

Investigation of MCAS-NAEP Comparisons and Other External Validity Evidence^{1,2}

Stephen Jirka and Ronald K. Hambleton

University of Massachusetts Amherst

June 30, 2005

¹ Center for Educational Assessment MCAS Validity Report No. 14. (CEA-585).
Amherst, MA: University of Massachusetts, Center for Educational Assessment.

² This work was carried out under a contract between the University of Massachusetts Center for Educational Assessment and Measured Progress and the Massachusetts Department of Education, 2004-2005.

Since the first administration of the MCAS assessments in 1998, several validity studies have been carried out. These include studies to investigate the level of differential item functioning (see, for example, Zenisky, Hambleton, & Robin, 2004), construct validity (Goodman, 2004), consistency of test content across years (O’Neil, Sireci, & Huff, 2000), the validity of the dual language forms (Sireci & Khaliq, 2002), and test content alignment with test specifications and curriculum frameworks (Hambleton & Zhao, 2004, 2005). In this report, we have compiled evidence about the validity of MCAS by considering data that are external to the assessments themselves: From the National Assessment of Educational Progress (NAEP), standardized achievement tests, Achieve Inc. data, and Keep the Promise data. For each area of evidence we have provided the following information: Background and Goals of the Study or Studies, Methodology, Main Results, and Conclusions.

Comparison of Trends in NAEP, Massachusetts-NAEP, and MCAS Results

Background/Goals of the Study

One of the criticisms that is sometimes directed at states is that the academic accomplishment and growth they report from year to year based on their state assessments are inflated because schools narrowly teach to the portion of the curricula they expect to be on the assessments, and students are prepped to maximize their scores by using test-taking skills. Some critics feel that students would not do nearly as well on assessments such as those offered by NAEP (National Assessment of Educational Progress) where the content domains being assessed may be more challenging and testing-taking skills would be less influential in the results. NAEP is the only nationally representative and continuing assessment of what students in the United States know and can do in various subject areas. Assessments in subjects such as reading, mathematics, and science have been conducted periodically since 1969. Massachusetts commissioned this study to compare NAEP results with the state testing results, look for the similarities and differences, and try to explain them.

The purpose of this study by Jirka and Hambleton (2004) was to compare average proficiency levels, proficiency classifications, and trends between NAEP results for the country and NAEP results for Massachusetts, and then NAEP results for Massachusetts and MCAS results. For NAEP, results are available as far back as 1992, one year prior to the passing of the Educational Reform Act in Massachusetts, and for the MCAS, results in ELA and mathematics have been available since 1998. Some cautions in these comparisons were noted by the authors. However, when state results are not generally in line with external indicators of achievement such as NAEP or nationally normed standardized achievement tests, it would be useful to know that, and to try and explain the differences. In some cases, those explanations may have implications for revising the Massachusetts curriculum frameworks and MCAS tests, and other times, they may have no implications at all.

The Jirka and Hambleton study was carried out at the Center for Educational Assessment that is a research and outreach unit in the School of Education at the University of Massachusetts Amherst. It is administratively located within the Department of Educational Policy, Research and Administration. The Center conducts research in the area of educational assessment, carries out service and outreach activities, and strives to fulfill the land-grant mission of the University of Massachusetts.

Methodology

The data used in this study were obtained from the National Center for Educational Statistics website and from the Massachusetts Department of Education website. The NAEP data included the mean scaled scores for each grade and subject area for both the nation and the state of Massachusetts, along with the performance category data. Break outs were also available for the disaggregated groups required under the new NCLB law. These data were collected on all NAEP administrations since 1992 in reading and mathematics where reporting was available for the Commonwealth of Massachusetts. The MCAS data contained mean scaled scores up until 2002 and performance level data for both the entire state and the various breakout group categories required to be reported since 1998.

After the data were collected, they were compiled and various graphical displays were composed to see if changes were evident between the NAEP scores for the country and for Massachusetts. Additionally, scores for MCAS were graphed and data were compared to determine if there were trends across the years and between the various groups. During this process, the NAEP and MCAS performance category descriptors were gathered and compared to determine if roughly the same types of knowledge and skills were required.

Main Results

In comparing student performance in Massachusetts and the rest of the states since 1992 on the NAEP Mathematics test, the findings seem clear: the average Massachusetts student outperformed the average student in the country in 1992 at grades 4 and 8, and maintained that difference (grade 4) or increased the difference (from a difference of 6 points to a difference of 11 points) (grade 8) over the 11 year period. The other interesting observation is that over that 11 year period of NAEP-state assessments, mathematics achievement across the country showed a sizable increase (15 points at grade 4, and 9 points at grade 8). Clearly mathematics achievement has been on the rise in both grades 4 and 8 since 1992, and Massachusetts has kept pace at grade 4 with the rise of mathematics achievement around the country, and moved ahead even further at the grade 8 level.

For Reading, we again find that the average Massachusetts student outperformed the average student in the country, and maintained that advantage over the 11 years of testing. Generally, the trends were flat in reading, though Massachusetts students did show a modest gain (2 points) compared to other states at grade 4, and a moderate gain at

grade 8 (4 points). (A one point gain, is roughly equivalent to a single month of instruction, so a 4 point gain is roughly equivalent to giving students an extra four months of instruction over four years of instruction or about one extra month per year. An extra month of schooling per year seems practically meaningful.)

The shortcoming of these results are, however, that they are reported in terms of scaled scores. Without more benchmarks on the scaled score scale and standard deviations of scores, interpretations are difficult to make because scaled scales are not meaningful with some benchmarks and effect sizes cannot be calculated. What is clear is that scores in Massachusetts and other states are on the increase, and at the grade 8 level especially, and in both Mathematics and Reading, Massachusetts appears to be increasing its advantage over other states.

Conclusions

The NAEP results in Mathematics, at both grades 4 and 8, show clearly that sizeable gains are being made around the country. And, in Massachusetts at grade 8, the gains are even greater. These findings are reflected too in the MCAS results. Over six years of MCAS testing, at grade 4, the percentage of students being classified as Needing Improvement or better increased from 78 to 84 percent, and the percentage of students being classified as proficient or better has increased from 34 to 40 percent. Comparable gains were observed at grade 8 too: over six years, the percentage of students being judged as “Needing Improvement” or better moved from 57 to 67 percent, and the percentage of students being judged as “Proficient” or better moved from 31 to 37 percent. These results are very much in line with NAEP results after adjustments are made for the shorter time period of MCAS assessments.

The NAEP results in English Language Arts/Reading are not as clear. There are moderate increases for grades 4 and 8 through the years. These gains are moderate, but they are gains. Massachusetts has percentages that are above the nation on all grades and performance levels. More dramatic increases are seen for the MCAS results. Grade 8 has steady increases from 86 to 92 percent for Needs Improvement, 55 to 67 percent for Proficient and above, and 3 to 5 percent for the Advance and above percentage of students. More fluctuation is seen in grade 10, but the overall gains are striking. Percentages of students at or above Needs Improvement increased from 72 to 89 percent, from 38 to 61 percent for the Proficient and above category, and from 5 to 20 percent for Advanced. These are gains of 23 and 15 percent, respectively. These results are comparable to the overall gains seen in NAEP over much longer time period.

This study is another piece of evidence that helps to support both the content and construct validity of the use of the MCAS. NAEP is a national assessment system that is considered to be rigorous and representative of most of the content standards of the states. By showing that the general trends of the results of the MCAS match the general trends of a national assessment, we can see that the MCAS is doing what it is supposed to be doing. It is testing students on what they need to know and the skills they should have in order to function in the world after they graduate from high school.

MCAS External Validity Reports

Background/Goals of the Studies

After the 1998 MCAS administration, the Department of Education commissioned two external studies to examine the concurrent validity of the MCAS tests. The studies examined the relationship between the performance of students in two large Massachusetts school districts on the 1998 MCAS tests and a locally administered, national, standardized achievement test. Gong (1999) examined the relationship between MCAS scores and performance on the Metropolitan Achievement Test (MAT-7) at grade 10 and the Stanford Achievement Test (Stanford-9) at grade 4. Thacker and Hoffman (1999) examined the relationship between MCAS scores and performance on the Stanford-9 at grades 4, 8, and 10. The two studies also examined the relationship between MCAS performance and students' enrollment in specific courses.

The study was carried out by Art Thacker and R. Gene Hoffman at Human Resources Research Organization. Founded in 1955, HumRRO is a group of multidisciplinary professionals, with a technically diverse supporting staff with experts in industrial and experimental psychology, management analysis, sociology, operations research, computer science, economics, and educational research. HumRRO conducts research, develops products, and provides services to improve individual, team, and organizational performance. Brian Gong is from the National Center for the Improvement of Educational Assessment, Inc. The Center for Assessment is a nonprofit organization founded in 1998 by Richard Hill and Brian Gong, professionals with combined experience of over 40 years in the assessment community. The Center offers consulting services in many areas related to assessment and accountability, such as the design and analysis of custom assessment and accountability systems, data modeling and analysis, workshops, briefings, documentation, and public relations aimed at the assessment community. The Center focuses on both technical and practical issues that promote the effectiveness of educational assessment and accountability programs.

Methodology

The two studies described were based on the results of individual school districts because the commercial standardized tests (MAT-7 and Stanford-9) were administered by the districts rather than by the state. There is one commercial standardized test, however, that has been administered to students statewide. Since 1996, third grade students in Massachusetts have been administered the Iowa Test of Basic Skills (ITBS) in Reading as the *Massachusetts grade 3 Reading test*. Fourth grade students who completed the 1998 MCAS tests also completed the ITBS reading tests as third grade students in 1997. Although the tests were administered approximately a year apart and only include reading, at this time these two tests provide the only opportunity for statewide comparisons of performance on MCAS and a commercial standardized test.

These reports used available datasets from the department of education to produce descriptive and inferential statistics. Correlations and regression equations were computed to determine the degree of a relationship between variables of concern and the effect of certain variables on others. Tables and graphs that display various statistical results were included in the report along with correlations, R^2 values, and standardized coefficients. The correlations indicated the strength of a relationship. R^2 was used to determine the strength of the predictions and the standardized coefficients in the prediction equations helped to show which variables had more weight in the predictions.

Regression equations were also used to determine the extent to which differences in performance of males and females on MCAS were similar to Stanford-9 scores. Two separate equations were constructed without gender, and then another set with gender. Differences in the weight were then compared.

Main Results

A brief summary of the results from these studies involving 55,000 students is given below. To begin with, the results of the studies indicated that students in each of the MCAS performance levels (*Warning/Failing, Needs Improvement, Proficient, Advanced*) generally performed similarly on a commercial standardized test. That is, students at higher performance levels on MCAS tended to perform better on the standardized test. On the Stanford-9, students who scored *Proficient* or *Advanced* on MCAS tended to score above the 75th percentile on the standardized tests. Students who scored at the *Needs Improvement* level on MCAS scored around the 50th percentile, and students whose MCAS performance was at the *Warning/Failing* level averaged consistently below the 25th percentile on the standardized test. Both studies found that there was a fairly strong relationship between the courses students reported taking and the students' MCAS scores. That is, students who reported taking higher level courses tended to perform better on MCAS. This relationship was strongest in mathematics.

Conclusions

The two studies revealed a strong relationship between student performance on the state assessment (MCAS) and a commercially available assessment (ITBS). Consistent with the results of the individual district results, students who scored at higher performance levels on MCAS tended to score at higher percentile ranks on the ITBS. These studies help demonstrate both construct and content validity of the state assessment system.

The results of these two studies are encouraging, but additional studies need to be done. The content validity of the tests needs to be examined every year, and studies examining consequential validity should be conducted to determine its impact on schools, districts, and the students themselves. Additionally, gathering evidence of concurrent validity from a larger sample of school districts, comparing the results of the grade 8 Science and Technology/Engineering test with the Third International Mathematics and

Science Study, and gathering information about the intended uses of the MCAS are all validity studies that should be considered.

Achieve Report on MCAS

Background/Goals of the Studies

During the spring and summer of 2001, Achieve conducted an evaluation of the state's K–12 mathematics standards and grade 10 Massachusetts Comprehensive Assessment System (MCAS) tests in English language arts and mathematics. The state's English language arts standards were not analyzed because Achieve believes these are already among the best standards in the nation and uses them as “exemplary standards” against which other states' standards are compared.³

The goals of this study were to determine how Massachusetts' mathematics standards compare with those of high-performing states and nations and to determine if the expectations for schools and students are rigorous yet reasonable. Additionally, the state wanted to determine how well the grade 10 MCAS tests in English language arts and mathematics measure the knowledge and skills in the curriculum framework standards and how challenging the assessments are.

Created in 1996 by governors and business leaders to serve as a clearinghouse and resource center on education standards, assessments and accountability, Achieve, Inc. provides states with candid feedback on the quality of their academic standards, assessments, accountability systems and other policies to promote high academic achievement. Achieve has analyzed the policies and expectations of over 10 states and is currently is working with several others.

Methodology

Achieve uses a protocol that is implemented by its experts to evaluate a state's set of standards. The technique it uses is called benchmarking, where the state's standards are compared to an exemplar. The exemplar used for English language arts is actually the Massachusetts standards, and that is why they were not analyzed for this report. For the analysis of mathematics, Achieve used standards from other states and nations as exemplars. Achieve does not believe that evaluating the standards results in one number that determines if the state standards match or not, but that there is an overall evaluation that takes many factors and pieces of information into account. Achieve compares a state's standards to state, national, and international benchmark standards recognized for their quality and or for producing high student achievement utilizing a commissioned panel which determined these exemplars. These exemplars currently include Texas, North Carolina, Arizona, Japan, and Massachusetts.

³ Achieve, Inc. (October 2001). *Measuring Up: A Standards and Assessment Benchmarking Report for Massachusetts*. Washington, D.C.

The alignment involves confirmation of the test blueprint, a comparison between the content of the item and the standard, and an analysis of the cognitive task required by the item and the standard. Also taken into account are a determination of whether doing well on a strand of items requires students to master challenging material and analysis of the balance and range of the items.

The process first entails a systematic comparison between the standards and the exemplars. Next, experts review these using a protocol with a set of guiding questions followed by two other panels that were convened to study the tests using the alignment protocol. Finally, the staff at Achieve synthesizes the results from the various reviews in to a final report. The report for Massachusetts was completed in October of 2001.

Main Results

According to the final report, Massachusetts has made substantial progress in developing and implementing the two essential components of standards-based reform, namely having strong standards and assessments that measure what the standards expect. More specifically, they found that, overall, Massachusetts' standards and high school tests are of high quality and are aligned, and, accordingly, this provides a solid foundation on which to build the state's education policy. Alignment of the grade 10 MCAS tests are "rigorous and generally well aligned with the standards." This ensures that students in Massachusetts are required to demonstrate important knowledge and skills before graduating from high school, one of the major goals of the No Child Left Behind Act. Overall, the mathematics standards represent an articulate statement of what students should minimally know and be able to do by the time they complete high school. As noted earlier, the English language arts standards are among the best in the country and are used as an exemplar when analyzing the standards of other states.

Achieve also found that the grade 10 tests are rigorous yet reasonable for the students and noted that the MCAS test are the most challenging of the tests that have been reviewed by them. The MCAS tests measure the important knowledge and skills demanded by the standards, are technically sound, encourage high levels of performance, and provide a template for effective classroom instruction. Releasing of all common items (items taken by all students at grade 10) each year is an exemplary strategy that enhances not only educational practice, but also the credibility of the state's educational improvement efforts. The majority of states do not release their common items. Achieve felt that many students should be able to pass these tests by the end of the 10th grade, and other students should be able to pass them by the end of high school, given sufficient support.

While Achieve liked many aspects about the mathematics standards, they did note some ways they could be strengthened. The standards are well-organized, clear, and precise, but they should be strengthened by placing more emphasis on developing students' conceptual understanding of mathematics. At this point, they are reasonable minimum competency standards and are generally comprehensive. Achieve noted that the standards are not yet as challenging as standards from Japan, Arizona, or Achieve's

Mathematics Achievement Partnership, of which Massachusetts is a founding partner state.

Conclusions

The main conclusions of the report are that both the standards and the assessment of Massachusetts are good. Achieve did recommend two main improvements that Massachusetts should consider in its assessment program. These recommendations are outlined below.

First, the MCAS high school tests include relatively minor flaws that should be fixed in subsequent rounds of testing. For example, the math test emphasizes some standards at the expense of others, and the English language arts test needs more treatment of nonfiction. The MCAS mathematics test measures important content that all high school students should be responsible for knowing, yet it can be improved further. Achieve observed that, while the test is generally well constructed, the balance of items does not provide adequate coverage for all of the important knowledge and skills detailed in the standards. The state should ensure that items assessing number sense and operations and data analysis are appropriately challenging on the next edition of the MCAS mathematics test. The state has already taken this into account and is working to ensure that the entire breadth of the standards are tested.

While the English language arts test is also strong and rigorous for students, it focuses too much on literature. Achieve believes that all students should be exposed to a deep and engaging literary curriculum, but they also should learn how to read and interpret informational texts, such as historical documents, scientific journal articles, and technical manuals. Achieve also pointed out that Massachusetts should consider varying the 10th-grade writing prompt from year to year to assess students' skills in producing different kinds of compositions, rather than just literary analyses. This along with including more informational texts will provide students with more of the skills they will need to participate meaningfully in the emerging "knowledge economy."

The second improvement involves the mathematics standards. Achieve believes that they should move beyond being just minimal requirements by requiring a little more rigor and depth. They should emphasize mathematical reasoning more and focus more on the essential content at each grade level. In order to be prepared for success in college and high-performance workplaces, students will need higher expectations. Achieve acknowledged that many Massachusetts schools and students are struggling to meet the state's current standards, but still believes the state should raise its mathematics standards. They added that the state may wish to publish companion materials to the mathematics standards that include numerous sample problems, activities, and descriptions of how educators can build students' conceptual understanding of mathematics and reasoning ability.

The implications of this report bode well for the Commonwealth of Massachusetts. As stated previously, the standards for English language arts are among

the best in the nation and world and are used as an exemplar when evaluating other states' standards. While the mathematics standards are not at that level, they are reasonable and achievable for students. With some minor modifications, they will be strengthened and will help ensure that the state maintains its status as having one of the premier assessment programs in the nation.

Keep the Promise

Background/Goals of the Studies

Keep the Promise is a three-year longitudinal study of at-risk high school students, involving Massachusetts' three largest urban school districts (Boston, Springfield, and Worcester) produced by Mass Insight Education. This organization is sponsored by business leaders in Massachusetts.

Incorporating both qualitative and quantitative data collection and analysis, Mass Insight hopes to address three main research objectives over the next three years. The researchers want to look at three areas involving student characteristics and behavior, remediation program design and effectiveness, and outreach and participation. More specifically, they aim (1) to track the behavior, experiences, and perceptions of students who are in the classes of 2003, 2004, and 2005 and who need extra academic help to develop the skills required to pass MCAS and earn a Massachusetts high school diploma; (2) to catalog and begin to evaluate the effectiveness of high school remediation programs with an aim toward identifying promising practices; (3) to gauge the quality and effectiveness of related outreach efforts to students and their families. Each one of these will be dealt with in the following paragraphs.

For student characteristics and behavior, the researchers want to answer questions that will help give a clearer understanding of students in the state over the next three years. They want to determine the profile of students who took MCAS 10th grade exams more than once before passing (or who never passed them), and how it compares to the profile of students who passed. They also want to find out to what extent students who fail to pass MCAS by their senior year continue to work toward passing MCAS and obtaining their high school diplomas. Additionally, they want to know what steps these students take in pursuit of their objective. Finally, they want to answer the question: How do attitudes and behaviors of at-risk students change, if at all, as the MCAS requirement becomes increasingly embedded in the school landscape?

For remediation program design and effectiveness the researchers want to know what program characteristics appear to deliver the best results, and whether there are certain types of programs more appropriate for some student populations than others.

For outreach and participation the researchers want to know if there are particular characteristics or experiences (e.g., demographic, academic) that make a student more or less likely to participate in voluntary remediation programs. They would also like to know what other factors enhance or diminish the likelihood that a student will utilize

academic support services (e.g., motivation, family pressures, practical or economic incentives/disincentives). Finally, they want to know the characteristics of programs and outreach efforts that do the best job of getting students to actively participate.

Methodology

The main methods utilized in the production of these reports were surveys, interviews, and analysis of existing data sources. *Seizing the Day* used an interview protocol that was initially developed by the UMASS Donahue Institute and shared with staff at Mass Insight and other research collaborators. It was refined by pilot studies with potential interview candidates. The target population were students in graduation years 2003 and 2004 who failed one or both parts of the 10th grade MCAS. They were invited by letters sent to parents, and all eligible students who met the criteria were invited. Ultimately, 134 students were interviewed.

For the reports *Beyond Tests and Good Intentions* and *What We Know Now*, the data offered in these first-year reports is limited strictly to ASSP-funded program data collected during the FY'02 (school-year and summer) and the FY'03 school years. Subsequent years of research will review all state and locally funded programs. Available MCAS and retest data is limited to the spring 2001, fall 2001, spring 2002, and fall 2002 administration results. The subgroup analysis presented in this report uses Student Information Management System (SIMS) data and MCAS student survey responses provided by the Massachusetts Department of Education.

For data comparison, observed differences were determined to be statistically valid or are possible due to chance by using Chi-Square or two sample t-tests. For all tests, a 95% confidence interval ($p < .05$) is an acceptable standard for determining statistical significance and is the basis for all claims of significance (or lack thereof) within these reports. If a statistical test demonstrates that the differences in responses between two subgroups are significant, then, in the simplest terms, such differences are likely not due to chance.

Main Results

There are three key findings in *Seizing the Day*. First, the reality of high stakes for high school students has led to increased effort and improved school behavior. Four out of five students who need extra help now take the steps to get it. Eighty-two percent of students who did not pass the first year now participate in remediation programs. Juniors are committing earlier to passing MCAS. A higher percentage of juniors now seek help than previous juniors. Students who are at-risk are putting more effort into their school work, according to the survey data. They pay attention more in class, spend more time on homework, and are absent less often. Most of the participants in these extra help programs are doing so during the day. Second, students rate remediation programs and teachers highly and say they play critical roles. Students report high levels of satisfaction with extra help programs and credit remediation programs with their retest success. Third, students receive mixed messages about their academic performance and

preparedness for postsecondary education. There appears to be a disconnect between classroom standards in large urban districts and the state's minimum achievement expectations. Most of the students who failed the MCAS the first time report having a C or higher. The students' perceptions of their own preparedness are inconsistent with their MCAS scores and college's expectations. Most of the students who fail plan on continuing their education in 2 or 4-year colleges, but professors surveyed say that the high school students they encounter are not prepared. Students are not clear about the rationale of the MCAS and receive mixed messages from parents and teachers. Some students believe they should receive a diploma for attending all their classes, and these students may be hearing negative perceptions of the test from parents and teachers.

Beyond Tests and Good Intentions has three main findings. First, through FY 2003, a wide array of academic support programs offered targeted, often individualized attention to at risk urban students on a scale never attempted before. The emphasis of most of these programs is on providing intensive small-group or individual instruction and innovative programming to these students. Second, educators have worked to adapt programs to increase student participation, and this has led to an emphasis on in-school remediation programming for at-risk students. These programs usually entail a stand-alone MCAS preparatory class in addition to the students' regular classes. Third, the three districts have adopted strategies that overlap in some ways but differ in others, and understanding the impact of these choices is critical in the current cycle of severe financial constraint. Some supplemented state funding with local funding, and one district has implemented a centralized approach with the inclusion of a full time MCAS specialist.

What We Know Now found many useful findings by analyzing student records. In general, the researchers found that at risk students are participating in extra help and are likely to pass a retest. The students have a positive perception of these programs. Programs that attract and retain all eligible students, target support to the student's weakest areas, are linked to the regular school day are better. Students who have poor attendance also struggle with passing the MCAS. Certain subpopulations such as LEP, special education, and minority students (generally), have done less well on the MCAS. However, students generally say they are trying harder in school, despite receiving mixed messages about the MCAS, and a mismatch between grades and test scores.

Conclusions

The information presented in the three, first-year Keep the Promise reports provides a good impression of the current state of educational reform for three large urban districts in Massachusetts, and this first year research suggests a confluence of several factors has produced the fairly astonishing 96% graduation rate achieved by the Class of 2003. These factors include: (1) improved curriculum and alignment with standards, (2) remediation — in school, before school, after school, and in summer, (3) increased motivation as students neared the spring of their senior year, (4) test-taking comfort as students became more familiar with the tests.

The appraisal of Boston, Worcester and Springfield’s academic support programs presented in all three first-year Keep the Promise reports does not offer data-supported conclusions about the most effective forms of remediation. However, the answers to those questions should come over time as Keep the Promise moves into its second and third years. Mass Insight posed several important questions about high school remediation as continuing areas for research. These questions include:

- At the district level, what are the impacts of different system-wide strategies with respect to remediation program design? Do the advantages of a centralized district approach outweigh the flexibility of permitting high schools latitude to design their own programs?
- At the school level, how can educators successfully reach and serve the students who are furthest behind, judging by their initial tenth-grade MCAS scores — those who represent a disproportionately high percentage of the 1,033 students who, at this writing, have not earned their CD across all three of these districts?
- At the remediation program level, what are the design features — technology? Low teacher-student ratio? incentive tie-ins with jobs outside of school? — that most effectively promote increased motivation and achievement?

References

- Gong, B. (1999). *Relationships between student performance on the MCAS (Massachusetts Comprehensive Assessment System) and other tests—collaborating district A*. Dover, NH: The National Center for the Improvement of Educational Assessment, Inc.
- Goodman, D. P. (2004). *A multitrait-multimethod validity investigation of 2002 Massachusetts Comprehensive Assessment System tests* (Center for Educational Assessment Research Report No. 531). Amherst MA: University of Massachusetts, Center for Educational Assessment.
- Hambleton, R. K., & Zhao, Y. (2004). *Alignment of MCAS grade 10 English language arts and mathematics assessments with the curriculum frameworks and the test specifications* (Center for Educational Assessment Research Report No. 538). Amherst, MA: University of Massachusetts, Center for Educational Assessment.
- Jirka, S. J., Hambleton, R. K. (2004). *Comparison of trends in NAEP, Massachusetts-NAEP, and MCAS results*. (Center for Educational Assessment Research Report No. 537). Amherst, MA: University of Massachusetts, Center for Educational Assessment.
- Mass Insight Education. (2003a, October). *Seizing the day: Massachusetts' at-risk high school students speak out on their experiences at the front lines of education reform*. Boston, MA: Mass Insight Education.
- Mass Insight Education. (2003b, October). *Beyond tests and good intentions: What the academic "ER" looks like in Boston, Springfield, and Worcester*. Boston, MA: Mass Insight Education..
- Mass Insight Education. (2003c, November). *What we know now: Early findings and important questions about urban high school remediation in Massachusetts*. Boston, MA: Mass Insight Education.
- O'Neil, T. O., Sireci, S. G., & Huff, K. L. (2000). *Evaluating the consistency of test content across two successive administrations of a state-mandated science assessment* (Center for Educational Assessment Research Report No. 459). Amherst, MA: University of Massachusetts, Center for Educational Assessment.
- Sireci, S. G., & Khaliq, S. N. (2002). *An analysis of the psychometric properties of dual language test forms* (Center for Educational Assessment Research Report No. 458). Amherst, MA: University of Massachusetts, Center for Educational Assessment.

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

Zenisky, A. L., Hambleton, R. K., and Robin, F. (2004). DIF detection and interpretation in large-scale science assessments: Informing item-writing practices. *Educational Assessment*, 9(1&2), 61-78.