science support the validity of MCAS, and are similar to those in Grade 4. (See Figure 6 through Figure 8.) Students with higher scores on the commercial standardized test score tend to score higher on the MCAS, with good separation between the median scores for each of the MCAS proficiency levels. The graphs clearly show the high standards of the MCAS. For example, the students who scored *Proficient* on the MCAS in English, mathematics, or science scored at or above the 60th percentile on the commercial test in each corresponding content area.

Figure 6: MAT7 Reading Scores by MCAS Proficiency Level, English Language Arts, Grade 10

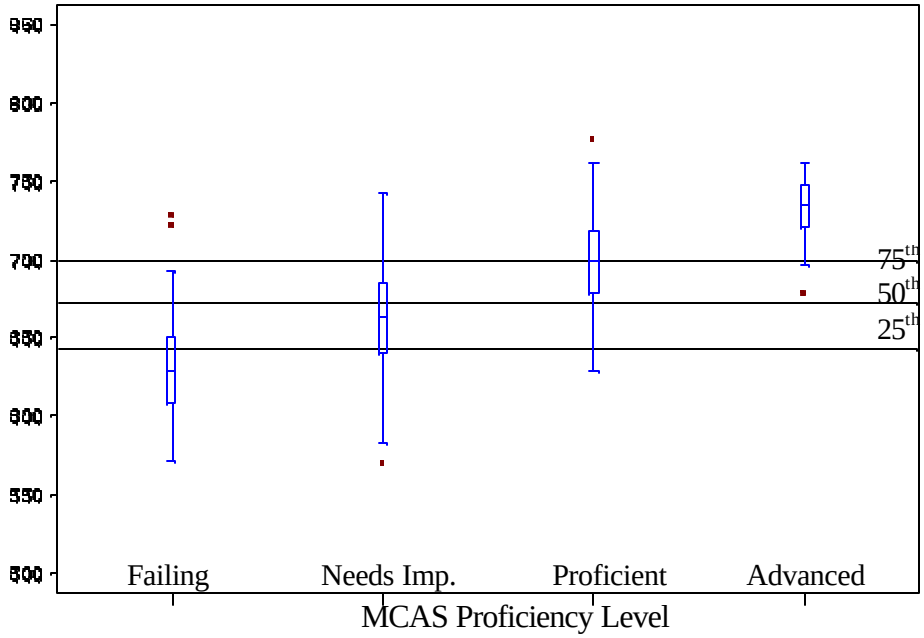


Figure 7: MAT7 Mathematics Scores by MCAS Proficiency Level, Mathematics, Grade 10

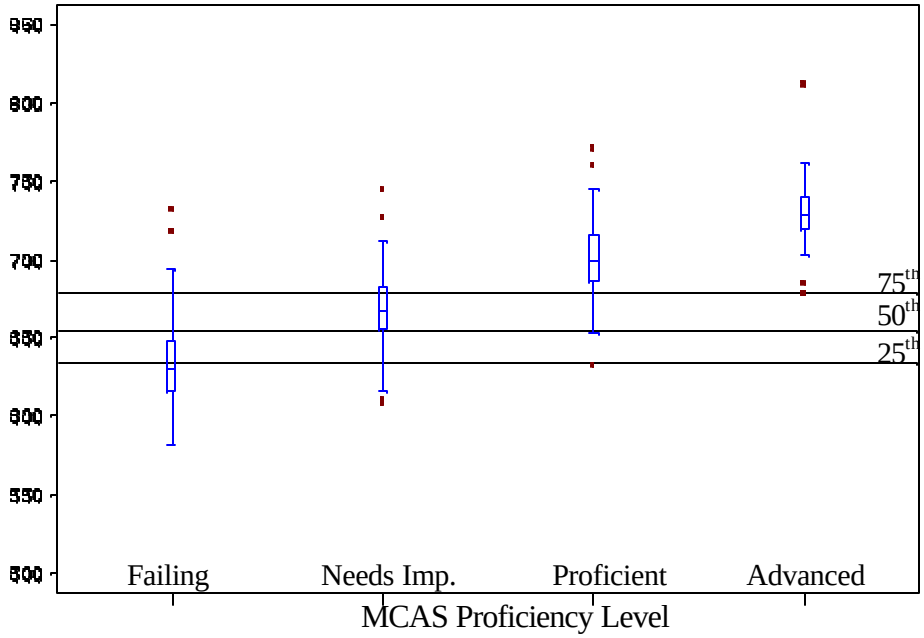


Figure 8: MAT7 Science Scores by MCAS Proficiency Level, Science, Grade 10

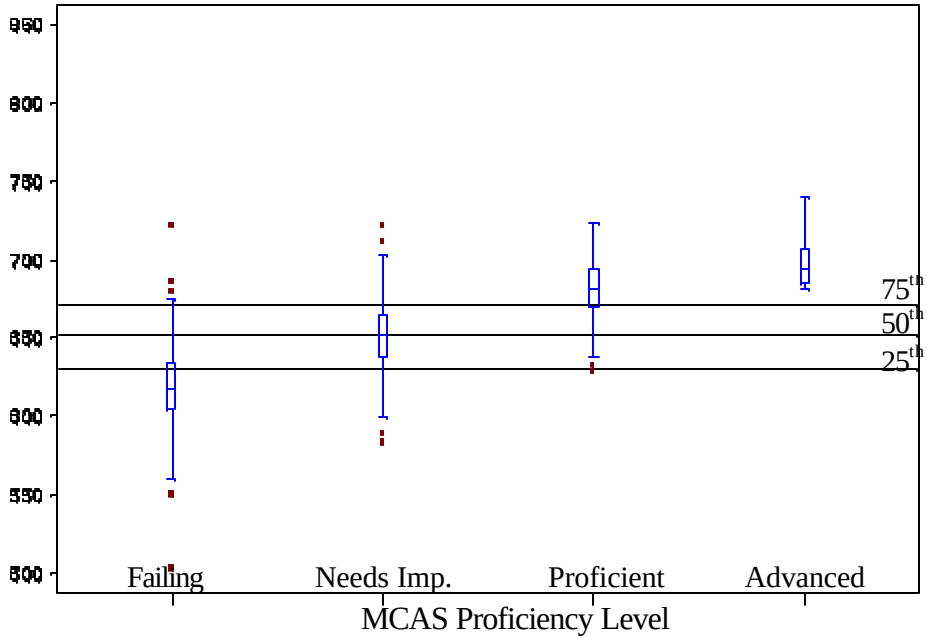


Table 7 provides commercial test information for each of the MCAS proficiency levels for this administration of the MCAS. The mean scale score is provided for the content area that corresponds to each MCAS area, e.g., MAT7 Reading for MCAS English language arts. The national percentile rank that corresponds to the mean scale score is also given for each MCAS proficiency level, by content area. For example, for students who scored *Failing* in MCAS English language arts, the average score on the commercial test was at the 17th percentile for that test. This indicates that the students scored above 17 percent (and below 82 percent) of a nationally representative group of students used to set the norms of the MAT7. The ELA average for the *Needs Improvement* level was at the 39th percentile; the *Proficient* was at the 76th percentile; and *Advanced* was at the 93rd percentile. Note that these NPRs would be expected to change with each administration of the MCAS.

The students took the commercial test in the fall of their Grade 9, then the MCAS in the spring of their Grade 10. The norms reflect the fall, Grade 9 administration.

Table 7 : Commercial Test Scores for Each MCAS Proficiency Level, Grade 10

Grade 10 Cohort, MAT7 Scores by MCAS Proficiency Levels						
Proficiency Level	Mean	NPR	S.D.	Minimum	Maximum	N
English/language arts (Reading)						
Failing	631.0	17	30.2	571	729	169
Needs Improve.	661.7	39	31.7	571	742	293
Proficient	700.6	76	30.0	629	777	199
Advanced	731.7	93	21.0	679	762	20
Math (Math)						
Failing	633.5	26	24.0	582	733	385
Needs Improve.	667.3	64	21.9	609	745	177
Proficient	699.3	89	22.9	634	772	99
Advanced	727.3	97	29.5	679	812	21
Science & Technology (Science)						
Failing	619.1	15	25.8	504	723	230
Needs Improve.	651.6	51	22.4	585	723	323
Proficient	680.3	83	22.0	630	723	106
Advanced	698.8	93	19.3	681	739	8

Mean, SD, minimum, maximum, and n are for the scale scores for students who scored at the designated MCAS proficiency level.
"NPR" is the national percentile rank corresponding to the mean scale score.

Relationships Between Student MCAS Scale Scores and Commercial Test Scores

It is possible to directly compare performance on the MCAS with performance on the commercial tests through using scale scores. Correlational analysis provides an indication of the strength of relationship between performance on the two tests.

Correlations

The pattern of correlations between the MCAS and commercial test scores supports the interpretation that MCAS is assessing largely but not exactly the same constructs as tested by the SAT9. On the other hand, there is some evidence of a method or test effect, where the portions of the SAT9 tend to be more highly related to each other than to the MCAS (with the exception of reading); similarly, the portions of MCAS tend to be more highly related to each other, regardless of subject area, than to the SAT9. The extended Writing on MCAS is somewhat distinct from the other results. Although the pattern of these correlations appear reasonable, these results are difficult to interpret conclusively regarding the validity of MCAS without additional evidence.

Table 8: Correlations between MCAS and SAT9 scores, District A, Grade 4

Variable	1	2	3	4	5	6	7	8
SAT9								
1. Reading								
2. Language	0.82							
3. Composition	0.74	0.88						
4. Math	0.76	0.77	0.69					
5. Science	0.76	0.72	0.65	0.70				
MCAS								
6. ELA	0.82	0.76	0.68	0.70	0.66			
7. Math	0.67	0.66	0.60	0.69	0.61	0.74		
8. Sci. & Tech.	0.71	0.65	0.59	0.64	0.64	0.75	0.75	
9. Writing (raw)	0.58	0.59	0.49	0.56	0.45	0.78*	0.52	0.51

* Writing is one component of the ELA score.

Table 8 gives the correlations between the various test content areas for District A, Grade 4. The SAT9 areas, shown in the upper left triangle of numbers, are highly correlated with each other. The MCAS areas, in the lower right triangle of numbers, are highly correlated with each other as well, with the exception of Writing. (Writing is highly correlated with English language arts, as expected since Writing is one of the components that makes up the ELA total.)

The correlations between the SAT9 and the MCAS are outlined in the box. Underlined correlations are between similar content areas in the two tests (e.g., SAT9 Reading and MCAS English Language Arts). As expected, the correlations between similar content areas tend to be higher than the correlations between dissimilar content areas. Both the SAT9 reading and the MCAS reading are highly correlated with each other and with scores on other content areas. This may be because reading is indicative of a general ability level, because reading is required to understand the test, or some other reasons.

The MCAS extended writing stands out as being less related to any of the other performances (with the exception of MCAS English language arts, of which it is a subcomponent). The writing exercise correlates noticeably lower with the other measures. The scale upon which the writing exercise is based, however, is calibrated somewhat differently from the other measures. It is unclear how much these lower correlations reflect this extended writing tapping somewhat different abilities and skills, and how much involve statistical artifacts.

The correlations provide some evidence that the two tests are different from each other and are not measuring exactly the same things in the same ways. The correlations tend to be higher within each test than on the same content areas between the two tests. One might expect the highest correlations to be between different measures of the same content, e.g., between reading as measured by the SAT9 and the MCAS, or math on the

SAT9 and the MCAS. This is true if the tests measure the same type of reading, or give the same emphasis to the mathematical topics within the test. There are many ways the two tests might be different: difference in the content assessed; format differences (all multiple-choice on the SAT9, multiple-choice and written answers on the MCAS); different levels of knowledge (MCAS items often ask for an explanation); different administration conditions, student motivation, and so on. It is not possible to know with these limited data what accounts for the differences between the two tests.

Table 9: Correlations between MCAS and MAT7 scores, District A, Grade 10

Variable	1	2	3	4	5	6	7	8
MAT7								
1. Reading								
2. Language	0.78							
3. Composition	0.70	0.89						
4. Math	0.75	0.74	0.67					
5. Science	0.75	0.66	0.59	0.70				
MCAS								
6. ELA	0.72	0.68	0.61	0.67	0.61			
7. Math	0.66	0.66	0.59	0.81	0.65	0.71		
8. Sci. & Tech.	0.72	0.64	0.59	0.72	0.71	0.77	0.79	
9. Writing (raw)	0.61	0.61	0.54	0.57	0.46	0.*81	0.56	0.57

* Writing is one component of the ELA score.

Possible Effects of Testing Method

An important question is whether MCAS is inordinately subject to testing certain skills, such as verbal ability, since it relies in part on students writing their responses. While this is an extremely complex issue, one indicator of a verbal test—rather than one that tested mathematics and science, for example—would be that reading proficiency would be a large factor.

It is true that reading is highly correlated with all aspects of both tests. (The exception is MCAS Writing, which has been discussed above.) The magnitude of the effects of reading skills (or general intelligence as manifested on the reading sections) differentially on MCAS versus SAT9 can be estimated as negligible to moderate. The MCAS is less susceptible to showing effects of reading skills than is the SAT9.

The effect of reading or English language arts competence on performance in the other content areas can be estimated through a statistical regression approach. A regression analysis predicts performance on one measure using one or more other "predictor" measures. In this case, the analysis examines whether considering a student's performance on English language arts or reading makes it possible to predict more accurately how a student scored in math or science. The results are shown in Table 10.

Table 10: Method Regressions, District A, Grade 4

Dependent Variable	Predictor Variables with Standardized Coefficients			
	Content	Method/Skill	R^2	ΔR^2
MCAS ELA	SAT9 Reading (.82)		0.67	
MCAS ELA	SAT9 Reading (.71)	SAT9 Composition (.15)	0.68	.01
MCAS ELA	SAT9 Reading (.60)	SAT9 Language (.27)	0.69	.02
SAT9 Mathematics	MCAS Mathematics (.69)		0.48	
SAT9 Mathematics	MCAS Mathematics (.39)	MCAS ELA (.41)	0.55	.07
SAT9 Mathematics	MCAS Mathematics (.33)	SAT9 Reading (.54)	0.64	.16
MCAS Mathematics	SAT9 Mathematics (.69)		0.48	
MCAS Mathematics	SAT9 Mathematics (.34)	MCAS ELA (.50)	0.60	.12
MCAS Mathematics	SAT9 Mathematics (.43)	SAT9 Reading (.34)	0.52	.04
SAT9 Science	MCAS Science (.63)		0.41	
SAT9 Science	MCAS Science (.31)	MCAS ELA (.43)	0.48	.07
SAT9 Science	MCAS Science (.20)	SAT9 Reading (.61)	0.59	.18
MCAS Science	SAT9 Science (.63)		0.41	
MCAS Science	SAT9 Science (.25)	MCAS ELA (.58)	0.59	.18
MCAS Science	SAT9 Science (.24)	SAT9 Reading (.51)	0.52	.11

The change in coefficient (ΔR^2) indicates the difference considering ELA or reading makes. In these analyses, the smallest change could be .00, while the largest could be around .60.

Each comparison involves three results: a) the relationship between an MCAS and a SAT9 test in the same content area, such as science; b) the relationship when MCAS ELA is added, and c) the relationship when SAT9 reading is added.

In the table, the first row corresponds to the equation in which the MCAS English language arts score is predicted solely on the basis of the SAT9 Reading score alone. The second row includes the SAT9 Composition score to predict performance on MCAS English language arts. The small change in R-square from the first two to the second row indicates that SAT9 Composition adds only 1% to the power of SAT9 Reading performance to predict MCAS English language arts performance.

In math and science, the MCAS does show increases when reading scores are considered as well. However, these increases on MCAS predictability are less than the increases on the corresponding SAT9 areas. For example, MCAS Mathematics increases from an R^2=.48 when related to SAT9 Mathematics, to an R^2 = .52 and R^2=.60 when considered

with SAT9 Reading and MCAS English LA, respectively. The SAT9 Mathematics shows larger increases. SAT9 Mathematics considered with MCAS Mathematics alone has an $R^2 = .48$. Adding MCAS Reading or SAT9 Reading increase the R^2 to .55 and .64, respectively. The pattern is similar in Science. The changes in R-square are at least as large for the SAT9, with the addition of SAT9 reading, as they are for MCAS with the addition of MCAS English language arts. One can conclude that reading method effects in MCAS are not greater than those in the SAT9.

The patterns are similar for Grade 10, although the increases in R-square tend to be smaller than were found in Grade 4. These results support the conclusion that MCAS is not unduly a "language-based" test, at least in comparison with the commercial tests which consist entirely of multiple-choice questions.

Table 11: Method Regressions, District A, Grade 10

	Predictor Variables with Standardized Coefficients			
Dependent Variable	Content	Method/Skill	R^2	ΔR^2
MCAS ELA	MAT7 Reading (.72)		0.51	
MCAS ELA	MAT7 Reading (.57)	MAT7 Composition (.21)	0.53	.02
MCAS ELA	MAT7 Reading (.47)	MAT7 Language (.32)	0.55	.04
SAT9 Mathematics	MCAS Mathematics (.81)		0.66	
SAT9 Mathematics	MCAS Mathematics (.68)	MCAS ELA (.17)	0.67	.01
SAT9 Mathematics	MCAS Mathematics (.56)	MAT7 Reading (.37)	0.74	.08
MCAS Mathematics	MAT7 Mathematics (.82)		0.66	
MCAS Mathematics	MAT7 Mathematics (.61)	MCAS ELA (.31)	0.71	.05
MCAS Mathematics	MAT7 Mathematics (.73)	MAT7 Reading (.12)	0.66	.00
MAT7 Science	MCAS Science (.71)		0.51	
MAT7 Science	MCAS Science (.60)	MCAS ELA (.15)	0.52	.01
MAT7 Science	MCAS Science (.36)	MAT7 Reading (.49)	0.63	.12
MCAS Science	MAT7 Science (.71)		0.51	
MCAS Science	MAT7 Science (.38)	MCAS ELA (.54)	0.69	.18
MCAS Science	MAT7 Science (.40)	MAT7 Reading (.42)	0.58	.07

Relationships Between MCAS Scores and Gender, Ethnicity, and Transience

Because the same students took both the SAT9 and the MCAS, it is possible to analyze differences in MCAS performance between males and females, and among different ethnic populations. Differences in test performance tend to exist for a variety of reasons. These differences, however, should not be increased by the test itself without strong justification. A key question, then, is whether differences between gender or racial/ethnic subgroups is greater on the MCAS for this sample of students than would be expected based on comparison with the differences on other tests, the SAT9 and MAT7 in this case.

Gender

The differences in MCAS performance between females and males were minor in general, and in keeping with differences observed on other tests. Differences were more pronounced in high school than elementary school. Most importantly, gender did not make a difference in student performance on the MCAS when performance on the commercial tests was considered. That is, the MCAS did not exhibit differential effects for or against males or females, when compared with their performance on the commercial tests.

The performance by females and males in Grade 4 on the MCAS followed the stereotypical pattern also found on the SAT9. On average, females scored slightly higher than males on the MCAS in writing and reading. Males scored slightly higher on MCAS mathematics. On the SAT9, females scored higher in reading, language, composition, and mathematics. Males scored higher on the SAT9 science. The differences are statistically significant ($p > .05$) for MCAS English language arts and MCAS writing, and for SAT9 Reading, Language, and Composition.

Table 12 gives information about the average performance for males and females.

Table 12: Performance of Females and Males on MCAS and SAT9, Grade 4

Test & Content Area	Statistic	Males	Females	p < .05
MCAS English LA	Mean	226.6	228.7	*
	S.D.	10.5	10.0	
	N	775	770	
MCAS Mathematics	Mean	231.8	231.4	
	S.D.	16.0	15.3	
	N	774	770	
MCAS Sci. & Tech.	Mean	236.0	235.8	
	S.D.	14.0	12.6	
	N	775	770	
MCAS Writing	Mean	9.6	10.8	*
	S.D.	3.3	3.2	
	N	769	769	
SAT9 Reading	Mean	630.9	638.1	*
	S.D.	44.5	40.4	
	N	756	729	
SAT9 Language	Mean	597.3	606.5	*
	S.D.	38.6	36.7	
	N	774	766	
SAT9 Composition	Mean	602.1	612.1	*
	S.D.	43.8	43.7	
	N	771	762	
SAT9 Mathematics	Mean	621.0	622.8	
	S.D.	39.2	33.5	
	N	771	767	
SAT9 Science	Mean	633.1	630.4	
	S.D.	38.1	32.8	
	N	773	765	
* difference significant at p < .05				

The differences between females and males was more pronounced in Grade 10, especially on the MCAS. Females scored significantly higher than males in English language arts and writing. Males scored higher than females in mathematics and science and technology. (See Table 13.) It is not clear how much of this difference may be attributable to differences in interests or courses taken by girls and boys.

Table 13: Performance of Females and Males on MCAS and MAT7, Grade 10

Test & Content Area	Statistic	Males	Females	p < .05
MCAS English LA	Mean	227.4	231.5	*
	S.D.	16.4	16.2	
	N	305	372	
MCAS Mathematics	Mean	221.1	217.4	*
	S.D.	19.2	17.3	
	N	306	372	
MCAS Sci. & Tech.	Mean	226.4	222.7	*
	S.D.	15.3	13.3	
	N	306	372	
MCAS Writing	Mean	11.5	12.5	*
	S.D.	3.3	3.1	
	N	289	350	
SAT9 Reading	Mean	667.3	667.3	
	S.D.	43.5	39.5	
	N	305	372	
SAT9 Language	Mean	641.5	644.8	
	S.D.	35.6	36.7	
	N	300	366	
SAT9 Composition	Mean	638.9	643.1	
	S.D.	44.8	45.8	
	N	302	371	
SAT9 Mathematics	Mean	657.9	652.0	
	S.D.	39.6	32.5	
	N	305	372	
SAT9 Science	Mean	651.9	640.0	*
	S.D.	34.4	29.3	
	N	300	362	
* difference significant at p < .05				

Regression analysis was used to examine the extent to which MCAS differences in performance for males and females were similar to differences in SAT9 scores. No significant gains in predictability were found on the basis of considering gender. That is, considering gender did not appreciably contribute to predicting performance on MCAS compared to the SAT9.

Regression analysis can be used to examine how variables in combination with each other can predict scores on a criterion or dependent variable. For each of the three content areas, a regression equation was calculated predicting MCAS scores from only the matching SAT9 content scores. Then, a second equation was created which added gender. If MCAS scores were exhibiting greater gender differences than expected from SAT9 gender differences, there would be meaningful increases in the predictability of MCAS scores when gender is added. Also, if gender differences were meaningful beyond the differences that occur in the SAT9 scores, gender would have a noticeable weight while the commercial test weight would remain similar to the equation without gender.

Table 14: Regression Results Showing Adjusted Strengths of Gender Effects, Grade 4

	Standardized Coefficient			Change in R^2
MCAS Test	SAT9 Control	Gender	R^2	due to Gender
English LA	.81		.66	
English LA	.81	-.09	.67	.01
Mathematics	.69		.48	
Mathematics	.69	.06	.48	.00
Sci. & Tech.	.64		.41	
Sci. & Tech.	.64	-.03	.41	.00
SAT9	MCAS	Gender	R^2	Change in R^2 due to Gender
Reading	.81		.66	
Reading	.81	.01	.66	.00
Mathematics	.69		.48	
Mathematics	.69	-.08	.48	.00
Science	.64		.41	
Science	.64	.05	.41	.00

Table 14 presents regression results for each of the MCAS content areas. The rows in the table can be read as the equations predicting the MCAS scores from the commercial test scores and gender. In these equations, variables are standardized so they all have a mean of zero and a standard deviation of one. Gender is coded such that negative weights mean that females tend to have slightly higher scores than would be expected from gender differences on the contrasting test.

Note that considering gender resulted in no increase in the R-squares, indicating that gender was not a factor in the MCAS beyond what it was in the commercial test for this sample of students. There were some small negative or positive weights for gender, depending on the content area, indicating that females scored higher on the MCAS than would be predicted from their performance on the SAT9 in English language arts and science. The boys scored higher than would be predicted on the MCAS mathematics. Curiously, the pattern was reversed on the SAT9, where females scored higher than expected on the SAT9 mathematics. This analysis does not necessarily mean that males or females are unfairly discriminated against by the MCAS. It does mean that compared to their counterparts, they do not do as well on the MCAS as predicted by their SAT9 scores. It is not possible to give conclusive reasons for the difference. The overall results, again, indicate no significant increase in r-square, indicating that gender did not contribute to a detectable difference in performance beyond what would be expected based on the other test.

The results are essentially the same for Grade 10. There are no significant detectable gender effects on MCAS over what is present in the commercial test. The Grade 10 results are shown in Table 15.

Table 15: Regression Results Showing Adjusted Strengths of Gender Effects, Grade 10

	Standardized Coefficient			Change in R^2 due to Gender
MCAS Test	MAT7 Control	Gender	R^2	
English LA	.71		.506	
English LA	.72	-.24	.521	.01
Mathematics	.82		.658	
Mathematics	.81	.07	.659	.00
Sci. & Tech.	.71		.510	
Sci. & Tech.	.71	.01	.510	.00
MAT7	MCAS	Gender	R^2	Change in R^2 due to Gender
Reading	.71		.506	
Reading	.72	.18	.514	.00
Mathematics	.81		.658	
Mathematics	.81	.00	.658	.00
Science	.72		.510	
Science	.70	.18	.517	.01

Ethnicity

No undue differential effects were found for the MCAS in terms of ethnicity. Students of various ethnic groups performed comparably on MCAS and on the commercial tests. However, there were large differences in achievement by various racial/ethnic subgroups on both tests.

The district's matched sample was diverse racially, with 55% of the students reporting themselves as white, 22% Hispanic, 9% African-American, 5% Asian/Pacific Islander, 2% Native American, and 7% as more than one of the above in Grade 4. (See Table 16.) By Grade 10, the proportion of white and Asian/Pacific Islander students had increased, and the proportion of Hispanic, African-American and Native American students had decreased. Note that these ethnic/racial identities were self-reported and not verified. Due to the process of matching students and the smaller numbers of students involved, these may not represent the actual proportions of students enrolled in the district.

Table 16: Number and Percents of Students in Sample by Ethnic/Racial Subgroup

	White	Hispanic	African-American	Asian	Native American	Multiple Ethnic/ Racial	Total
Grade 4							
Number of Students	816	348	136	86	28	104	1518
Percent of Total	53.8%	22.9%	9.0%	5.7%	1.8%	6.9%	100.1%
Grade 10							
Number of Students	359	86	32	74	6	37	594
Percent of Total	60.4%	14.5%	5.4%	12.5%	1.0%	6.2%	100.0%

There were substantial differences between subgroups on both the MCAS and SAT9. The differences between the mean performance of subgroups on the MCAS were similar to those on the SAT9. Thus, it appears that the MCAS does not result in larger subgroup differences than the SAT9. In addition, a regression analysis indicates that there were small, if any, differences due to ethnicity. The ethnicity effects for performance on the MCAS tended to be smaller than those on the SAT9. (See Table 17 and Table 18.)

On the MCAS English language arts and science and technology, the white students on average scored significantly higher than the other racial/ethnic subgroups ($p < .05$), and outscored all other racial/ethnic groups on mathematics except Asian/Pacific Islanders. Hispanic students were significantly lower than Asian students in every area. There were few other significant differences in pair-wise comparisons.

The pattern was similar for the SAT9. White students were the highest on every SAT9 content area but one, followed by the Asian and multiple ethnic/racial subgroups. On the SAT9 math, the Asian subgroup scored the highest.

It is not possible from these data to ascertain how much of the difference in performance between subgroups is due to factors such as "real" achievement, language facility, reading/writing, "cultural bias," motivation, etc.

Table 17: Regression Results Showing Adjusted Strengths of Race/Ethnicity Effects, Grade 4

MCAS Test	SAT9 Control	Race/Ethnicity	R^2	Change in R^2 due to Race/Ethnicity
English LA	Reading		.667	
English LA	Reading	Race/Ethnicity	.672	.01
Mathematics	Mathematics		.478	
Mathematics	Mathematics	Race/Ethnicity	.497	.02
Sci. & Tech.	Science		.406	
Sci. & Tech.	Science	Race/Ethnicity	.438	.03
SAT9	MCAS	Race/Ethnicity	R^2	Change in R^2 due to Race/Ethnicity
Reading	Reading		.667	
Reading	Reading	Race/Ethnicity	.678	.01
Mathematics	Mathematics		.478	
Mathematics	Mathematics	Race/Ethnicity	.498	.02
Science	Science		.406	
Science	Science	Race/Ethnicity	.417	.01

The results were similar for Grade 10, in that a regression analysis found minor changes in R-square when race/ethnicity was considered for MCAS. Effects for MCAS were smaller than for the commercial test.

Table 18: Regression Results Showing Adjusted Strengths of Race/Ethnicity Effects, Grade 10

MCAS Test	MAT7 Control	Race/Ethnicity	R^2	Change in R^2 due to Race/Ethnicity
English LA	Reading		.548	
English LA	Reading	Race/Ethnicity	.556	.01
Mathematics	Mathematics		.684	
Mathematics	Mathematics	Race/Ethnicity	.685	.01
Sci. & Tech.	Science		.552	
Sci. & Tech.	Science	Race/Ethnicity	.564	.01
SAT9	MCAS	Race/Ethnicity	R^2	Change in R^2 due to Race/Ethnicity
Reading	Reading		.548	
Reading	Reading	Race/Ethnicity	.587	.04
Mathematics	Mathematics		.684	
Mathematics	Mathematics	Race/Ethnicity	.700	.02
Science	Science		.552	
Science	Science	Race/Ethnicity	.574	.02

There were large average score differences between many of the racial/ethnic subgroups. It is important to note that the existence of differences is not necessarily an indication of bias on the part of the test. Differences on the test may accurately reflect differences in learning due to different opportunities to learn. For example, if a subgroup of students took a stronger mathematics course of instruction than another subgroup, their test results would be expected to be different, and not because of any bias in test construction or scoring. This issue of differences in racial/ethnic group performance on the MCAS is addressed more extensively in the section on school courses.

Table 19 provides the numerical information regarding performance on the MCAS and SAT9 content area tests by ethnic/racial subgroup for Grade 4. Table 20 provides similar information for Grade 10.

Table 19: Performance on MCAS and SAT9 by Ethnic/Racial Subgroup, Grade 4

Test/ Content	Statistic	White	Hispanic	African-American	Asian/ Pacific Islander	Native American	Multiple Ethnic/ Racial
MCAS							
ELA	Mean	231.3	222.0	224.1	227.7	220.6	226.4
	S.D.	10.0	8.0	8.3	9.1	9.3	8.9
	N	816	348	136	86	28	104
Math	Mean	236.6	223.5	225.8	234.8	225.4	228.8
	S.D.	15.8	12.3	12.9	13.7	14.0	15.3
	N	815	348	136	86	28	104
Sci. & T	Mean	240.6	228.9	230.3	236.9	230.5	233.7
	S.D.	12.5	12.3	12.2	11.5	12.0	13.5
	N	816	348	136	86	28	104
Writing	Mean	11.1	8.9	9.3	10.5	8.4	9.8
	S.D.	3.2	3.0	3.2	3.2	3.2	2.8
	N	815	344	136	86	27	104
SAT9							
Reading	Mean	650.0	608.2	618.1	632.6	608.3	632.2
	S.D.	41.4	31.4	35.1	31.9	38.4	44.1
	N	815	327	130	84	27	101
Langua.	Mean	613.7	581.0	588.4	608.3	579.5	600.1
	S.D.	37.3	30.1	35.1	34.4	39.1	36.6
	N	815	346	136	86	28	102
Comp	Mean	619.1	586.6	592.9	607.6	589.7	607.1
	S.D.	43.9	36.9	38.5	43.8	38.0	45.9
	N	811	345	135	86	27	102
Math	Mean	631.9	603.7	607.3	640.9	600.4	617.0
	S.D.	35.5	30.2	34.3	33.3	38.9	34.0
	N	816	344	136	86	28	101
Science	Mean	642.1	614.8	617.9	633.3	619.3	628.7
	S.D.	35.6	30.4	33.6	28.0	28.8	30.9
	N	815	345	136	86	28	102

Table 20: Performance on MCAS and MAT7 by Ethnic/Racial Subgroup, Grade 10

Test/ Content	Statistic	White	Hispanic	African-Ameri-can	Asian/ Pacific Islander	Native Ameri-can	Multiple Ethnic/ Racial
MCAS							
ELA	Mean	234.5	219.7	224.3	227.9	217.3	230.0
	S.D.	14.9	15.3	13.2	15.6	21.2	12.8
	N	359	86	32	74	6	37
Math	Mean	223.7	209.0	209.3	221.6	208.3	217.1
	S.D.	18.9	14.3	11.6	17.2	9.2	15.3
	N	359	86	32	74	6	37
Sci. & T	Mean	228.7	214.8	218.1	223.9	216.7	225.7
	S.D.	13.4	12.8	11.0	12.4	13.5	12.2
	N	359	86	32	74	6	37
Writing	Mean	12.7	10.2	10.6	11.3	10.0	12.5
	S.D.	3.1	3.0	2.6	3.8	4.1	2.7
	N	350	79	31	70	5	37
SAT9							
Reading	Mean	681.1	635.8	647.7	649.9	654.3	664.0
	S.D.	38.9	35.9	35.2	37.4	45.2	34.9
	N	359	85	32	74	6	37
Langua.	Mean	652.0	619.4	626.0	639.4	632.0	642.1
	S.D.	34.8	32.2	33.8	34.7	42.2	30.7
	N	358	82	29	72	6	36
Comp	Mean	650.7	616.5	620.3	641.1	625.2	634.3
	S.D.	44.9	43.6	38.0	45.7	43.9	40.6
	N	359	85	30	73	6	37
Math	Mean	664.1	628.7	631.7	656.1	635.5	648.3
	S.D.	34.8	29.8	30.2	34.7	26.3	31.5
	N	359	85	32	74	6	37
Science	Mean	655.3	622.5	629.3	636.1	634.0	646.3
	S.D.	30.6	31.6	24.4	28.4	31.4	25.7
	N	356	82	29	71	6	36

The average scores of the various racial/ethnic subgroups were evaluated for statistically significant differences, where one group scored higher than another. The SAT9 and MAT7 showed a few more significant differences between groups than did the MCAS. This was consistent with the regression analysis which indicated that the MCAS was not more subject to differential performance than were the standardized, commercial tests. The results of Bonferroni pair-wise comparisons are shown in Table 21, where $p < .05$. Each line in the table denotes one group that scored significantly higher than the other

designated group(s). Indented lines denote further ordering. For example, in MCAS English language arts, "W>A B H M N" indicates that whites scored significantly higher than Asian, black, Hispanic, Multiple, and Native Americans. (The letters are listed alphabetically.) "A > H" indicates that Asians scored higher than Hispanics. "M > H" indicates that students with multiple racial/ethnic backgrounds scored higher than Hispanics. Thus it may be inferred that "Whites scored higher than Asians who scored higher than Hispanics."

Note that the MCAS and commercial tests have virtually the same characteristics, where significant differences were found in the same content areas between the same groups. This supports the conclusion that the test results consistently show real learning differences, and are not merely due to some characteristic of the test. The few differences unique to the test (MCAS or SAT9/MAT7) are highlighted in bold print.

Table 21: Performance by Racial/Ethnic Subgroups, Grades 4 and 10

Grade 4			
MCAS		**SAT9**	
Content area	Significant Mean Differences	Content area	Significant Mean Differences
English language arts	W>A B H M N A> H M> H	Reading	W>A B H M N A> H M> H
Mathematics	W> B H M N M > H A > B H	Mathematics	W> B H M N M> H A> B H **M N**
Science	W> B H M N M> H A > B H	Science	W> B H M N M> H A > B H
Writing	W> B H M N A> H N		
		Language	W> B H M N M> H A > B H N
		Composition	W> B H N A > H M > H
Grade 10			
MCAS		**MAT7**	
Content area	Significant Mean Differences	Content area	Significant Mean Differences
English language arts	W > A B H **A > H** M > H	Reading	W > A B H M > H
Mathematics	W > B H A > B H	Mathematics	W > B H A > B H **M > H**
Science	W > B H A > H M > H	Science	W > **A** B H M > H
Writing	W > A B H M > H		
		Language	W > B H A > H M > H
		Composition	W > B H A > H

Legend: A = Asian/Pacific Islander; B = African American/Black, H = Hispanic/Latino, N = Native American, W = white, M = Multiple (more than one of the previous). Means different, $p < .05$ using Bonferroni t-tests. W>ABH indicates that whites scored higher than Asians, blacks, and Hispanics. Groups are listed alphabetically; other significant group differences must be designated by an additional line, e.g., A>H.

Although student performance on the MCAS and the commercial tests are fairly highly correlated, it might be asked what the differences are for racial/ethnic subgroups on the two tests. In particular, do students in various racial/ethnic groups do better on one test than another? The answer is that differences are generally small. White students tend to do better on Grade 10 MCAS English language arts, mathematics, and science than would be expected from just considering their performance on the MAT7, while minority students tend to score lower on MCAS than would be expected. In Grade 4, Asian/Pacific Islander students tend to do better on MCAS English language arts, mathematics, and science than would be expected from just considering their performance on the SAT9, and white students do better on English language arts and science. Results were mixed, but in general the other minority groups scored lower on the MCAS than would be predicted just considering their SAT9 scores. The results are given in Table 25.

Table 22: Residuals from predicting MCAS from commercial tests, by race/ethnicity, by grade

Residuals from predicting MCAS scores from Commercial Tests						
	Asian/ Pacific Islander	African American	Hispanic/ Latino	Multiple	Native American[1]	White
Grade 4						
English	2.28	-0.71	-1.78	0.42	-9.54	0.13
Math	1.14	-0.86	0.15	-0.07	-3.45	-0.13
Science	2.11	-1.70	-2.79	0.61	-4.89	0.38
Grade 10						
English	-0.21	-0.36	-0.68	-0.96	-1.94	0.55
Math	-2.64	-1.68	-2.77	-1.00	0.04	1.85
Science	0.41	-2.47	-3.12	-1.41	-2.62	1.96

[1]-Due to small number of students, results not reliable.

The residuals are obtained by computing a regression equation between the MCAS score and the commercial test score. The regression equation allows one to compute a predicted MCAS score, based on the commercial test score. The residual is the difference between the predicted score and the observed score. Thus the residuals show whether a racial/ethnic group, on average, scored higher (positive residual) or lower (negative residual), and the number of MCAS points different than was predicted when taking into consideration the performance on the commercial test.

Most of the differences in Grade 4 are 2-4 points between the scores that were over-predicted and under-predicted. In Grade 10, most differences are 3-5 points. These differences are about one-fifth to two-fifths of an effect size, which is not an uncommon amount. It is not clear whether these differences reflect bias in one of the tests (MCAS *or* commercial test), differential performance by subgroups on the lower ends of the tests, or

other factors. Additional data over time could help resolve whether there is differential improvement on the MCAS for various racial/ethnic subgroups, irrespective of their initial standing on the commercial tests.

Transience

Students indicated on the student questionnaire how many years they had been attending the same school and the same district. The extent to which students were consistently in the school and district over time, or student transience, was related to student performance on the MCAS, but not more so than on the commercial tests in general. However, some anomalous results for both Grade 4 and Grade 10 indicate that this issue should be researched more extensively.

In Grade 4, the longer the student reported she or he had been enrolled at the same school, the higher the probability the student would score *Proficient* or *Advanced* on the MCAS. The students who reported they had been enrolled in the school since the 1st grade were scored *Failing* about half as often as those who had attended only beginning the year of the MCAS test, and were over 40% more likely to score *Proficient* or *Advanced*. The pattern was similar for moving within the district.

The results are similar for the effect of transience on SAT9 performance.

Table 23 gives the percent of students within each transience category who scored at various levels on the MCAS and the SAT9, for selected content area tests. Note again that these are based on self-reported data from Grade 4 students regarding their enrollment over the previous three years. Note, for example, that the percentage of students who were at the *Failing* proficiency level in English language arts increased as the student spent less time in the school. 15% of the students were at the *Failing* level who reported being in the same school since the first grade to the fourth grade (when MCAS was administered). That percentage increased to 25% of the students who were at the *Failing* level for those who had moved in during the second grade, and 27% were at *Failing* who moved in during the third grade. Curiously, only 17% were at *Failing* who reported attending the school for less than one year. The numbers are smoother for mathematics and science.

Table 23: Performance by Transience, Selected Measures, Grade 4

Percentage of students (school / district)										
	Total Percent		MCAS				SAT9			
			Failing		Proficient or Advanced		25th Percentile or lower		75th Percentile or higher	
Attended school/ district since	S*	D*	S	D	S	D	S	D	S	D
			English Language Arts				SAT9 Reading			
1st grade	59	80	15	18	15	14	31	34	20	36
2nd grade	11	6	25	35	10	7	41	51	12	25
3rd grade	12	7	27	30	9	7	48	50	11	26
Less than one year	18	7	17	25	9	7	49	42	9	24
			Mathematics				SAT9 Mathematics			
1st grade			17	20	33	30	31	34	18	35
2nd grade			25	34	22	17	38	48	10	17
3rd grade			32	40	24	20	45	50	15	22
Less than one year			37	35	16	22	51	50	8	26
			Science				SAT9 Science			
1st grade			8	10	50	46	23	25	20	45
2nd grade			12	21	37	32	27	30	10	35
3rd grade			17	18	38	31	30	35	9	36
Less than one year			20	14	30	36	34	30	8	45
* S = reported being in the same school D = reported being in the same district										

Results for Grade 10 are given in Table 24. The results for Grade 10 tend to support the view that increased transience is associated with lower scores. However, there are anomalies, especially when the results are compared for how long a student was in the school versus how long the student was in the district. For example, students who said they moved into the school in the 10th grade (and the MCAS was administered in the spring of the 10th grade year) scored *Failing* on English language arts at twice the rate as those who said they had attended the school since the 9th grade, (43% compared to 21%). However, the pattern was reversed for the district results, with recent enrollees outperforming those who have been in the district longer than one year. The district results for mathematics are still different.

Table 24: Performance by Transience, Selected Measures, Grade 10

Percentage of students (school / district)										
Attended school/ district since	Total Percent		MCAS				MAT7			
			Failing		Proficient or Advanced		25th Percentile or lower		75th Percentile or higher	
	S*	D*	S	D	S	D	S	D	S	D
			English Language Arts				MAT7 Reading			
7th grade	6	83	-	16	-	39	-	27	-	24
8th grade	1	3	-	43	-	14	-	71	-	7
9th grade	88	10	21	47	35	17	31	49	27	13
Less than one year	5	5	43	26	10	19	48	40	7	19
			Mathematics				MAT7 Mathematics			
7th grade			-	28	-	22	-	26	-	28
8th grade			-	14	-	7	-	57	-	7
9th grade			53	15	19	9	29	45	25	11
Less than one year			87	33	3	11	55	37	10	19
			Science				MAT7 Science			
7th grade			-	26	-	20	-	29	-	21
8th grade			-	71	-	7	-	77	-	15
9th grade			30	47	19	13	32	51	20	10
Less than one year			73	30	3	11	70	26	4	15
* S = reported being in the same school D = reported being in the same district										

These Grade 10 results may reflect problems with the student self-reports, the wording of the questionnaire, complex patterns of movement between private/parochial schools and public high schools, immigration, or other factors.

Although the descriptive patterns of MCAS proficiency levels are suggestive, the correlations are low to moderate. The correlations between the student's test scores and the student's report of how long she/he had been in the school and in the district are given in Table 25.

Table 25: Correlations between Student Transience and Test Scores

Correlations					
Grade 4			**Grade 10**		
Test Area	Transience in School	Transience in District	Test Area	Transience in School	Transience in District
MCAS			MCAS		
English	-.20	-.12	English	-.03	-.20
Mathematics	-.20	-.12	Mathematics	-.03	-.14
Sci. & Tech.	-.21	-.12	Sci. & Tech.	-.03	-.14
SAT9			MAT7		
Reading	-.21	-.12	Reading	-.06	-.21
Mathematics	-.18	-.11	Mathematics	-.01	-.15
Science	-.16	-.10	Science	-.04	-.16

Note the very low correlations in Grade 10 for school transience. These are probably due to the wording of the student questionnaire and the placement of the test. The questionnaire asked whether the student had been in the same school since the 7th, 8th, 9th, or 10th grade. Since the high schools in the district studied were 9-12 schools, the only valid answers were 9th grade (88% of the students) and 10th grade (5% of the students). That restriction of range results in correlations around zero, as would be expected. The correlational patterns for Grade 4 do not have this restriction and are more as might be expected.

It sounds reasonable that, all other factors being equal, greater stability would be related to higher educational achievement. However, there are obviously many possible reasons for transience. And there are curious anomalies in these data, such as the higher failure rate of the cohort that reported being in the district since the 2nd grade. It would be worthwhile to do a more extensive analysis of transience, using district enrollment data rather than relying solely on student self-reports. More information is needed to understand fully the relationship of transience and achievement.

Relationships Between Student MCAS Performance and Courses Taken

There were moderately strong relationships between performance on the MCAS and students' reported course-taking patterns in Grade 10. Higher student performance in mathematics and science on the MCAS was related to students taking more advanced courses. This indicates that MCAS is sensitive to students' opportunities to learn, and that performance on the MCAS will likely rise as students learn the curricula aligned with

the Massachusetts curriculum frameworks. The relationships between students' reports of their curriculum and performance on the MCAS were much weaker for Grade 4.

Grade 4

The student questionnaire contained several questions related to specific topics of study and time spent. For example, students were asked whether they had studied living things, Earth and space science, motion and energy—which indicates some breadth of coverage among the life, earth, and physical sciences curriculum. Similar questions were asked for mathematics and history/social studies. The relationship of school curriculum to ELA test scores were not analyzed since the MCAS student questionnaire did not contain questions at Grade 4 about English language arts topics or study time. (See the Appendix for copies of the student questionnaire.)

The analyses revealed that the strongest relationships with students' test performance was with whether they reported having studied certain topics. The relationship with test performance was stronger with mathematics than with science, and with topics than with reports of how often they studied. For example, studying certain Math topics correlated with the MCAS Math performance scale scores about r = 0.28, and frequency of studying Science correlated with science scale scores about r = 0.11. The correlations are shown in Table 26.

It is interesting that the students' reports of whether they had been exposed to the listed topics (even very briefly described) correlated more highly with MCAS performance than with performance on the SAT9 same content area test. The relationship between elementary curriculum, instruction, and performance on the MCAS needs to be explored more extensively than is possible with the limited data from the student questionnaire.

Table 26: Correlations between Curriculum/Instruction and Test Performance, Grade 4

Correlations				
	MCAS Math	SAT9 Math	MCAS Science	SAT9 Science
Math Topics	.28	.15	.27	.15
Math Time	.10	.04	.13	.04
Science Topics	.17	.11	.16	.10
Science Time	.08	.02	.11	.00
The topics and time variables are from the Student Questionnaire.				

It is not unusual for students' self-reports of studying to be weakly related to actual performance. There are many possible reasons for this. For example, time spent is not an indicator of quality: students may spend an hour a day studying mathematics because they are exploring topics on their own, or because they are very slow at completing the required homework. Frequency of studying "science" in class does not indicate what was

learned or how well. In addition, the curriculum within the district may be more uniform, which might result in low variance in the reported curricular exposure, and hence to low correlations.

Grade 10

The Grade 10 MCAS results were strongly related to students' course enrollment status in specific courses. The student questionnaire had questions involving enrollment in mathematics (integrated/algebra 1, geometry/integrated math II, algebra II/integrated math III, advanced math), and science (integrated, biology, earth, physics, and chemistry).

Higher performance on the MCAS was related to students' taking more advanced courses. The correlations ranged from highs of around r=.50 for advanced mathematics and biology, to negative correlations for integrated science and earth science courses. The negative correlations and other results in science are discussed more below. In addition, the course-taking patterns have strong implications for understanding and addressing racial/ethnic group differences on the MCAS. The racial/ethnic group differences are also discussed more below.

The course-taking variable ranged from 1=have not taken and do not plan to take, 2=plan to take, 3=am currently enrolled, and 4=have completed. The correlations are shown in Table 27 between enrollment status for the specified high school mathematics and science courses and performance on the MCAS mathematics or science and technology sections.

Table 27: Correlations between MCAS performance and Enrollment in Math and Science Courses, Grade 10

MCAS score	Courses				
	Integrated I /Algebra I	Geometry/ Integrated II	Algebra II /Integrated III	Advanced	
Mathematics scale score	.24	.44	.36	.50	
	Integrated	Biology	Earth	Chemistry	Physics
Science scale score	-.35	.51	-.24	.47	.01

The negative correlations in integrated and earth science courses indicate that those students who indicated being currently enrolled or having completed the course had lower MCAS science and technology scores. This result may reflect that being enrolled in Integrated science in Grade 10 is related to lower performance because a sophomore

would usually have taken a disciplinary science sequence. (Over 95% of the students who reported they had not taken Integrated Science and did not plan to, reported that they had already completed a Biology course. That is, the academic track is this district apparently is to complete Biology in Grade 9, then enroll in Chemistry in Grade 10.). Similarly, those students enrolled in Earth science may not be taking the main academic sequence in science. The lack of correlation for Physics reflects the fact that Physics apparently is the last science course for these students; fewer than 5% of the students reported being enrolled in a Physics course currently or in the past. The correlation of .01 indicates that students' stated intentions whether or not to enroll in Physics sometime in the future are not related to their MCAS science and technology performance in Grade 10.

The mathematics correlations are higher in general. The course-taking sequence in mathematics is also more uniform (i.e., all the courses fit into a single sequence). This may reflect the nature of the mathematics discipline as taught in high schools, in contrast to science.

The moderately strong correlations between student performance and single questions on the student questionnaire support the view that MCAS is related to specific courses of study. This indicates that MCAS is tapping into what students are studying in their academic courses. More importantly, it indicates that coursework is an important determinant in student performance on the Grade 10 MCAS. This will be an essential point as achieving a passing score becomes required for student graduation. Schools are and can be providing students with the opportunities to learn the academic content knowledge and skills assessed by MCAS.

About 98% of the Grade 10 students in the district reported being currently enrolled in a mathematics course; about 71% reported being currently enrolled in a science course. (Note that some students may not have responded to the questionnaire.) There is a specific sequence of courses in mathematics, as indicated both by the percentages of students who are enrolled and also the very high percentages of students who reported having completed the previous course in the sequence. The sequence is less clear in science. Since MCAS is designed to reflect the Massachusetts common curriculum frameworks, one question will be to what extent the students learn the knowledge specified in the frameworks in their courses. The course enrollments are shown in Table 28.

Table 28: Percentages of students enrolled in math and science courses, Grade 10

	Course Enrollment Patterns					
	Integrated I /Algebra I	Geometry/ Integrated II	Algebra II /Integrated III	Advanced Math		Total
Currently enrolled	11	57	18	12		98
Completed previous course		94	86	99[1]		
	Integrated	Biology	Chemistry	Earth	Physics	Total
Currently enrolled	11	9	44	5	2	71
Completed previous course		87	43			

[1]- 8% already completed Algebra II or Integrated III; additional 91% reported being concurrently enrolled in one of those courses

Grade 10, Race/Ethnicity

Examining student enrollment in mathematics and science courses is one example of seeking to understand what factors contribute to performance on the MCAS. This is an essential part to establishing the validity of the MCAS, and to ensuring that the results from the MCAS are used constructively to improve education in the state. To extend the example, course enrollments were examined for racial/ethnic groups.

As discussed previously in this paper, there were large performance differences on the MCAS between racial/ethnic groups. White students generally scored higher, and Hispanic/Latino students generally scored the lowest of the various subgroups. Comparable differences were found for the commercial tests. The question might be asked about the causes of these differences. Examining the course enrollment patterns for racial/ethnic subgroups provides some important insights.

There were large differences between racial/ethnic groups and the courses enrolled in science and mathematics that correspond to the subgroup performance on MCAS. In general, white and Asian students reported enrolling in more advanced science and mathematics courses. African American, Hispanic, and Native American students were much more likely to report not being enrolled in any science course in grade 10. (See Table 29.)

Table 29: Percentages enrolled in Science courses, by Racial/Ethnic Group

Percent Currently Enrolled, Science						
	Asian/ Pacific Islander	African American	Hispanic/ Latino	Multiple	Native American	White
Integrated Science	13	19	23	8	17	8
Biology	8	9	7	22	20	8
Earth Science	7	6	6	0	17	4
Chemistry	54	25	15	30	17	53
Physics	4	3	5	0	0	0
Total Enrolled	96	62	56	60	71	73
Total Not Enrolled	4	38	44	40	29	26

Mathematics course enrollment percentages were higher than science for all groups except Native Americans. (The very low numbers of students reporting themselves as Native Americans makes results for this subgroup statistically unstable, and should be interpreted with considerable caution.) However, a sizable proportion of some minority groups reported not being enrolled in any mathematics course in grade 10. (See Table 30.)

Table 30: Percentages enrolled in Mathematics courses, by Racial/Ethnic Group

Percent Currently Enrolled, Math						
	Asian/ Pacific Islander	African American	Hispanic/ Latino	Multiple	Native American	White
Integrated I/ Algebra I	11	16	23	0	17	8
Geometry/ Integrated II	51	72	56	68	50	57
Algebra II/ Integrated III	23	9	8	11	0	21
Advanced	18	6	1	5	0	16
Tot. Enrolled	103	103	88	84	67	102
Total Not Enrolled	-	-	12	16	33	-

When courses are grouped into "lower" and "upper" in an academic sequence, the disparities are even clearer. (See Table 31.) In Science, about 53% of the white students are enrolled in Chemistry or Physics; the large majority of students in these courses report haven previously completed biology. About half as many minority students (with the exception of Asian students) report being enrolled in these courses, and a much higher proportion of minority students do not report being enrolled in a science course in Grade 10 at all. Asian students report a much lower rate of not being enrolled in any science course, and a higher rate of being enrolled in the core discipline courses.

Table 31: Percentages of students enrolled in "lower and upper" science courses, by racial/ethnic group

Percent Currently Enrolled in "Beginning" and "On-Sequence" Courses, Science						
	Asian/ Pacific Islander	African American	Hispanic/ Latino	Multiple	Native American	White
Not enrolled[1]	4	38	44	40	29	26
Beginning[2]	21	28	30	30	37	16
On-sequence[3]	58	28	20	30	17	53

[1] Student did not report being currently enrolled in any of the science courses identified in the student questionnaire

[2] Integrated Science or Biology course

[3] Chemistry or Physics course

The pattern in mathematics was similar to that in Science. (See Table 32.) Students in all racial/ethnic groups reported being enrolled in mathematics courses more than science. However, Asian students enrolled at a higher rate in upper mathematics courses, while the other racial/ethnic minority students enrolled in upper mathematics courses at a rate of much less than one-half that of white students. Correspondingly, minority students (other than Asian students) reported being enrolled in lower math courses two to three times more frequently than did white students.

Table 32: Percentages of students enrolled in "lower and upper" mathematics courses, by racial/ethnic group

Percent Currently Enrolled in "Lower" and "Upper" Courses, Math						
	Asian/ Pacific Islander	African American	Hispanic/ Latino	Multiple	Native American	White
Currently enrolled[1]	103	103	88	84	67	102
Lower[2]	11	16	23	0	17	8
Upper[3]	41	15	9	16	0	37

[1] Percent of responses indicating current enrollment in a course on the student questionnaire; may sum to more than 100% due to multiple enrollments.

[2] Integrated math or Algebra I.

[3] Algebra II/Integrated III or Advanced math, including trigonometry, pre-calculus, calculus, or discrete mathematics.

A regression analysis showed that course enrollment contributed more to predicting MCAS science scores than did race/ethnicity, after performance on the MAT7 was considered. (See Table 33.) Including race/ethnicity did not contribute to prediction significantly either on its own ($\Delta r^2 = .01$) or in combination with courses ($\Delta r^2 = .00$).

Table 33: Relationship of Ethnicity and Courses to MCAS Performance, Science

MCAS performance	Commercial test control	Other factors	R-square	Change in R-square
Sci. & Tech.	MAT7 Science		.55	
Sci. & Tech.	MAT7 Science	Biology	.60	.04
Sci. & Tech.	MAT7 Science	Biology, Geometry	.61	.05
Sci. & Tech.	MAT7 Science	Race/Ethnicity	.56	.01
Sci. & Tech.	MAT7 Science	Biology, Geometry, Race/Ethnicity	.61	.00

Neither courses nor ethnicity significantly added to the predictability of MCAS high school mathematics scores over what was predicted by performance on the MAT7 Math. As with science, race/ethnicity was not a significant factor statistically, when courses were considered.

The pattern for MCAS mathematics performance shows no strong effects either for courses or race/ethnicity, beyond the commercial test. Again, the significant result is that ethnicity was not found to be a factor in MCAS more than in commercial tests.

Table 34: Relationship of Ethnicity and Courses to MCAS Performance, Mathematics

MCAS performance	Commercial test control	Other factors	R-square	Change in R-square
Mathematics	MAT7 Math		.68	
Mathematics	MAT7 Math	Race/Ethnicity	.68	.00
Mathematics	MAT7 Math	Integrated/ Algebra I, Geometry	.69	.01
Mathematics	MAT7 Math	Integrated/ Algebra I, Geometry, Race/Ethnicity	.69	.00
Mathematics	MAT7 Math	Biology, Geometry	.70	.02

The fact of no statistical increase does not mean that there are not real practical differences. Merely reporting having been enrolled in a course does not necessarily mean the student is well-prepared. There are large differences in MCAS performance, even for students that claim to have taken a similar course. For example, when performance was analyzed for students who said they had already completed a biology course, white students got *Failing* ratings half as often, and scored *Proficient* scores two to three times more often than students in other racial/ethnic groups. (See Table 35.) The phenomenon of racial/ethnic group preparation through academic courses may be better understood through examining whether there are different courses, different learning within the courses, self-report differences, or other significant variables.

Table 35: Percentages of students at "Failing" and "Proficient" MCAS levels in science, who completed biology, by racial/ethnic group

Percentage of students who reported having completed a biology course						
MCAS Science Level	Asian/ Pacific Islander	African American	Hispanic/ Latino	Multiple	Native American	White
Failing	19	22	45	25	-	8
Proficient	8	22	9	13	-	34

The purpose of this analysis is to indicate that the evidence supports the validity of MCAS. For example, course enrollment data provide some reasonable explanations about large differences in subgroup performance. Such an examination of MCAS and related data can provide districts, schools, parents, students, and other policy makers with valuable insights related to how to account for scores on the MCAS. More importantly, educators may understand better how to improve schools as places of learning for the students who attend. Then the purpose of the MCAS will be fulfilled.

Discussion and Recommendations

It is a credit to the educational institutions and students that the results turned out as strongly as they did. Often in the first year administration of a new state program the student performances are weakened because of lack of familiarity, engagement, or other factors.

These results support the MCAS as a valid assessment. Performance on the MCAS is appropriately related to performance on familiar, commercially available tests. The MCAS reports in alignment with the state's curriculum framework and desired student performance standards. MCAS is comparable to that on commercial tests in terms of differential performance by gender and racial/ethnic groups. MCAS appears sensitive to student transience and course-taking, as would be expected.

These results and those from the companion study of another district are encouraging at the beginning of this new testing program. However, additional work must be done to continue to establish the validity of the MCAS program. Additional areas of validity work are discussed below.

While the results from this one district are provocative, **these analyses should be extended to a state-wide sample**. The nature of the cooperative school district indicates that more districts and a more heterogeneous sample will be needed to establish the validity of the MCAS program for all the schools across the Commonwealth. It was recognized that student attrition between years, difficulties in matching the student identification number across data files, inconsistent predictor measures, and collaborating districts that are not entirely representative of the state severely constrain the generalizability of the study. However, it was felt that the modest efforts required to establish this "first look" would be useful, both in documenting the beginning of the MCAS system, and in informing future efforts to design more systematic validity studies. Districts, the Department, and interested educational organizations should take this opportunity to plan a coherent validation agenda. In particular, documentation and evaluation of instructional impact is complex and expensive, yet invaluable. That should be undertaken immediately to ensure valid looks at the system as it is implemented and evolves over time. It is especially needed since one of the recognized criteria for tests with high student takes—such as the MCAS graduation requirement—is that students (and therefor teachers and administrators) have had sufficient time, information, and opportunity to prepare adequately.

In particular, we recommend that the issues of **transience and differential achievement** be closely examined, after validating the mobility data. Transience is not considered in the proposed school accountability scheme, but this study suggested that it could have a significant effect. Schools typically say that they have very little control over student transience, and can do little to remediate its effects on learning. This issue should be thoroughly investigated, including relation to SES (socio-economic status) and ethnicity.

In general, the student self-reported variables—frequency of study, curriculum coverage, ethnicity—should be validated. This could be done in conjunction with cooperating districts' existing databases. Curriculum and student engagement are difficult to assess accurately. Studies involving surveys, observations, and other methods to get at qualitative as well as quantitative aspects of student experience are necessary.

No data on student economic status were available for this study. However, economic status should be looked at in the future, as that has persistently been found in research to be highly related to student achievement.

It is highly recommended that the state follow up on issues of **ethnicity, course-taking patterns, and eventual achievement**. With the pending graduation requirements, it will be necessary to address whether students have had sufficient opportunities to learn. Differential impact by ethnicity is a potentially serious issue that must be addressed. Unlike transience, schools have a greater degree of control over various aspects of this, including quality of courses, advisement, adjusting prerequisites, course offerings, communication and research about likely "payoffs" of certain course patterns, and so on.

It would be possible to analyze results for subcomponents of **MCAS (multiple choice, open-ended, short-answer, and essay writing)**. It will be important to establish their validity and effectiveness.

Similarly, these results come from one point in time, at the very beginning of the MCAS program. It would be expected that results would **change over time** as students and educators become more familiar with the MCAS assessment program. A similar analysis should be done in the future to document longitudinal effects. It will be important to monitor the changes (e.g., gains) in the various measures. For example, how quickly will a district's norm-referenced test scores change without state accountability, how quickly will its MCAS scores change, and how will that be interpreted and used?

Similar analyses will be appropriate to establish the validity of the **other content areas** (e.g., social studies) as they enter the operational assessment and accountability systems.

This study focused on possible differences between subgroups at the district level. However, many results in school reform will first be manifested at the **school level effects**. Improvement for girls, students of color, and economically disadvantaged students should be monitored at the school level for the purpose of identifying schools with programs that make an unusually large positive difference. These schools should be identified and studied so the powerful aspects of their programs can be made available to other schools.

The **districts have a key role** in school accountability and improving learning. Studies might appropriately look for causal relationships (and effective models) in district resource allocation, district professional development and policies encouraging sound teacher practices, district use and interpretation of reports and results.

The accountability system is currently being designed, but many validity issues can be anticipated regarding **accountability decisions**, such as their relation to student demographic variables.

Research impact of **accountability** on school instruction and organization. School accountability is intended to lead to improvement in student learning. It is reasonable that school improvement will include changes in instruction, curriculum, and school organization that can be empirically studied. In terms of test validity, it is also important to document such things as whether instruction, inappropriately geared to the test, results in narrowed curriculum (and whether this is related to open-ended, multiple choice, or writing formats).

Appendix

MCAS Student Questionnaire

Copies are provided of the Student Questionnaire for grades 4 and 10.