Executive Summary of Two Companion Validity Studies Regarding MCAS

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

By B. Gong

And

Relationships Between MCAS and SAT-9 for One District in Massachusetts

By A. A. Thacker and R. G. Hoffman

This research analyzes aspects of the validity of the Massachusetts Comprehensive Assessment System (MCAS), the 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 research examined the relationships between student performance on MCAS and performance on commercial, standardized tests that had been administered for the same students in two districts.

Research was completed in each of the two districts separately, but cooperatively. The two reports, one from data collected in the Worchester district, the other from Boston, are sufficiently related to be considered companion pieces. However, differences in testing times and formats for the two districts made separate analysis necessary. In Boston, students took the Stanford Achievement Test, 9th edition published by Harcourt Brace Educational Measurement (SAT-9), in grades 4, 8, and 10, the same year they took the MCAS. In Worchester, students took the Metropolitan Achievement Test, 7th edition, also published by Harcourt Brace (MAT-7), in the 5th grade and the SAT-9 in the 9th grade.

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 student's course-taking patterns and self-reports of studying specific topics or frequency of addressing a subject at school.

The initial analysis contained in both these reports supports the view that MCAS is a valid assessment. The following conclusions are also supported:

- **MCAS is challenging**. The MCAS proficiency levels represented notably high Students at or above the proficient level on MCAS were generally above the 75th percentile on SAT-9. Massachusetts is only beginning to establish and assess the curriculum frameworks, so these initial low scores on MCAS are to be expected and are not necessarily inappropriate.

- **MCAS is correlated with SAT-9 and MAT-7**. Student performance on the MCAS was related to performance on familiar, norm-referenced tests. Correlations are high, but not so high as to indicate that MCAS, SAT-9 and MAT-7 are measuring exactly the same

domain. An appropriate, moderate relationship was found between performance on MCAS and performance on the other tests.

- **Gender differences on MCAS are slight and are similar to differences on SAT-9 and MAT-7.** Consideration of gender did not appreciably affect the relationships between performance on MCAS and the commercial tests. The similarities and differences in student achievement could not be attributable to consideration of gender.

- **Ethnic differences on MCAS are similar to SAT-9 and MAT-7, but exhibit minor differences for some ethnic groups**. While these results are not completely consistent across the two districts, there is some evidence that differences between scores for black, white, and Hispanic students were slightly larger on MCAS than SAT-9. This tendency was mild, but reasonably consistent across content areas and grade levels in Boston. The tendency was even less apparent in Worchester. Further studies will be required to establish if this pattern is consistent across the state or anomalous to a specific district. These districts are comprised of a particularly ethnically diverse population, so it is possible that another factor, associated with ethnicity, could be causing these results. For instance, school-level factors may play a large role depending on the level to which the school has adopted the Massachusetts curriculum frameworks and on the specific proportions of the differing ethnic populations served by the school.

- **Transience is negatively related to scores on MAT-7, SAT-9, and MCAS**. Data regarding transience was not sufficient to establish any differences between the tests, however, some unexpected or anomalous results suggest that further research is needed in order to fully evaluate the effect of transience on MCAS.

- **Student course taking patterns, especially in mathematics, are related to student scores on MCAS and the commercial tests**. Students who reported taking higher level classes tended to outperform students who did not. This trend held true for mathematics and science in 8^{th} and 10^{th} grade, although some anomalous results with regard to science courses will need to be clarified with the districts. Fourth-grade students' reports of the time they spent on specific topics were also related to scores on MCAS and the commercial tests. Mathematics courses and time reports were most clearly related to test scores. Elective courses, such as foreign language, arts, and technology courses were also related to higher test scores.

- **Some curious results found during this data analysis will require either more information from the district, a statewide data analysis, and/or a different research approach to adequately explain.**

These results, although based on a limited sample of students and districts, suggest that MCAS is a reasonable measure of student academic ability. It was found to exhibit high standards for student performance and is the only test option designed to evaluate whether students are meeting the Commonwealth's expectations as described in the curriculum frameworks.

RELATIONSHIPS BETWEEN MCAS (MASSACHUSETTS COMPREHENSIVE ASSESSMENT SYSTEM) AND COMMERCIAL STANDARDIZED TESTS FOR TWO COLLABORATING DISTRICTS: A SUMMARY OF FIVE IMPORTANT ISSUES

FEBRUARY, 1999

Prepared for

Advanced Systems in Measurement and Evaluation

and

The Massachusetts Department of Education

by

Arthur A. Thacker and R. Gene Hoffman
Human Resources Research Organization
295 W. Lincoln Trail Blvd.
Radcliff, KY 40160

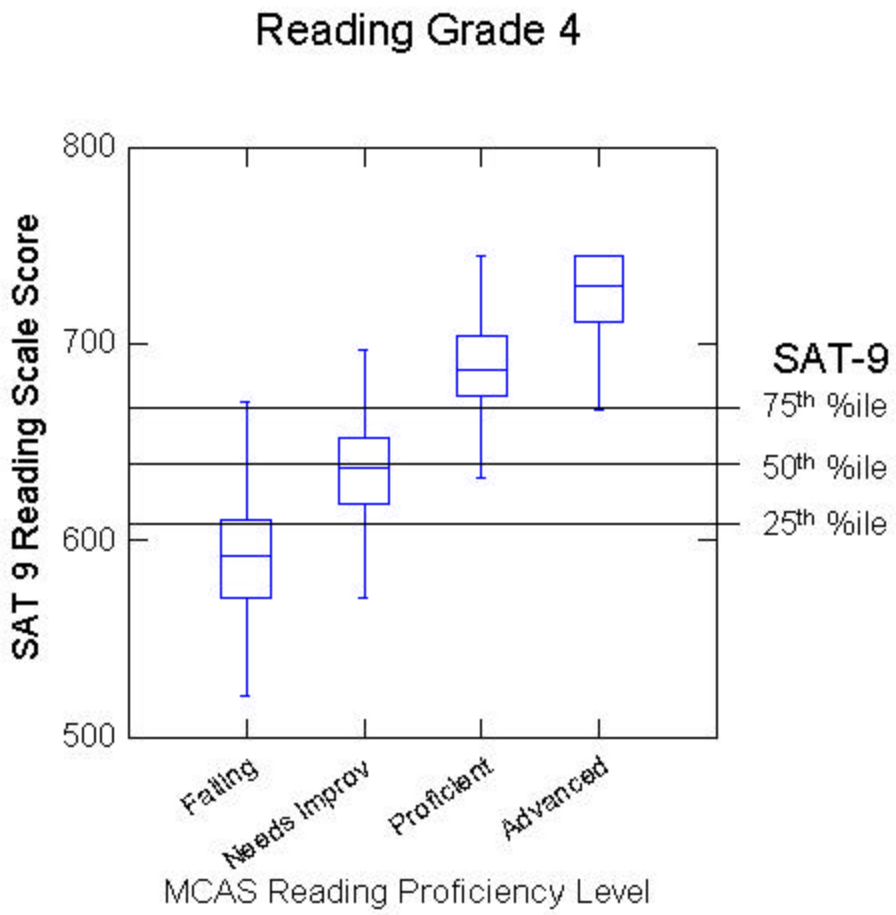
and

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

The comments in this report are the authors' and are meant solely for the consideration of Advanced Systems of Measurement and Evaluation and the Massachusetts Department of Education (MDE). They do not represent policy positions of MDE.

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	SAT-9 Reading Scale Scores				Percentile	
Proficiency Level	Mean	S.D.	1st Quartile	3rd Quartile	of the Mean	N
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