

Whether high-stakes standardized testing should serve as the leverage point for school reform in Massachusetts and the U.S. was the focus of a conference at the John F. Kennedy School of Government at Harvard University in October 2001.

“Testing Testing: School Accountability in Massachusetts and Beyond” – sponsored by the Rappaport Institute for Greater Boston and the Program on Education Policy and Governance at the Kennedy School brought together national experts and key education officials in Massachusetts and Boston.

Paul E. Peterson, director of the Program on Education Policy and Governance, began the conference by distinguishing between two approaches to school reform. Bottom-up reform relies on vouchers, charter schools, and other forms of school choice. Top-down reform sets strong statewide standards and requirements and impels local districts to meet those standards. MCAS, the Massachusetts test administered in the fourth, eighth, and tenth grades, lies at the center of Massachusetts reform efforts. By requiring students to pass the tenth-grade MCAS test to qualify for graduation, the Commonwealth wields a strong “stick” to go along with the “carrots” of \$2 billion in additional funding since the passage of the Education Reform Act of 1993.

Charles C. Euchner, executive director of the Rappaport Institute, provided a summary of the history of testing in Massachusetts and the general issues that surround high-stakes testing nationwide. Borrowing from Albert O. Hirshmann’s class argument in *Exit, Voice, and Loyalty*, Euchner noted that consumers have three choices when they are dissatisfied with a public good like education. They can either exit to another provider (a private school or another district), use their voice to agitate for better service, or remain committed to the existing service whatever its flaws. Euchner asked: “In this case, what is the proper balance of these three in terms of making schools efficient and accountable? Like businesses, schools need accountability for when people take the exit option and go elsewhere for education. Or perhaps families need a better voice? Or people need more loyalty to them?”

There are four key approaches to school reform:

- **Capacity-building:** Providing schools with more resources for smaller class sizes, teacher training, and curriculum development.
- **Standards-based reform:** Mandating curriculum frameworks, use of diagnostic tools, higher levels of teacher training.
- **Performance-based reform:** Requiring students, teachers, and their schools to demonstrate a certain level of mastery on tests and other systems of evaluation.
- **Market-driven reform:** Providing choice to students and their parents through vouchers, charter schools, and other kinds of school choice.

Education reform in Massachusetts is a combination of the second and third approaches. Under the 1993 law, the state spends more money to promote equity in school funding in exchange for rigorous testing, clear graduation standards, and remediation when necessary. The reform also focuses on improved teacher training and more authority for principals. Since much of the effective schools literature focuses on leadership at the school level, principals are the most important link in the whole school reform process.

MCAS has put a range of questions on the front burner in Massachusetts: How does the MCAS affect use of school resources? How does the MCAS affect the structure and management of schools? Are the testing conditions fair? Is the test too hard – or too easy? Why do Massachusetts students do well on national exams but do poorly on the MCAS? Are schools “teaching to the test”? If so, is such an approach bad?

Ultimately, Euchner argued, the Commonwealth and its cities and towns need to decide what they consider the proper goals of public schooling. Is it to provide a broad liberal arts foundation for all? Is it training students for the jobs of a new economy? Or is it preparing a diverse student body for a variety of unknowable life challenges?

Critiquing high-stakes testing

Walter Haney, Professor of Sociology at Boston College, argued against teacher testing as the centerpiece of school reform. Haney argued that typing test results to graduation is a misuse of testing. Decades of research on college entrance exams have shown that test scores alone are poor predictors of college success; using tests along with other criterion like essays and portfolios offers a fuller portrait of a student's abilities and promise. News reports have documented inconsistencies and errors in test scoring. Research also shows wide variability in individual and school test scores from year-to-year, indicating imperfections in tests' ability to measure achievement.

Haney then disputed the "Texas Miracle," the claim that high-stakes testing has produced better test scores, fewer dropouts, and higher overall achievement. Haney's analysis of the longterm progress of students from the sixth to the ninth to the 12th grade found significant drops in the number of students.

Haney also criticized the MCAS itself. He said the cutoff score of 220 for passing the MCAS was arbitrary, with no evidence of its reliability as a measure of subject mastery. He also criticized MCAS questions to be ambiguous, of poor quality (i.e., reading comprehension questions that do not require reading the passage to get the right answer), or misaligned with state curriculum frameworks. Haney cited extensive research that shows that school achievement is highly correlated with the socioeconomic status of students and that MCAS might be more a measure of SES than educational attainment.

Finally, Haney noted that when students have access to technology and can type their essay answers, they perform much at a higher level than when they must write longhand.

High-stakes testing's impact on schools and students

Brian Jacob, Assistant Professor at the Kennedy School, provided findings from his research on high-stakes testing in Chicago. Chicago implemented a "get-tough" policy in 1995 that embraces two kinds of accountability – one for students and one for schools. Students would no longer be automatically promoted to the next grade, but must pass tests in math and reading to be promoted at the end of 3rd, 6th, and 8th grades. Students who fail the test must repeat the grade.

School accountability would be based solely on reading scores, and the City has the power to reconstitute a school if fewer than 15 percent of students met national norms for reading. Poor performance would bring intensive intervention including principal mentoring, reorganization of the school, teacher training, and help from external partners. If these efforts failed, the school would be shut down until a new staff could be hired under new leadership and with a clear plan for going forward.

Reading and math scores increased dramatically on the Illinois Test of Basic Skills (ITBS) after the addition of high-stakes for all three grade levels. But when comparing with a state assessment test that was given without high-stakes, these dramatic increases did not show.

A survey of teachers showed that teachers reported spending a significant amount of time on test preparation after the high-stakes were implemented, and they spent more time on tasks aligned with testing such as interpreting short reading passages. More students reported greater personal attention from teachers with the strongest increases for students at high or moderate risk. Students indicated greater parental support and higher levels of engagement, again with the greatest increases for high or moderate risk students.

Robert Costrell, Director of Research for the State Office of Administration and Finance, noted that the most pronounced improvement was in the lowest achieving groups. He also expanded on the curriculum-based design of the MCAS. On a curriculum-based exam, students will be held accountable for all aspects of the curriculum frameworks that the state designates – thus, if teachers teach to the test, they will effectively be teaching all of the necessary frameworks.

The view from the State House

In his luncheon address, James Peyser, chairman of the Massachusetts Board of Education, focusing on the connection between curriculum and testing.

“Ed reform is not just about assessment, it’s about curriculum frameworks and standards for what kids should be able to do by a certain point. There are increased resources to enrich curriculum and expand professional development. Finally, there is a system of school-level accountability. School accountability is not about having a club in the closet to beat up principals and teachers. It’s about identifying people who are not able to improve on their own and teachers who can be helped through professional development.”

Peyser discussed three broad goals of MCAS programs:

Expanding opportunity: The Commonwealth plans to provide Remediation Grants to school districts to run summer programs, Saturday programs, or afterschool programs to help students struggling with the MCAS. Governor Swift plans to target districts with a disproportionate number of failing students for the program. The state also plans to provide individualized tutoring support through grants to local organizations and an “Extra Help Guarantee” grant of \$1,000 families of struggling high school seniors who have not passed the MCAS. Peyser added:

Standards should be held constant and time should be a variable – meaning that we shouldn’t just cut them off after 12th grade. For some students, this means continuing their study after 12th grade is over. The Governor has committed to continuing learning opportunities for any student who has met local graduation requirements but who has not yet passed the MCAS.”

Ensure fairness: Standardized tests have long been criticized for random variation in scores, for poor question design, and for unfairly judging students who might just be having a bad day. The state hopes to insure fairness by allowing students to take the exam five times and, possibly, by establishing an appeals process that would consider a student’s grades, academic engagement, and effort in addition to their MCAS scores. If a student failed by only a few points but had consistently high grades, good attendance, took the test in good faith and was a positive participant in her classes, that student would be considered for an appeal. The state already has policies to take into account the needs of special needs students.

Press for excellence: The major goal of MCAS, Peyser said, was not to teach toward minimum standards, but to drive schools and students to seek higher levels of achievement throughout public education. Teacher recruitment and retention is a critical element in school improvement. The Massachusetts Institute for New Teachers (MINT) helps uncertified entrants into the teaching profession get training and certification during their first year of teaching. The state is also working with colleges and universities to encourage graduates to become teachers and to provide alternative certification programs. The state is also funding induction and mentoring programs designed at retention and professional development for high-quality teachers.

Peyser concluded his address with a positive assessment of school,s’ response to MCAS: “It’s actually working. When you go out to schools, you see that people, student and adults, are behaving differently. Not that this wasn’t happening before, but now it’s happening more consistently and more rapidly. Teachers collaborating more, students are getting additional learning time, and we are now seeing the first signs of real improvement. NAEP scores showed an 8-percentage point drop the in number of students doing poorly from 1996-2001. National average drop was around 3 percent. This is because we are much more engaged in education reform than other states.

Testing’s impact on communities

Economist Sandra Black of the University of California at Los Angeles examined at the impact of test scores on housing prices. It is well known that people choose their neighborhood based on the quality of the schools, but how families make their judgments remains a matter of debate. Some families may look at average achievement of students in that district, though a more accurate measure would be to look at the “value added” of the school.

Realtors shared student achievement data with prospective buyers since the 1970’s, but the greater availability and prominence of the data today likely make it more important in a family’s overall decision of where to buy a home. Black also looked at the impact of testing on neighborhood composition. Research has shown that wealthier families move into neighborhoods with better schools; because these

neighborhoods then get more expensive, poorer families cannot access these schools. This is likely to contribute to reinforcing class designations across generations.

‘Nice’ and ‘mean’ accountability

Frederick Hess of the University of Virginia laid out two visions of accountability: “nice” accountability and “mean” accountability. He defined “nice” accountability as something that relies on suasion and informal pressure to bring about change. “Mean” accountability, on the other hand, is coercive and relies on punitive measures to impel people to change. Mean accountability fosters agreements on what the goals and standards should be and can even increase the professional discretion of educators if the focus on a single outcome takes away the need for day-to-day monitoring. It may also ensure that the hardest-to-educate groups are being served.

Hess noted a common trend in standards movements: “watering down” requirements. People who are upset and motivated make their discontent known and demand accommodations. Policymakers respond to this by adding provisions rather than removing standards, but these provisions often serve to weaken the standards anyway. One method that policymakers can use, however, is to institutionalize the policy and hope that this will wear down the resistance over time. As accountability enters the grammar of schooling, people realize that a system with no accountability cannot succeed. Assessments become the “gold standard” for performance, and schools that do not do them are seen as lacking focus. Thus, over time and with a gradual phasing-in process, standards can be a valuable tool in changing the way schools do business.

Tim Knowles, Deputy Superintendent for Teaching and Learning at the Boston Public Schools, echoed that sentiment by saying that the standards and accountability movement in Massachusetts has “driven fundamental changes in the way we do business at BPS.” He cited extended days and a move toward small learning communities at the high schools, and intensive support and training for teachers. MCAS, he said, has “galvanized the community in ways that have not been seen before. It has raised attention, raised concern, and brought resources to bear that never existed before.”

But Knowles cited two areas of concern. First, he is concerned over the disparity that exists in a state where public students are accountable to an exam and private school students are not. If families can “afford to buy a diploma by attending a private school, they will.” Low-income students will be forced to take the exam and more affluent students will not. This inequality often brings up the question of using vouchers to send low-income students to private schools – a situation that puts the state in the position of using public monies to fund students who will not be accountable to the public’s standards. On this point, an audience member proposed giving a voucher to any student who failed the MCAS in order to help them catch up. Knowles replied: “The real problem is that kids who are failing need very clear pathways about where they are going and what’s going to happen to them if they fail. Could the community colleges help kids catch up? Perhaps. But the notion of incentives for private schools to own up to the MCAS challenge is an interesting one.”

Knowles also pointed out that in order to improve student outcomes, teaching practices needed to change. Unfortunately, the MCAS is scored at the level of the student, and there is very little data available that speaks directly to teaching practices or how teachers rank against other teachers statewide. This does not mean tying teacher performance to student achievement, rather it means having a vertically equated test so teachers can evaluate their progress over time and change teaching methods accordingly.

Ed Doherty, president of the Boston Teachers Union, echoed those sentiments and added that MCAS has put undue pressure on teachers, especially urban teachers. Doherty quoted frustrated teachers who were considering leaving the profession: “I’m tired of reading in the paper every day that I’m failing at my job. I’m tired of the disrespect. I’m tired of the bashing. I’m working as hard as I can and I love my kids, but I’m tired of being told that I’m failing.”

Doherty continued, “As long as this keeps getting hammered in, we will drive people out of the profession.” He added that teacher accountability is a difficult issue because much of the success of teachers depends on the students. “All students can learn,” Doherty said. “All students can learn if they come to school every day, get support at home, come to school fed, have a good attitude, etc. Unfortunately, this is not the case in many urban areas, and schools by themselves cannot do it all. Schools cannot make up for many public policy problems that face these families: housing, welfare, employment ...”

The quality of teaching in the Boston Public Schools, Doherty said, is excellent. He asserted that the quality of teaching at Dorchester High School as good as the quality of teaching at the more prestigious Boston Latin School. Faced with the myriad problems of the students in Dorchester, however, it becomes a much tougher job. All of this makes teacher standards very difficult, and nobody has devised an appropriate monitoring system – either qualitative or quantitative – to help evaluate teachers.

The week after

Five days after the conclusion of this conference, 10th grade MCAS scores were released for the class of 2003. They were dramatically higher than anyone, policymakers or critics, anticipated. The “Testing Testing” conference gives us many perspectives on why this might have happened, from students being better test-takers to successful remediation programs, from smaller high schools and longer school days to teachers who spent more time on test preparation or teachers who “taught to the test.” All of these theories require further research to be proven conclusively, but the fact remains that Massachusetts students seem well on their way to meeting some of the highest standards in the country.