Education REFORM

Ten Years after the Massachusetts Education Reform Act of 1993

2 **An Overview of Massachusetts Education Reform**

3 **Impact of the Education Reform Act on Equalizing Education Finance**
Linda Driscoll

6 **Bridges and Barriers: Equity Issues in High Stakes Testing**
Joseph B. Berger

11 **Setting Passing Scores on Tests**
Ronald K. Hambleton

15 **Are MCAS Scores Comparable Over Time?**
Lisa A. Keller

19 **VRROOM: The Virtual Reference Room**
Robert W. Maloy, David Hart, Paul Oh, and Victoria Getis

23 **School Accountability**
Susan Bowles

28 **Managing to Lead in the Decade of Education Reform**
Francis L. Gougeon

31 **Capacity to Implement Education Reform**
Kathryn A. McDermott

34 **Conclusions: The Impact of Education Reform after Ten Years**
Andrew Churchill

Education CONNECTION 2003
University of Masssachusetts Amherst School of Education

An Overview of
MASSACHUSETTS EDUCATION REFORM

In June 1993, Governor William Weld signed into law the Massachusetts Education Reform Act (MERA). MERA greatly increased the state role both in funding public education and in guiding the local educational process. The state's role changed to incorporate setting curriculum frameworks and holding schools accountable for student performance. Because MERA was designed to be a systemic reform of education, all of the various state activities and policies needed to fit together into a coherent whole based on state educational standards. This comprehensive set of reforms included the following elements.

Increased State Funding for Public Education. MERA doubled state funding of K–12 education from $1.3 billion in 1993 to $2.6 billion in 2000. The achievement of this ambitious target has been a major accomplishment.

A "Foundation Budget" for All Districts. MERA laid out the concept of a minimum budget necessary for each district to adequately educate all of its students. Poorer communities that were spending below foundation-budget levels would receive more than those at or above that threshold. By 2002, all districts were at or above foundation level. Revision of the foundation formula to more fully reflect actual spending needs is under consideration.

Learning Standards. The legislation instructed the Board of Education to develop curriculum frameworks and other standards such as vocational standards, and to support local districts' implementation of standards through alignment of curriculum and instruction. Frameworks have been developed in arts, English language arts, foreign languages, health, mathematics, history/ social science, and science/technology/ engineering.

Student Assessment. MERA required an assessment of student learning based on the frameworks and specified a competency determination as a requirement for graduation. The Massachusetts Comprehensive Assessment System (MCAS) has been developed and implemented for those purposes. MCAS tests have been administered in multiple grades in English language arts, mathematics, reading, science/technology/engineering, and history/social science. Students in the class of 2003 and beyond must perform at least at the "Needs Improvement" level in English language arts and mathematics to graduate.

An Accountability System for School and District Performance. MERA required the state to hold schools and districts accountable for student performance and to provide remedies for persistent underperformance. This has proven to be quite a challenge for the state, with several changes of jurisdiction occurring. Over the past year, a "performance index" process has been developed based on MCAS scores and has been judged to be in compliance with the federal No Child Left Behind Act.

Changes in Local Education Governance and Management. MERA included significant changes in the way schools and districts are run. School Committees' power over personnel issues was reduced, with superintendents and principals given more authority. All schools were to have School Councils composed of parents, teachers, students, and administrators.

Enhancing Educator Quality. A portion of state aid to local districts was earmarked for teacher professional development. Teacher licensure has been revised, and teacher tests for new teachers have been instituted.

Ensuring Readiness to Learn Through Early Childhood Education Programs. Between 1996 and 1999, spending on early childhood education increased by 247 percent.

Implementing Choice and Charter Schools. MERA expanded inter-district choice and authorized the state to approve charter schools—public schools exempt from local control and union contracts.

Impact of the Education Reform Act on Equalizing Education Finance

Linda Driscoll

As a first time school superintendent in June 1993, I remember the excitement I felt as I learned of a new funding formula conceived through the recently passed Massachusetts Education Reform Act (MERA) of 1993. Designed to expand the state's role in financing schools and establish an adequate and fair per pupil spending requirement, this formula seemed to be a great improvement over past funding mechanisms. My colleagues and I had hopes that this formula would improve the budget making process. Prior to MERA, Chapter 70 funds from the state to support education were not predictable, adequate, or equitable. Determining the municipality's share could be both confusing and political.

This new method, although quite complex, set a minimum spending level for each district, determined the municipality's ability to pay, and designated state funds to make up the difference. The per-pupil rates were based upon 19 components including such areas as teachers' salaries, support, administration, professional development, books, and maintenance. Student enrollment was then weighted by such factors as elementary school, high school, special education, bilingual, and low-income populations. Therefore, districts with similar enrollment numbers could have very different foundation amounts if their school populations contained large numbers of bilingual and low-income students. Special education funds, however, were not based upon actual student numbers but upon prescribed percentages. Three and a half percent was the assumption for in-district special education and one percent for students who are tuitioned out. Finally, a wage adjustment factor either increased or decreased the amount designated to salary lines based upon the community's living standard. Thus, a district's foundation amount was determined. The town's ability to pay its share was determined by an even more complicated procedure. It contained 35 points among which were included the municipality's equalized valuation, growth rate, per capita income, and excess debt.

This formula, although complicated, did set a minimum spending level for each district based upon enrollment needs and determined what percentage of this need would be expected from the community, based upon its ability to pay. The difference would come from the state. The basic points made sense, and the factors of foundation budget, minimum contribution, and Chapter 70 aid are clear.

This formula, begun in FY94, was authorized for a seven-year period, scheduled to end in FY00. It has not been reauthorized and debate over the formula is ongoing from such organizations as the Massachusetts Superintendents' Association, Massachusetts Municipal Association, League of Women Voters, Massachusetts Department of Education, Massachusetts Association of School Committees, the Governor's Office, and the Foundation Budget Review Commission. Since FY93, state aid to schools, commonly referred to as Chapter 70 aid, has grown from $1.3 billion to about $3.2

billion in FY03. The foundation per-pupil expenditures rose from $5,577 to $7,201 in that same time span. These figures represent a change in state aid to education from 28.5 percent in FY93 to 46.9 percent in FY03. Massachusetts is now closer to the national average of 49 percent. In fact, the Department of Education's budget for all programs increased an average of $285 million each year during the first eight years of education reform.[1]

So, has this large increase in state aid accomplished the aim of ensuring that the schools are adequately funded in an equitable way? The answer to that question depends upon how the goal is defined. If the goal is to assist all school systems to meet the foundation level of spending, then that target has been reached. However, if the goal is to ensure that all schools have the level of resources necessary to provide the kind of excellence in curriculum and instruction that leads to all students being successful, then we still have much to accomplish. The notion of providing merely *adequate* resources while expecting *excellence* from all students is central to much of the debate concerning school finance in Massachusetts.

Has the goal of reducing disparity of per-pupil expenditures across property-rich and property-poor communities been achieved? Although there are still some large discrepancies, with some richer communities spending well above their minimum contribution, it is evident that this gap has lessened. In FY01, two-thirds of districts spent 10 percent higher than their foundation, 31 districts were above 150 percent, and two districts were approaching 300 percent.[2] Has the money been fairly allocated to municipalities and regions? It is obvious that the poorer urban districts have benefited greatly financially from school reform. However, the regional districts, because of their unique situation, have struggled with disparities in allocations from sending towns. Also, some high effort communities work harder spending a larger percentage of their taxes for support of the schools. There are districts whose state portion of their expenses have increased dramatically, with the Chapter 70 funds reaching as high as 99 percent of the foundation. Other districts' state aid has remained very close to their FY93 percentages, with some receiving as little as 12 percent from the state.

There is little doubt that the schools in Massachusetts, on the whole, have benefited from the financial increases due to education reform, but can we do better and can this formula be more effective and fairer? Ambitious goals have been set by the state of Massachusetts through the curriculum frameworks, high standards, and high stakes accountability as measured by the Massachusetts Comprehensive Assessment System (MCAS). In order to continue to strive towards those goals and to enable all children to acquire the proficiencies these standards demand, the foundation budget needs to be raised. It simply is not adequate to reach these goals.

Whether the present foundation is scrapped and begun again from scratch or modified in its present condition, several pertinent issues must be addressed. At the top of this list is *special education*. Simply put, it is under funded. It also has taken a disproportionate share of new funds allocated through education reform. For 56 percent of districts, special education (SPED) increases exceeded 50 percent of new state aid, and for 29 percent of districts, SPED increases exceeded 100 percent of new state aid from FY93 to FY99.[3] The percentage of money allocated to SPED in the foundation budget is inadequate. It needs to be adjusted based upon reality, not projections. Also, the Circuit Breaker law, where the state reimburses instructional and tuition costs for high cost students both in and out of district, should be fully funded. *Student teacher ratios* need to be adjusted. The numbers deemed effective nearly ten years ago, when the education reform law was written, are no longer valid. *Per pupil spending* categories need to be changed. Elementary and preschool allocations, in particular, are unrealistic. The population of students in preschool contains many at-risk students who need more resources to provide for their programming needs. *Teachers' salaries* need to be factored into the foundation at actual amounts. *Enrollment numbers* should be better computed, especially in those

districts with high growth. This lag in determining foundation numbers results in budgets being based on lower than actual student counts. A *technology factor* needs to be added to the foundation budget. The financial demands of funding technology have increased greatly since MERA's inception. *Low-income students* should be counted using more accurate data than free/reduced lunch counts that are gathered through an application process. Money must be included to address the cost of *implementing high standards and accountability*. The cost of remedial and after-school tutorials to ensure that all students meet these standards is not implicit in the foundation. The foundation budget was not developed with an understanding of the costs of redesigning standards-based curriculum and re-provisioning schools for these changes.

It is not within the scope of this article to discuss the many issues regarding equity on the municipal side. Because of vagaries in the formula, economically similar communities can have widely varying minimum contributions. Let it suffice to mention that many towns are struggling to provide for their students within the mandates of education reform and that adjustments are warranted.

After almost ten years of operating schools with the foundation budget, much has been learned about the pros and cons of such a funding mechanism. The goal of assisting all schools to meet foundation has been met. This reform movement—spurred by the *McDuffey* lawsuit,[4] formulated through "Every Child a Winner,"[5] supported by motivated lawmakers and a growing economy—has improved the financial conditions in many schools. There is a greater understanding of what needs to be done to improve the structure and fairness of the formula. But can it happen, and more importantly, will it happen? When revenues to the state decline, as is now the case, Chapter 70 funds are subsequently decreased. Even ten percent declines, as were the recommendations this past year, can be devastating to communities that receive up to 90 percent of their funds in Chapter 70 aid. Even if the state keeps its commitment to fund Chapter 70 to the degree that schools stay at foundation, it is not enough to accomplish expectations, particularly now with the federal requirements of the No Child Left Behind Act. If the increasing partnership between towns and the state slips backward, it is difficult to imagine that the municipalities and regions will be able to assume the increased costs. It goes without saying, however, that if the federal government would increase educational aid over the average seven percent it now allocates, the financial picture definitely would be brighter.

If the formula is reauthorized and funded adequately to meet the above mentioned adjustments, Massachusetts stands to be well on its way to fulfilling the promise of "Every Child a Winner."

ENDNOTES

[1] Memorandum to School Superintendents from David Driscoll, Commissioner of Education, October 22, 2001.

[2] Report of The Foundation Budget Review Commission, June 2001.

[3] Impact of Special Education on Education Reform, Mass. Special Needs Task Force, November 2001.

[4] *McDuffey v. Secretary of the Executive Office of Education,* 415 Mass. 545 (1993).

[5] Entitled *A Proposal for Legislative Action Plan for Systematic Reform of Massachusetts' Public Primary and Secondary Educational System* by the Massachusetts Business Alliance for Education (MBAE), 1991.

Linda Driscoll (M.Ed. 1982; Ed.D. 1996) has taught Public School Finance at the University of Massachusetts Amherst. Recently appointed Superintendent for Erving School Union 28, she previously served as Principal at the West Street School in Granby, Massachusetts, and has prior experience as a Superintendent of Schools and Special Education Director.

Bridges and Barriers: Equity Issues in High Stakes Testing

Joseph B. Berger

Over the past several decades, the call for educational reform and accountability has been one of the most visible public policy issues in the nation.The need for excellence in education is undisputed, yet the nature of the current wave of educational reform movements has been the source of public controversy and ongoing debate. Testing and accountability have been at the core of recent attempts to improve and reform education at K–12 and postsecondary levels and serve as lightning rods for much of the controversy. While testing and accountability are not new to education, the prominent role they play in the current reform era is unprecedented. Moreover, not only are states and educational institutions being increasingly mandated to use standardized tests for a variety of assessment and accountability purposes, but there is also a growing trend toward implementing policies that attach "high stakes" to student performance on these tests. High stakes testing refers to the practice of attaching consequences of the highest order to an individual's performance on a particular test. Most notable are the high stakes attached to standardized tests at the high school level; in some states, students must now pass these tests in order to receive a high school diploma. Many plans attach stakes to the performance of schools and districts, however, thus far, states have focused more on holding individual students accountable than on finding ways to hold schools and districts responsible for ensuring that students have been adequately prepared to meet the challenges of high stakes tests.

Despite the controversy, the use of high stakes testing as a reform and accountability tool is becoming increasingly common. As of the end of 2002, 18 states have implemented high stakes testing as a requirement for high school graduation and six other states are in the process of phasing in such a requirement.[1] Other states are considering the implementation of high stakes testing. It is also clear that the recently enacted No Child Left Behind Elementary and Secondary Education Act (ESEA) federal education policy that heavily emphasizes required testing in public schools will add momentum to the movement towards the increased role of testing and may lead to a greater likelihood that stakes will be attached to those tests as well.

The controversies surrounding high stakes testing have been polarized for and against this policy. While many people support the important role that testing can and should play as one part of more comprehensive efforts to improve and reform education, others are concerned about the unintended and potentially negative consequences that may exacerbate existing educational inequities. More specifically, testing has been promoted as an essential component of the educational reform and accountability movement as an impetus for motivating schools and students to perform at higher standards while also providing a means for assessing progress. From this pro-testing perspective, the use of tests will help build better bridges to future success for students. Tests are seen as an important feature of a larger reform effort that will ultimately result in students who are better prepared to succeed in college and career. It has been suggested that accountability systems that use testing as a means for ensuring that students are learning are essential to making sure such bridges to future opportunity are available to all students, particularly those who have been traditionally disadvantaged by existing inequities in the educational system. Such comprehensive reform and testing efforts at the state and national levels are intended to provide impetus for

chronically under-achieving districts, schools, and students to be identified as such so that targeted interventions can be implemented to close the achievement gaps that have thus far persisted in the absence of such reforms. Proponents believe that the increased emphasis on testing has benefits for students, schools, and society as a whole. Additionally, proponents assert that the implementation of high stakes tests will help schools better assess their own performance, help districts and states improve their ability to evaluate the educational performance of schools, and provide a means for holding schools accountable for the learning that occurs in their classrooms.[2]

Others have expressed concern that an over-emphasis on testing leads to the creation of more barriers that will deny future educational and vocational opportunities for many students, particularly those who are already disadvantaged by an inequitable educational system. In particular, high-stakes graduation tests may have a negative effect on access to postsecondary education for students who have traditionally been underserved in K–12 educational settings and under-represented in higher education. Blacks, Hispanics, English-language learners (ELL), and low-SES students are more likely to attend schools that fail to provide high-quality curriculum and instruction than White and high-SES students. It is therefore not surprising that low-SES and minority students fail high-stakes graduation tests at much higher rates than do their high-SES and non-minority peers.[3] Additionally, studies suggest that students who are doing well academically are the most likely to drop out of school as a result of failing a high-stakes graduation test.[4] Thus, failure on the high-stakes graduation test may dissuade many otherwise high-performing minority students from attaining a postsecondary education. This already unacceptable inequity is a growing problem, as these groups also represent a major portion of the fastest growing segment of the population—a segment of population that is projected to constitute a majority of our citizenry and workforce sometime later this century. Many constituents and stakeholders are also concerned that heavy reliance on such tests will lead to increased "teaching to the test," in which curriculum and instruction focus solely on subject matter covered on the standardized tests; potential misalignments with other tests used throughout the educational pipeline that may result in students getting mixed messages about their academic progress; inappropriate use of tests to determine college admissions and placement; and barriers that limit students' aspirations and access.

There is also a need to align education reform efforts across the K–16 educational spectrum so that changes in primary and secondary education lead to greater levels of opportunity and achievement for students in postsecondary education. Standards-based reform cuts both ways for access to college. On the positive side, it promises to redress the inequities of the previous educational system, ensuring that if all students in a state are proficient in the same minimum standard then they will have equal opportunity to master the same high-level curriculum. This can potentially increase the pool of college-ready students. Conversely, when implemented with high stakes, reform will negatively impact many students failing to meet minimum thresholds on standardized exams. Because of their lack of success on such standardized exams, often such negative impacts will befall minority and low-income students, preventing them from graduating from high school and moving on to postsecondary education. Whatever the long-term consequences of standards-based reform, in the absence of a major policy change, the short-term consequence of high stakes tests will almost certainly be a smaller pool of students eligible for postsecondary education. Current trends suggest that those students from minority and low income backgrounds who are already most under-represented in postsecondary education are most likely to have their access to college reduced even further.[5]

Massachusetts is one of the states that has generated a great deal of controversy regarding its use of high stakes testing. In 1993, Massachusetts instituted the Massachusetts Education Reform Act (MERA), the

most comprehensive statewide standards-and-accountability reform in New England. In implementing MERA, the state created the common core of learning, state-level curriculum frameworks in major subjects. The curriculum alignment was widely supported, and schools and districts have been working to align their curricula with the content of the common core.

In 1998, the Massachusetts Comprehensive Assessment System (MCAS) was initiated and is currently used to test students in grades four, eight, and ten in four core subject areas (English language arts, mathematics, science, and social studies). A requirement that students pass the tenth grade MCAS English and mathematics exams in order to receive a high school diploma has been phased in, with the Class of 2003 being the first class held to this standard in order to graduate. This requirement, like other high stakes testing, has generated much controversy and is just as polarized in Massachusetts as is the broader issue throughout the rest of the nation. The Massachusetts Department of Education (DOE) considers the graduation requirement important in the effort to raise the academic achievement of its students; however, many parents and educators think the requirement is an unnecessary obstacle to postsecondary admission and employment opportunities. With the phase-in of the high-stakes impact of the test, proponents of MCAS argue that students will do better on MCAS in the future because they will have been exposed to curriculum based on the state frameworks and will take the test more seriously. Many stakeholders are particularly concerned about the impact of MCAS as a high stakes test on students who have been traditionally disadvantaged in public schools.

A recent report[6] by the Center for Education Policy at the University of Massachusetts Amherst on the progress of education reform in Massachusetts summarizes some of the results of MCAS. In spring 2001, 68 percent of the Class of 2003 achieved competency determination status by passing both portions of the tenth-grade MCAS. Eighty-two percent passed the English language arts assessment and 75 percent passed the mathematics assessment. After the fall 2001 retest, approximately 75 percent had achieved competency determination status, and after the spring 2002 test, approximately 81 percent had passed both sections. Clearly, these numbers do indicate that the majority of students can pass over these bridges and on to high school graduation and all of the doors that can be opened with a high school diploma.

However, MCAS performance varies across different sub-populations. On one level, this can be seen as MCAS fulfilling its mandate to build future bridges by focusing on achievement gaps and allowing necessary resources to be targeted toward the neediest students and schools. Unfortunately, the high-stakes nature of the assessment turns these disparities into barriers that cause certain types of students to be more likely to suffer negative consequences. For example, students from higher-income districts were more likely to pass MCAS than their peers in lower-income districts. Approximately 95 percent of the students in the highest-income districts had passed the test in 2001, while only about 59 percent of the Class of 2003 had passed in the lowest-income districts. Overall, 81 percent of tenth graders had passed both required MCAS tests after the third attempt (spring 2002). Passing rates for "regular education" students were significantly higher at 87 percent than for students with disabilities who had a 55 percent passing rate and Limited English Proficient (LEP) students were much lower, at 35 percent.

Achievement gaps also exist across racial and ethnic groups. White students in the Class of 2003 had an 87 percent passing rate after the spring 2002 tests. Asian and Native American students had a passing rate of 83 percent after three attempts. African-American and Hispanic students had passing rates of 56 percent and 50 percent, respectively. These findings indicate gaps that are far too wide between these groups, suggesting that students from these various disadvantaged groups are disproportionately likely to encounter barriers as they attempt to pass MCAS as a high stakes test.

One factor that seems to be an important predictor of success on MCAS is, not surprisingly, the level of courses students have taken prior to taking the test. Opponents and supporters of MCAS agree that the test is not a "basic skills" test; it includes material normally covered in algebra and geometry courses. A study released in April 2000 by the independent group MassInsight Education found strong correlations between enrollment in higher-level mathematics courses and success on MCAS. More than half of the

students who failed the MCAS mathematics test in 1999 were enrolled in Algebra I or an even lower-level course in tenth grade—the norm for college-bound students is to take Algebra I in eighth or ninth grade.[7] According to the DOE, Latino tenth-graders are three times likelier to be enrolled in low-level mathematics courses than Whites. Black tenth-graders are two and half times likelier to be taking low-level mathematics. Even more troubling, 12 percent of Latino tenth-graders are not enrolled in a mathematics course at all.[8] While overall test scores improved dramatically in 2001 there are still large-scale achievement gaps that remain, a point acknowledged by the DOE.[9]

The scores for under-represented minority students on MCAS, as on other standardized tests around the country, remain lower than the scores of their non-minority peers. The results of the analysis of MCAS in one study, however, suggests that even after controlling for SAT scores, minority students fail about 10 percent more than non-minority students.[10] This means that there is an additional 10 percent of bias (above and beyond any existing bias in SAT) in a test that students need to pass in order to graduate from high school. It is also troubling that minority students are twice as likely to fail MCAS as are non-minority students. This type of systemic result cries out for a determination of whether this is test discrimination or other discrimination and an answer regarding how we can rid such exam results of the disparity. At a minimum, a lack of answers leaves the state vulnerable to legal challenges to its testing system. Holding students accountable for a test that appears biased is not right; states should only consider attaching high stakes to these tests once the inherent bias of the tests has been thoroughly examined and eliminated.

Currently, MCAS is the only high stakes test in New England that requires students to pass both portions of the exam in order to be eligible to receive their high school diplomas. In the future, there will be even more components than just English language arts and mathematics that students must pass in order to graduate from high school. This particular approach is only one way of determining the consequences of high stakes for students. Because students learn in different ways and succeed using various combinations of skills, policymakers should strongly consider structuring tests to reflect this reality. Some students compensate for weaknesses in one academic area with strength in another. Colleges and universities have long recognized this fact, as evidenced by their willingness to allow for higher scores in one portion of the SAT to compensate for lower scores in another. The NCAA, an arm of the college's athletic departments, established a tradeoff between standardized test scores and cumulative high school averages. The same should be done with high stakes standardized tests such as MCAS. This could be accomplished in several ways. For example, a sliding scale could be established where a score above a certain level in mathematics could balance a score below a certain level in English. Alternatively, the cut score could be determined as a cumulative score across the various portions of the exam, rather than by the unique results of each portion. While there are trade-offs associated with any approach to scoring tests and determining passing standards, the importance and consequences of these tests mandate that all potentially useful approaches be given serious consideration. Only in this manner can one of our most fundamental values—that students are given opportunities to grow and mature at their own pace—be protected.

Embedded within this recommendation is a call for the use of multiple assessments. The use of multiple indicators has been criticized on a number of counts, including that they are fuzzy and that there is no agreement on which sets of measures should be used.[11] Yet, there is no consensus about the effectiveness of any one test. Indeed, the findings from several studies reveal that any one test only measures a partial facet of student learning; the use of additional information provides a more complete and accurate picture of what students have learned and may yet achieve. Single tests are efficient, but are less effective than multiple measures in assessing student learning and predicting future success.

Higher standards and higher stakes mean greater challenges for students. The appropriate balance between level of challenge and subsequent level of support is recognized as a fundamentally necessary condition to facilitate growth, development, and learning. Given the importance of the challenge students face when taking standardized exams, particularly those with high stakes, it is of paramount importance that appropriate levels of support be provided to

help students prepare for these tests. This means that states need to ensure that curriculum frameworks are fully implemented for every child, that at-risk students are receiving additional resources, and that effective professional development is being provided for educators responsible for student learning. All of these things should be fully implemented prior to attaching any stakes to the performance of individual students on such tests. This recommendation also focuses on the need to hold adults accountable first. Most states that have implemented high stakes testing, not just Massachusetts, have found it more difficult to find ways of holding teachers, schools, and districts accountable for preparing students to succeed than they have found it to hold students accountable. Yet, testing may result in unfair and unequal treatment of students if learning environments are not all equal in supporting the students who must pass these tests or have further educational opportunities denied to them. High stakes should be applied only at the point that schools and districts are providing the education to which every child is entitled.

There is little doubt that testing is here to stay. However, policymakers have a wide range of choices about how to test students and how to use the tests as a fair and effective means for improving educational outcomes. Policymakers need to remember that testing is the means to an end, not the end in and of itself. This type of recognition can serve as the basis for making sure that tests are constructed and used in ways that best benefit students. In order to do this, testing should not only be evaluative; it should also be rigorously and continually evaluated. The empirical results of such evaluations should be used to guide on-going adjustments to the tests and the assessment systems of which they are often a part. Fine tuning will be necessary until tests become better bridges across which all can realize their educational aspirations and not bigger barriers that impede or deny opportunity.

If we truly are going to "Leave No Child Behind" then we need to make sure that all children are properly prepared to cross the bridges we make for them. If the students are unprepared, or if the bridges are not well constructed or too difficult to get to or across, then we will have constructed barriers instead of bridges. The results of high stakes testing in Massachusetts and elsewhere suggest that these exit exams are serving as both bridges and barriers. Many students are crossing that bridge, but many others are encountering it as a barrier. Unfortunately, many of those who encounter a barrier instead of bridge are minority and low-income students who have traditionally encountered more than their fair share of barriers. Policymakers and educators need to make sure that educational policies, such as high stakes tests, are indeed bridges to future success and not barriers to opportunity. Now is the time to make sure that all students have equitable chances to cross the bridge and create successful futures on the other side that will benefit themselves, their families, their communities, and their employers, as well as our larger society.

ENDNOTES

[1] *Measuring Up: A Report on Standards and Assessments for Massachusetts.* Washington, DC: Achieve, Inc., 2001.

[2] *op. cit.*

[3] National Research Council. "Linking Test and Learning" *News Report, 49*(3). 1999. p.3.

[4] *op.cit.*

[5] Coelen, S. P. & Berger, J. B. *Educational Opportunity and the Impact of Graduation Testing in New England.* Amherst, MA: Center for Education Policy, School of Education, University of Massachusetts Amherst, 2001.

[6] Churchill, A., McDermott, K., Bowles, S. Lee-Davis, C., Effrat, A., Berger, J., Carey, J., Brooks, C. & Klock, R. *2002 Annual Report on the Progress of Education Reform in Massachusetts.* Boston: Massachusetts Education Reform Review Commission, 2002.

[7] *High Stakes and High Standards: Public Attitudes toward Education Reform.* Boston: Mass Insight, 2000.

[8] Gaston Institute. "MCAS Failure Rates for 1998 and 1999 Compared in New Study." *Maurice Gaston Institute Report.* Boston: University of Massachusetts Boston, 2000.

[9] *Spring 2001 MCAS Tests: State Results by Race/Ethnicity and Student Status.* Malden, MA: Massachusetts Department of Education, 2001.

[10] Coelen, S. P. & Berger, J. B. *Educational Opportunity and the Impact of Graduation Testing in New England.* Amherst, MA: Center for Education Policy, School of Education, University of Massachusetts Amherst, 2001.

[11] Reville, S. P. "Multiple Measures?" *Education Week. 21*(11). 2001. p. 52.

Joseph B. Berger is an assistant professor in the Department of Educational Policy, Research and Administration. He currently serves as the interim chair of the department and is an Associate Director of the Center for Education Policy.

Setting Passing Scores on Tests…

Not Too High…Not Too Low…But Just About Right

Ronald K. Hambleton

Hardly a day goes by without an article, editorial, or letter to the editor in the newspaper about state testing. In my state, the testing program is called the Massachusetts Comprehensive Assessment System, better known as MCAS. MCAS is about educational reform—reforms in curricula, teacher and principal training, class size, after-school programs, school funding, testing, etc., but many think MCAS is only about testing. Just about every state has its own version of the MCAS. With the new federal legislation, the No Child Left Behind Act, states have little choice but to embark upon major educational reforms, especially in the subject areas of reading and mathematics.

Today's newspaper headline in Amherst was about 10,000 high school seniors across the state of Massachusetts who have their last chance this week to pass the English language arts and mathematics tests prior to graduation next June. The day before, the newspaper reported on a mathematics item on last year's test that appeared to have two correct answers. And the day before that, a news report focused on a group of students who managed to meet the state's graduation criteria for English language arts and mathematics proficiency by an alternative method to passing the state tests—high attendance, coming close to passing the tests, and doing well in their day-to-day classroom work. News coverage is relentless. Almost everyone wants to improve education, and just about everyone has opinions (and voices those opinions) about how this can and can't be done.

Many people have questioned the "passing scores" on state tests. In Massachusetts, one of the popular explanations for the high failure rate among tenth grade students has been that the passing scores are set too high. Given all the talk about passing scores, it is surprising that very few people seem to know much about them. Who sets the passing scores? How are they set? What level of test performance is actually required to achieve a passing score of 220 on the MCAS? In my work around the state I have yet to meet anyone—a school board member, a teacher, or a parent—who has more than the most superficial knowledge about *how* the passing scores are actually set, or about the meaning of a passing score of 220. The remainder of this paper will be about setting passing scores. The specifics pertain to Massachusetts, but most states set their passing scores in much the same way as Massachusetts.

I would make the point first that most states today do not simply report scores as pass-fail. With the more demanding curricula in place, many states want to recognize several levels of academic accomplishment. Students are classified, based upon their test performance, into four levels of proficiency: Advanced, Proficient, Basic, and Below Basic. In Massachusetts,

four proficiency levels are used: Advanced, Proficient, Needs Improvement, and Warning. Therefore, not one but three cut-scores (often called "performance standards") are set on the test score scale to classify students into the four performance categories. Performance standards are scores on the test score scale that are needed to separate students into performance levels. On the 1998 MCAS grade 10 mathematics test, the performance standards were set at approximately 40 percent correct (for Needs Improvement), 57 percent correct (for Proficient), and 76 percent correct (for Advanced). The performance standards are about 5 percent higher on the grade 10 English language arts test. Students must achieve at least the Needs Improvement level to "pass." These are the levels of performance that have remained in place since 1998. In subsequent years, test scores are adjusted to reflect the fact that later tests may be a bit easier or harder (this is called "equating" and is discussed in an article on page 15), and then the 1998 performance standards are applied.

Let's take a look now at how states go about trying to set defensible and valid performance standards. First, performance standards on a state test are almost always recommended by a panel of 15 to 25 persons to the State Board of Education, which has the responsibility for setting the final performance standards. In Massachusetts, the recommendations from the panels were accepted without revision by the Board of Education. Members of these panels include teachers and school administrators and, sometimes, especially for tests at the higher grade levels, the public. The composition of the panels varies from state to state, but teachers normally make up the largest group. When performance standards were set in Massachusetts in 1998, 209 persons participated on 12 panels (an average of about 17 per panel). Fifty-one percent of the panel members were teachers. The prevailing view is that teachers have the best information about (1) the challenges of the curricula and (2) realistic expectations for students. Other informed individuals and stakeholders are usually included on these panels: in Massachusetts, 22 percent of panel members were school administrators, 17 percent were from the business community, 7 percent were from higher education, and 3 percent were school committee members or members of local/state government. Every effort is made to form panels that bring balance, experience, and diversity to the process. Factors such as gender, ethnic background, age, years of teaching experience, subject area expertise, and geographic location in the state are routinely considered in forming panels. The acceptability of a panel to stakeholders is just about the most important consideration in the process of setting defensible and valid performance standards. The 209 panel members in Massachusetts were divided into 12 smaller panels and assigned to set performance standards on one of the tests at grades 4, 8, and 10.

Second, panelists must have clear descriptions of the performance levels. In Massachusetts and other states, these performance levels are very carefully spelled out (and approved by the Board of Education), and provide a basis for setting the performance standards on the tests and for interpreting student performance after tests have been administered. For example, at the grade 10 level in

mathematics the Needs Improvement category is described as follows: A student who achieves this performance level
1. Demonstrates partial understanding of our numeration system.
2. Performs some calculations and estimations.
3. Identifies examples of basic math concepts.
4. Reads and constructs graphs, tables, and charts.
5. Applies learned procedures to solve routine problems.
6. Applies some reasoning methods to solve simple problems.
7. Identifies and uses basic mathematical terms.
Descriptions like the one above are prepared for each performance level, and used by the standard-setting panels to determine the test scores needed to be classified at the Needs Improvement, Proficient, and Advanced levels. Similar descriptions are available for other performance levels in grade 10 mathematics and for the performance levels on the other tests.

With the panels formed, and descriptions of performance levels available, step three can begin: the judgmental process of setting performance standards. Following what is normally a half-day of training, panels begin their work. In Massachusetts, panels used a method for setting performance standards that involves sorting student test booklets into four performance levels. Panelists review the actual test booklets of a wide sampling of students who took the test and sort the booklets into the four performance levels. Panel members work independently at first, and later they discuss their classifications with other panel members and revise their classifications when revisions seem appropriate. There are no right or wrong classifications. What is valued are the judgments and thought processes of the panel members. Normally this task is completed without knowing the scores attached to the student test booklets. In this way, panel members can focus their attention on the quality of work they observe and match it to the performance levels. Poor work can be assigned to the lowest category, Warning, and outstanding work can be assigned to the highest category, Advanced. Other levels of student work must also be judged and classified.

A standard-setting study will often run for two or three days. By the end of the first day, panelists in Massachusetts had the opportunity to review and classify student test booklets, discuss their classifications with other panel members, and reclassify any student test booklets when they felt changes were in order. The standard-setting task took two days and, on the second day, panelists were given more student test booklets to review. This time, the student test booklets were relatively close to the tentative performance standards set the previous day. This second set of student test booklets allows for a more precise determination of the placement of the performance standard between performance levels. Each performance standard is obtained in the following way: a performance standard is taken to be the average score of student booklets that are judged to be "on the borderline." These are the student booklets that panelists are equally divided about—they could be at one performance level or one lower. In setting the performance standard for Needs Improvement, borderline students would be those students where panelists were divided about their placement—Warning or Needs Improvement. The scores of these students mark the placement of the performance standard. So, if the test booklet scores for these students happened to be 18, 18, 19, 19, 20, and 20, the performance standard would be set at 19, the average of the six borderline student scores. The process is repeated for each performance standard.

In a final step, panelists are asked to provide an evaluation of their training, the process they went through to set performance standards, and their confidence in the resulting performance standards. Other evaluative criteria include the intra-consistency of panelists' ratings and the level of agreement among panelists in their performance standards.

Figure 1 (on page 14) shows the student test scores (0 to 100 percent) on the vertical axis with the three performance standards marked. Once a student's score is known, the performance category to which the student is assigned is obvious. The process described above is repeated for each test. Performance standards are set by a panel and approved by the Board of Education (e.g., 40 percent, 60 percent, 75 percent), and then they are used for

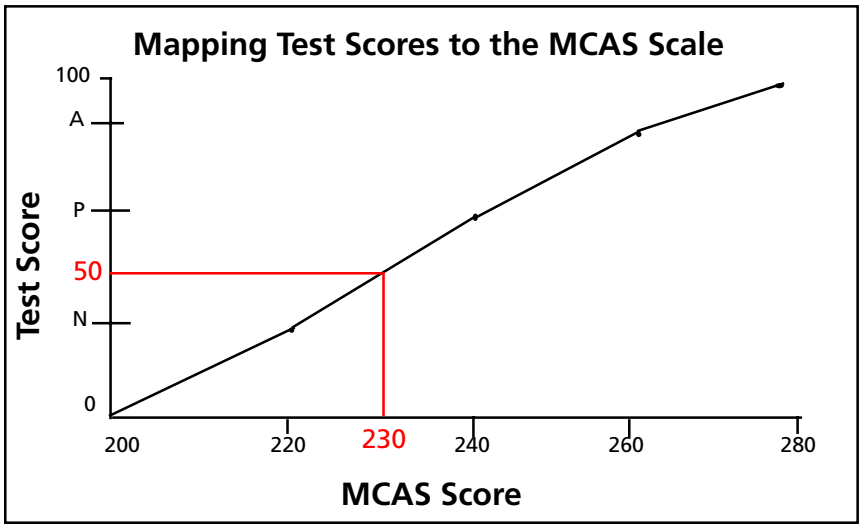


Figure 1

classifying student performance. Ultimately, each student's score is transformed to the "MCAS scale" via the mapping shown in the figure. Thus the student who does just well enough to meet the performance standard for Needs Improvement receives a score of 220. A student who does just well enough to meet the performance standard for Proficient receives a score of 240. Scores between these two performance standards would receive scores between 220 and 240, and so on. In the figure, a score of 50 would receive an MCAS score of 230. MCAS scores in excess of 260 are Advanced students.

In Massachusetts, regardless of grade level or subject, students are assigned to a performance level using the performance standards set by the panel (and approved by the Board of Education), and then their scores (after necessary adjustments are made for the non-equivalence of the tests from one year to the next) are mapped onto the MCAS scale. Thus the level of test performance represented by 220 depends on the panel, which in turn is working from the performance level descriptions and the test itself. In grade 10 mathematics in 2002, a student needed a score of 35 percent on the test to obtain a passing score of 220—students in the Needs Improvement, Proficient, or Advanced levels are considered to have passed the test. In grade 10 English language arts in 2002, a student needed a score of about 55 percent. The best way to think about a score of 220 is to recognize that a student with this score has accomplished just enough to meet the minimum expectations of the Needs Improvement performance level.

There are no "true performance standards" waiting to be discovered from careful educational research. Setting performance standards on educational tests like the MCAS involves careful professional judgment being made by persons who are viewed as suitable for providing the judgments. Validity of the performance standards is not judged by the percentage of students who pass or fail (who would know what these percentages should be?) but instead is judged by (1) the qualifications and demographics of the persons who are charged with the responsibility of setting the performance standards, (2) the clarity and suitability of the performance level descriptors, (3) the reasonableness of the process the panel goes through, and associated statistics that describe their judgments, and (4) the confidence the panel members have in the process itself and in the final performance standards that they set. For follow-up reading about setting performance standards, Cizek's book, *Setting Performance Standards: Concepts, Methods, and Perspectives,* would be an excellent starting point. For details on the standard-setting process in Massachusetts, readers are referred to the technical report published in 1999 by the Massachusetts Department of Education.

REFERENCES

Cizek, G. (Ed.). *Setting Performance Standards: Concepts, Methods, and Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers, 2001.

1998 Massachusetts Comprehensive Assessment System Technical Report. Malden, MA: Massachusetts Department of Education, 1999.

Ronald K. Hambleton is a Distinguished University Professor in the Department of Educational Policy, Research and Administration. He serves as Director of the Research and Evaluation Methods program of study and Co-Director of the Center for Educational Assessment.

Are MCAS Scores Comparable Over Time?

Lisa A. Keller

One of the central goals of the Massachusetts Education Reform Act (MERA) of 1993 is to ensure that all students achieve high academic standards, and the Massachusetts Comprehensive Assessment System (MCAS) tests were developed to monitor just how well the schools are meeting this objective. Therefore, the ability to compare students' performance on MCAS over time is of central importance in evaluating the impact of MERA. While the *concept* of comparing scores over time is simple, the *process* that enables a valid comparison is quite complex.

The cut scores used to determine the proficiency level of the student (Advanced, Proficient, Needs Improvement, Failing) were set in 1998. Given that different tests are administered each year, how can we be sure that the changes in performance are due to changes in student achievement and not changes in test difficulty? That is, how can we be sure that students are accurately categorized in 2002, or 2003, or any subsequent year? Are students actually improving in abilities or is the test getting easier? The issue central to these questions is how to disentangle the impact on performance of the different tests and differing student abilities.

The MCAS was developed as an accountability tool connected to the education reform with the goal of monitoring student-, school-, district- and state-level progress. The MCAS tests are administered in grades 3–8 and in grade 10, and measure the content areas of mathematics, science and technology, English language arts, social studies, history, and reading. Due to the comprehensive nature of MCAS, not all tests are administered at all grade levels. The tests are divided among the years to ensure that students and teachers are not burdened with excessive testing. At the tenth grade level, students must pass both the English language arts test and the mathematics test to qualify for a high school diploma. (Students must also fulfill the graduation requirements of their particular school in order to receive a diploma and thus the graduation decision rests on many factors, not this one test.) In this way, the MCAS can be used as a monitoring tool. In the earlier grades, tests are administered to monitor educational progress of students so that remediation can take place prior to the tenth grade tests. The hope is that by implementing this assessment system, education can be improved for all students, and that a high school diploma will be a true indicator of a level of competency determined to be meaningful by educators in the Commonwealth.

Of central importance is the comparability of scores from year to year. Again, since the scores used to determine which category a student belongs in were established in 1998, it is essential that test scores from subsequent years are comparable to the scores that the same students would have obtained if they took the test in 1998. Test developers undergo extensive training in writing test items and choosing those items most appropriate

for the test. Additionally, all items are field-tested to ensure that they are functioning properly, and that the questions are appropriate and fair for all students. After the field-testing, items are selected based on statistical and content qualities. These features of the test questions allow test developers to build tests that are very similar from year to year, both in content and difficulty. However, despite this rigorous process, it is impossible to construct identical tests every year. As a result, the difficulty of the test may vary slightly from year to year. Because of this, comparing test scores over time becomes a challenge. To address this issue, statistical adjustments are made to account for the differences that exist. The process of adjusting scores to account for the differences in difficulty of test forms is known as *equating*.

In the case of MCAS, scores are reported on a 200–280 scale. That is, the minimum score on a subject test is 200 and the maximum score is 280. Regardless of administration (fall 2001 or fall 1999) a score of 240 should have the same meaning, and the goal of equating is to make sure that it does. There are several methods that can be used to equate test scores. Each of the methods is theoretically sound, and there is no "best" method in the sense of a most accurate method. Research has shown[1] that the different methods produce very similar results. Therefore, the choice of method depends more upon the nature of the testing situation and the types of score comparisons desired. For example, it may be of interest to compare scores across several years, as in the case of MCAS. However, in other situations, it may be desirable to compare scores only within a given year. Depending on the specific situation, one method may be preferred over another for ease and practicality of implementation.

Test forms may vary in difficulty from year to year, but students may also differ in ability from year to year. In fact, the hope is that students are becoming more proficient at the skills contained in the content standards. Thus, if the same test were administered to two different groups of students, the performance may be different because one group is more able than the other. Hence, when the questions on the test change, differences in scores can result from two sources: the difficulty of the test or the ability of the students. When scores are equated, it is necessary to make adjustments that account for the differences in difficulty of the test and not the differences in ability of the student groups. This is a difficult task, which requires sophisticated statistical techniques. Since the process is so technical, the details of each of the equating methods will not be presented here. However, the main concepts behind equating will be discussed.

The goal of equating is to adjust test scores to account for differences in difficulties among test forms. To do so, a reference test must be identified. Scores on any other test form will be adjusted such that they are comparable to those of the reference test. So, for example, in MCAS, when test scores from

Given that different tests are administered each year, how can we be sure that the changes in performance are due to changes in student achievement and not due to changes in test difficulty? The issue central to this question is how to disentangle the impact on performance of the different tests and differing student abilities.

a 1999 administration are compared to scores from the 1998 administration, the 1998 administration is the reference form. Therefore, to compare the scores from the 1999 to those obtained in 1998, the scores from 1999 are adjusted and made comparable to the scores from 1998. In this case, we can consider the 1998 test form as Form A and the 1999 test form as Form B (note that this concept can be generalized such that any administration other than 1998 can be considered Form B). The group that takes Form A will be referred to as Group A, and similarly, the group that receives Form B will be referred to as Group B.

To simplify the concept some, consider the following scenario. Assume Form A is identical to Form B. Group A performs worse, on average, than Group B. The conclusion would be that the differences in scores are due to differences in abilities of students: those who took Form B knew the material better than those that took Form A. This conclusion is reasonable, and it is this fundamental concept on which the idea of equating is built. In the instance that the two test forms (A and B) are not identical, the situation becomes more difficult, as now both the students and the questions change. This situation is more typical, and it is, in fact, the case with MCAS. Because it is, in general, impossible to decompose the differences in scores into the two sources (students and questions), it is necessary to hold one factor constant: either questions or students. Clearly it is not possible to have students take both Forms A and B. It is also unreasonable to keep the same questions on Form A and Form B, as the test items used to determine a student's score must be made available to the public. Therefore, a compromise is made.

In the case of MCAS, as well as many other large-scale assessments, the compromise is to keep a *subset* of items common from year to year. This subset of items is often referred to as an *anchor test*. By keeping some items common, the performance of the students on the anchor test can be compared. This scenario then reduces to the simple scenario above, where Form A and Form B are identical. In this case, the anchor test is identical for both groups of students; thus the relative ability of the two groups can be assessed using this anchor test. For example, in comparing the scores of Group A to Group B on the anchor test, if Group B scores higher than Group A, then the conclusion is that Group B is more able than Group A (on this set of items). The differences in ability are now known and can be taken into account when the scores are compared. The remaining differences in the scores are then attributed to the differences in difficulty of the two test forms, and adjustments can be made for these differences.

In order for this process to be reasonable, the subset of items that is chosen for the anchor test must be similar in content and format to the items on the entire test. That is, the anchor test should be a miniature version of the whole test.

So, if the whole test consists of multiple-choice items as well as short answer items, then the anchor test should also contain both types of items. Similarly, if there are five content areas represented in the entire test, then these same five content areas must also be represented in the anchor test. In controlling the content and format of items in this way, the differences in ability on the anchor test should be representative of the differences in ability in the entire test.

The final issue to resolve is the mandatory disclosure of items used to score students. Federal law requires the disclosure of all information in a student's record. Since the MCAS test scores are reported at the student level, parents have the right to see the items used in the computation of a student's score, which causes a problem in developing an anchor test. If the items of the anchor test were released to the public, then students in 1999 would have access to the items that were used in 1998. The students could then memorize the answers to these questions, resulting in a much higher performance on the anchor test than would be expected. Therefore, the differences in the abilities of the two groups would appear to be much larger, resulting in an incorrect assumption about the relative difficulty of the two tests. Then, when the statistical adjustment is made, the scores of the students in 1999 would be inflated, resulting in the misclassification of students. To remedy this problem, the items in the anchor test are not disclosed. Since they are not disclosed, they cannot be used toward the student's actual score. This practice results in a test composed of two parts: items used for scoring purposes and items used for equating purposes.

In the case of MCAS, scores must be equated back to the 1998 form so that growth in student achievement since 1998 can be accurately assessed.

Once the statistical adjustments have been made, the resulting scores from Group B on Form B are comparable to the scores Group B would have obtained on Form A if they had taken it. Therefore, a comparison of scores across forms (i.e., years) is possible. In the case of MCAS, scores must be equated back to the 1998 form so that growth in student achievement since 1998 can be assessed.

As part of the Department of Education's ongoing concern regarding the valid use of test scores, the Center for Educational Assessment (CEA) at the University of Massachusetts Amherst has been contracted to provide quality control for the equating of MCAS test scores. Each year, the CEA does the equating of test scores. The results of these analyses are then compared to those of the company contracted to develop and analyze the test to ensure that the equating was done properly. It is only after equivalence between the two sets of analyses can be established that the equating is considered final. In the event of a discrepancy, measures are taken to determine the explanation. Once an explanation is found, a judgment is made as to which method is most accurate, and the results of the more appropriate method are used.

ENDNOTE

[1] Hanson, B. A. & Béguin, A. A. "Obtaining a Common Scale for Item Response Theory Item Parameters Using Separate Versus Concurrent Estimation in the Common-Item Equating Design," *Applied Psychological Measurement, 26* (1999). pp.3-24.

Kim S. H., & Cohen A. S. "A Comparison of Linking and Concurrent Calibration under Item Response Theory," *Applied Psychological Measurement, 22* (1998). pp.131-143.

Li, Y. H., Griffith, W. D., & Tam, H. P. *Equating Multiple Tests via an IRT Linking Design: Utilizing a Single Set of Anchor Items with Fixed Common Item Parameters during the Calibration Process*. Paper presented at the annual meeting of the Psychometric Society, Knoxville, TN, June, 1997.

Lisa A. Keller (Ed.D. 2002) is a lecturer in the Department of Educational Policy, Research and Administration. Her main research interests are in the areas of test equating, computerized adaptive testing, and item response theory.

VRROOM: THE VIRTUAL REFERENCE ROOM

Using the Internet to Support Teaching the Massachusetts Curriculum Frameworks

Robert W. Maloy, David Hart, Paul Oh, and Victoria Getis

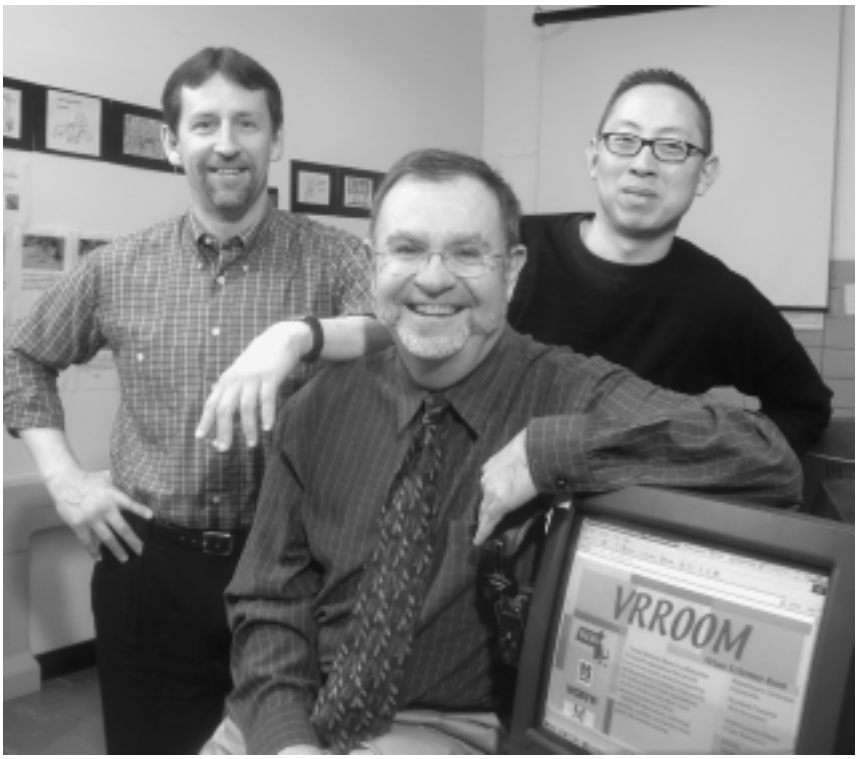
Ten years after the passage of the Massachusetts Education Reform Act (MERA) of 1993, classroom teachers, college students preparing to become teachers, and teacher educators in colleges and universities face a dramatically changed educational landscape. State-wide K–12 curriculum frameworks in English, mathematics, science/technology, history and social science, foreign language, arts, and health not only outline what academic content to teach at different grade levels, but demand that all students be educated to high levels of academic achievement. Massachusetts Comprehensive Assessment System (MCAS) tests add a high stakes dimension to these frameworks since students who do not pass the tests cannot graduate from high school.

Curriculum frameworks and high stakes tests mean that experienced educators and new teacher candidates must emphasize teaching of academic content at every grade level. Increasingly, they turn to the Internet for content resources and instructional strategies. But the Internet is not a one-stop source of valuable information. General search engines put teachers in touch with vast amounts of material, but not necessarily the resources needed to create effective lessons. Teacher-oriented sites provide step-by-step lesson plans, but it is unclear as to how they are judged for quality. As the hits on a Google search pile up, teachers quickly encounter online information overload without the time to determine what is useful and what is not.

Enter VRROOM—the Virtual Reference Room—an online resource for Massachusetts teachers and new teacher candidates being developed by the Center for Computer-Based Instructional Technology (CCBIT) in the Department of Computer Science in collaboration with teacher education programs in the School of Education at the University of Massachusetts Amherst. (It is available online at http://ccbit.cs.umass.edu/vrroom.) Its goal is to improve the academic content knowledge of college students who are preparing to teach and support the professional development of veteran teachers by providing rapid electronic access to challenging and engaging Internet-based academic resources.

Our VRROOM collaboration extends beyond the campus, bringing together organizations devoted to using the Internet, video, and other technologies to improve teaching and learning in public schools throughout Massachusetts, including public television station WGBY (Channel 57 in Springfield), the Massachusetts School Librarians and Media Specialists Association (MSLMA), the MetroWest Library Group of Greater Boston, and the Virtual Education Space (VES) initiative of the Massachusetts Department of Education.

Prior to last year, each partner organization had been working independently to link academic content and technology resources to the state's K–12 curriculum frameworks. Now working in VRROOM, with a common homepage, the partners intend to provide educators around the state with easy and rapid access to curriculum materials and instructional



David Hart, Robert W. Maloy, and Paul Oh

ideas, all linked to the curriculum frameworks (see VRROOM Partners on p. 21).

STANDARDS CONNECTOR

The University's contribution to VRROOM is the Standards Connector. It features active links to web-based resources for teaching the core content of the state's English, mathematics, science/technology, and history and social science K–12 curriculum frameworks. (Eventually, we plan to add links for the state's foreign language, arts, and health frameworks). Now in its second year, Standards Connector is a web site in progress, consisting of a growing collection of primary documents important to the study of world and U. S. history; interactive resources for improving the reading and writing skills of children and adolescents; and interactive mathematics and science activities for different grade levels.

The Standards Connector was initially intended as a resource for certification candidates in the School of Education's elementary and secondary teacher education programs. These college students were already familiar with the Internet, and many used the web to research curriculum ideas. Our idea was to give students in teaching methods classes a way to connect excellent web-based resources to the curriculum frameworks they would be expected to teach when they did their student teaching.

We began by entering the content of the state's curriculum frameworks into a database. With the assistance of graduate student researchers and University faculty, we began identifying Internet-based resources for meeting these content expectations. Next, we gathered feedback about the site from new teacher candidates attending courses in the School of Education and teachers attending regional meetings of the New England Council for the Social Studies and the Massachusetts Computer Using Educators group. From their comments, we incorporated the following key elements into Standards Connector's design and function, building a resource not only for new teacher candidates, but for classroom teachers as well.

• **Web Sites Linked to the Curriculum Frameworks.** Entering the system, users find the key academic content taken verbatim from the state's frameworks presented in green down the left side of the screen with annotated web sites listed along the right side. Web links are active so users can quickly move from Standards Connector to the online resource that they want to explore. Cooperating teachers and new teacher candidates reported having the curriculum frameworks and Internet resources for teaching academic content together in one place was very helpful for lesson planning and classroom instruction.

• **Web Sites Emphasizing Academic Content.** New teacher candidates and veteran teachers told us they valued having access to information from which they can create their own lessons rather than just seeing someone else's already assembled lesson plans. So rather than identifying sites consisting mainly of lesson plans, we looked for web materials that offer primary sources (speeches, photographs, memoranda, art, etc.), visual displays (maps, charts, graphs, interactive models), and infor-

mation summaries (biographies of key figures, background material on important events and concepts).

• **Strategies for Evaluating Web Resources.** To provide new teacher candidates, teachers, and K–12 students with criteria and skills for evaluating material on the Internet, we included a section of essential questions for users to ask when evaluating online materials. The goal is for every teacher and every student to be a thoughtful, critical evaluator of web resources, aware of the direct and indirect messages that web sites may contain. At the same time, to make our site as instructionally reliable as possible, all selections included in Standards Connector have been checked by faculty members or graduate students and vetted by a second reviewer.

• **Multiple Resources.** To offer educators a variety of resources as starting points for curriculum planning, we developed a system for identifying the kinds of web-based material posted in the web site. Internet resources are designated as either "teacher resources," "student resources," "lesson plans," "unit plans," "web-based activities," or "historic documents." Some sites have more than one designation. In addition, we built a separate teacher resource section with more than 50 general lesson plan sites containing thousands of curriculum ideas, each annotated to describe its emphasis, commercial or non-profit status, and strengths and weaknesses. A list of search engines is also part of the "teacher resources" section.

VRROOM Partners

Standards Connector has been created by the Center for Computer-Based Instructional Technology in the Department of Computer Science in collaboration with teacher education programs in the School of Education. It includes web sites linked to specific academic content expectations in the state's curriculum frameworks. Available at http://ccbit.cs.umass.edu/standardsconnector/.

Selection Connection, developed by Massachusetts School Library and Media Association, features notable children's and adolescent literature linked to the required content areas in the curriculum frameworks.

Video Lending Library, of public television station WGBY (Springfield), offers an extensive collection of PBS television programs, broadly coded to the curriculum frameworks. All videos loaned at no charge to teachers in western Mass. Available at http://www.wgby.org/edu.

MetroWest Massachusetts Regional Library System, the newest member of the VRROOM Collaborative, provides resources relating to the frameworks for librarians in greater Boston.

E-TEAMS

One of the interesting contradictions surrounding technology in schools is that while most teachers use computers (93 percent of teachers in grades 4–12 report use of computers professionally at school, at home, or both.[1]), only about one in four teachers report "significant" use of the web at school. Among veteran teachers, the percentage is even lower (13 percent for those with 20 or more years in the field).[2] In addition, there is a continuing digital divide between affluent and less affluent communities, with just over half of teachers in low-income and minority districts having access to the Internet at school.[3] As a result, many students are not seeing how technology can support the learning of core academic subjects.

Given these findings, a collection of electronic resources for teaching the curriculum frameworks is only part of a solution to the problem of integrating technology more completely into teaching and

Standards Connector as a Resource for Classroom Teachers, New Teacher Candidates, Parents, and Students

- presents excellent resources (print, Internet, video) directly linked to curriculum frameworks.
- contents screened by UMass School of Education graduate students and faculty.
- offers lesson plans, unit plans, Web-based activities, teacher resources, and student resources.
- serves as teaching tool when user implements lesson plans or demonstrates concepts through Web-based activities.
- extends learning as students explore resources and Web-based activities in classroom and beyond classroom.
- familiarizes students and teachers with technology as they navigate site and links.
- helps new teacher candidates learn effective uses of technology in classroom.
- up-dated so teachers can learn about new frameworks and find teaching tools linked to those frameworks.
- provides homework assistance and research resources for students.
- familiarizes parents with curriculum content mandated by state and taught in schools.
- available free to school districts and individuals.

learning in schools. An extensive database coded to the curriculum frameworks is only useful if teachers, new teacher candidates, and public school students use the resources to create lively, engaging classroom activities. Indeed, studies have shown that teachers who fully integrate computers, the Internet, and other technologies in their classrooms are more likely to engage in inquiry-based, problem-solving forms of teaching with their students.[4]

In an effort to support classroom teachers in increasing their digital literacy, the VRROOM partners have embarked this year on the creation of E-TEAMS to serve high-need school districts throughout the area. An E-TEAM consists of the following members:

- **a public school teacher** who is a participant in the National Teacher Training Institute (NTTI) program offered by public television station WGBY,
- **a content coach,** generally a graduate student from the School of Education, well-versed in the uses of technology to enhance teaching and learning, and
- **a resource person** from either the Center for Computer-Based Instructional Technology staff or the School of Education faculty.

The coach acts as a technology mentor to the teacher who has the responsibility of creating technology-infused lessons over the course of the year that will be added to a national database of NTTI lessons. In this mentoring role, the coach provides expertise and research support, teaches model lessons in school classrooms, answers questions, and serves as a guide to web and video resources. This way, teachers and coaches together can take advantage of the instructional potential of the Internet and other digital media.

CONCLUSION

Last spring, at a campus-wide meeting of faculty and administrators, Mark Weston of Apple Corporation likened recent developments in educational technology to the evolution of an ultra-modern highway system. In the 1990s, affordable, powerful computers burst on the scene, providing educators with truly fast vehicles. As the decade progressed, the Internet, electronic mail, and other systems allowed for the networking of these computers, giving teachers and students superhighways on which to drive their machines. Today, a new era of change is upon us, one in which educators must decide how to best realize the potentials of computers and networks. In this revolution, the focus will shift to the *content* of technology and the *competencies* of users—that is, to how we provide fuel for the cars and skills for the drivers.

The analogy of technology and highways extends to curriculum frameworks and education reform. In schools today, teachers must focus on academic content and the competencies of students. VRROOM, the Virtual Reference Room with its Standards Connector resource web site and its new idea of classroom E-TEAMS, seeks to advance the content knowledge and technological competencies of new teacher candidates, classroom teachers, and K–12 students. Technology alone cannot ensure that successful teaching and learning will happen in schools. Given guidance, support, and resources, educators and students alike will make greater use of the electronic tools at their disposal to experience lessons and activities full of innovation and learning.

ENDNOTES

[1]Becker, H. J., Ravitz, J. L., & Wong, Y. *Executive Summary: Teacher and Teacher-Directed Student Use of Computers and Software,* November, 1999. [Online]. Available: http://www.crito.uci.edu/TLC/findings/ComputerUse/html/body_startpage.htm.

[2] "Percentage of Public School Teachers Reporting Significant Use of Computers, Email, and the Internet at School, by Years of Teaching Experience." *Connection*, Fall, 2001. p. 17.

[3] "Digital Teaching." *Connection*, Fall, 2001. p. 17.

[4] Becker, H. J., & Ravitz, J. L.,"Internet Use on Teachers' Pedagogical Practices and Perceptions." *Journal of Research on Computing in Education*. Summer 1999, Vol. 31, No. 4. pp. 356-384.

Robert W. Maloy (MAT 1971) is a lecturer in the Secondary Teacher Education Program where he coordinates the history and political science teacher certification program, co-directs the 180 Days in Springfield Project, co-directs the TEAMS Tutoring Project, and co-directs the Virtual Reference Room (VRROOM) and Standards Connector initiatives.

David Hart is Executive Director of the Center for Computer-Based Instructional Technology (CCBIT) in the Department of Computer Science. CCBIT created and maintains the VRROOM and Standards Connector web sites.

Paul Oh is a technology resource teacher at the Mark's Meadow School in Amherst, Massachusetts, and K-12 Projects Director at CCBIT.

Victoria Getis, formerly K-12 Projects Director at the Center for Computer-Based Instructional Technology, is an instructional designer at the Institute for Technology Enhanced Learning and Research at The Ohio State University.

School Accountability

Susan Bowles

Accountability has become a fact of life in public education systems across the United States. We have student accountability measures that lead to adult accountability measures for schools and districts, but what exactly is "adult accountability?" Policymakers identify adult accountability systems, like school accountability, as an essential characteristic in improving the public education system. Bolstered by federal legislation and grounded in state education reform movements aimed at providing adequate funding levels for all students and schools, particularly those in low income communities, accountability is the stick that makes the carrot of increased funds from the state to all schools palatable to legislators and the public alike.

Heading into uncharted territory, states take the lead in holding adults in the local schools and districts accountable for student outcomes, but what if a school fails to meet the standards set by the state? The answer to this question is different in each state and has been implemented only in the most dire cases. School accountability is the first step on the journey to hold the adults in the public education system accountable for student outcomes. The stakes are high. If the past decade of standards-based reform is to be heralded as a success, states must move beyond the student accountability systems of tests toward the hard work of improving schools for all our children.

In 1993, the Massachusetts Education Reform Act (MERA) was passed by the state legislature. Aiming to ensure adequate funding to all students within the public education system, the state leaders mandated a comprehensive accountability system be put in place to account for the increase in state aid to schools. A decade after the passage of the legislation, the state has just begun to hold schools accountable for student performance. To understand school accountability in Massachusetts, it is helpful to use the analogy of the state as the doctor and the school as the patient. Under an accountability system, the state (as the doctor) is required to assess the health of the school (as the patient): to observe symptoms, to make a diagnosis, and to prescribe treatment. Clearly, comprehensive reform of public education as mandated by MERA has added to the state's traditional role in public education. School accountability is creating a scenario in which leaders in the state education agencies are becoming the doctors of public education with under-performing schools becoming their patients. But, what is the cure that the doctor can provide for the patient? The complexities of the public education system make a universal remedy impossible. The research points to unique needs and custom approaches for nursing an under-performing school back to health. This need for custom approaches may require greater amounts of expertise and greater financial resources than the states currently have.

The adult accountability system in Massachusetts starts at the school level, utilizing student scores on the Massachusetts Comprehensive Assessment System (MCAS) to determine

schools in need. Mirroring initiatives in other states, MERA mandates a foundation level of funding for all schools in the Commonwealth (the carrot) in exchange for a comprehensive accountability system based on student outcome measures (the stick). The Commonwealth is at a critical juncture in the implementation of accountability. If MERA is to be deemed successful, it will need to move beyond these student tests to transform the delivery of public education in schools that currently are not meeting state standards.

The state legislature has kept its promise to provide state funds to schools as set out in MERA. However, the larger system of adult accountability envisioned within the legislation is ready to begin without the state resources needed to implement the system. In fiscal year 2002 the state spent nearly 77 percent of the education budget on funds going directly to schools. Facing a declining state economy, the percentage of the state budget went up for schools between fiscal year 2001 and 2002, however, the funds for the adult accountability system decreased from .16 percent of the state education budget to .07 percent.[1] Consequently, as of spring 2002, the state had conducted diagnostic team visits to only 32 out of 80[2] schools, or 40 percent of schools considered to be critically failing and not meeting improvement goals for student performance in the first rating cycle. Of those schools visited, only a handful with the most flagrant problems have been labeled "underperforming" and placed under the auspices of state and district attention, support, and technical assistance. Essentially, the state must rank schools in need from the bottom-up to function within the parameters of limited resources, personnel, time, and money that currently exist. Rather than ramping up the state adult accountability system, there has been a decrease in resources available. This undermines the concept of systemic reform driven by the comprehensive accountability system envisioned in MERA.

The first rating cycle for schools within the Commonwealth was based on 1999–2000 MCAS results. An analysis of the first rating cycle reveals 80 schools considered "critically low" and "failing to meet" improvement goals, and an additional 195 considered "very low" and "failing to meet" improvement goals. Because of the high number of schools and the limited capacity of the state, a balance had to be struck to determine those schools that are most in need.

The first school diagnostic intervention team visits occurred in spring 2000. During this part of the process a team of state and local educators familiar with the improvement plans of the potentially underperforming school, visit the school for a half a day, observing classrooms and interviewing staff and leaders within the school. As of fall 2002, a total of 32 schools had been selected for review, and the commissioner of education declared a total of eight schools underperforming. Interestingly, another 15 schools were initially declared under-performing, but when the superintendents and principals of these schools were notified, they appealed the decision resulting in a new "under review" cat-

egory. The remaining nine schools were found to be able to improve without outside intervention from the state.

This dynamic of the superintendent and principal challenging the under-performing determination made by the state is worth investigation. While there may be exceptions, most principals, superintendents, and parents want to provide the best possible education for students. However, when schools were declared to be unhealthy (under-performing), their communities, superintendents, and principals fought back rather than accept the "support" from the state. A principal interviewed about a team visit to her school said,

> …[O]ne thing I can tell you about the review: [it is] stressful… [I]t is how you market what you personally believe and … your mission and vision of taking your institution to another level. You have to be able to articulate that, and, above all, you have to be in the position that if you think the [team visit] was not done properly, be ready to defend and be ready to rebut what you think is not right because we are talking about how [this will] effect the lives of students, the profession of teachers, the belief system of the public in an institution…

Plainly threatened by the state visit, the principal's comment reveals a crucial difference in the state and the local vision of accountability. In the eyes of the state, a school may have symptoms of illness and be diagnosed as having a curable disease. The state (as the doctor) will nurse the school back to health by providing extra care and negotiating a prescription for treatment with the school (in the role of the patient). The school, however, views any diagnosis made by the state as life threatening and even terminal. The natural response of the school is to seek a second opinion before accepting diagnosis and to take drastic measures to fight for its life.

This difference in understanding may be attributed to the recurring issue of local versus state control of the public education system. Adult accountability measures within MERA move the responsibility for educating children, once entirely in the hands of the local community, to the state. School accountability alters the traditional power boundaries and contributes to the perception that rather than *helping* the school, the state will be *usurping* control of the school away from the community. The media contributes to this fear, as substantiated by a principal in one of the potentially under-performing schools. He voiced his dismay, referring to the way "the state and the media had used this visit to shame [the] school within the community." In a district during the second year of state team visits, a superintendent remarked, "The media can make this process devastating for the commu-

School Intervention Decision-making Framework

Each state accountability system has its own distinctive qualities, but to develop any accountability system each must work through a decision-making framework. The following model developed after interviews and research on the activities of states currently embarking on the journey of adult accountability.

Performance Criteria are the standards and measures according to which a school (or district) is declared to be under-performing. These criteria may include: student test scores, attendance rates, dropout rates and/or student scores disaggregated by race, ethnicity and SES.

Strategic Criteria are other criteria used to determine in which school or district situations the state will intervene and to what degree. Factors such as state capacity and political considerations heavily influence these decisions.

Diagnostic Intervention is a first stage of intervention in which the state analyzes the situation of a school or district to determine the appropriate tactics and targets for corrective action. (Note: Many states have a general approach, such as a focus on "instruction" or "leadership." Interventions will often be customized within the boundaries of the general approach.)

Corrective Interventions emerge from the diagnostic process and include a wide range of potential treatment/remedies, including self-study, financial assistance, expert assistance, professional development, rewards for progress, on-site audits/monitoring, probation, suspension of accreditation, transfer or replacement of staff, transfer of/choice for students, takeover, reconstitution, or closure.

Exit Criteria are the criteria according to which a decision to conclude or change the intervention is made and the school is determined to be healthy or on the road to recovery.

Source: Bowles, S., Churchill, A., Effrat, E., & McDermott, K. *School and District Intervention: A Decision Making Framework for Policymakers.* Center for Education Policy: Amherst, MA, 2002.

nity…" She added that because of the potential devastation, "…the superintendent must step up to work with staff and buffer the school." A doctor and patient relationship must be a partnership with some level of discretion if it is to work—a difficult proposition in the public arena. Perhaps it is with this understanding that since the first year of school accountability visits, the state has made special efforts to soften the blow within the communities whose schools are potentially under-performing. Efforts to disseminate information, enlist district superintendent support,[3] and provide more information to principals to manage fears and anxiety within the community and among the staff within the school, reveal the state's ability to improve upon school accountability practices.

The state and district are the medical team that watches the school closely and works with the school to develop a regimen for self-improvement. If the state is the doctor, after the second year of visits the district is conscripted to be the nurse for a school that is found to be under-performing. Each of the under-performing schools receives technical assistance in using data to drive improvement planning, additional funds, and greater attention from the state and, ideally, the district. District superintendents in partnership with the state Department of Education monitor and provide technical assistance to the school principals and staff members. Still in its earliest phases of implementation of adult accountability, the state has taken an evolutionary approach to developing policies and strategies to support schools, learning from each experience and approaching school improvement as something in which the school, as the patient, must find the strength to improve from within, albeit with the direction and support of the state, serving as the doctor.

If the patient does not show signs of recovery, the doctors take drastic measures for a new, more intense level of treatment. According to MERA, after two years of showing no improvement by the subjective judgment of the Board of Education, a chronically under-performing school may be reconstituted. By definition, school reconstitution is the disbanding of the staff and faculty within a school and starting from scratch in the most extreme measures. In Massachusetts, reconstitution means that the state Board of Education may fire a principal, and the local district superintendent may appoint a principal with "extraordinary powers" to hire and fire employees. In the 2002 Quality Counts published by *Education Week,* Massachusetts is reported as one of 15 states with reconstitution powers. To date, the Massachusetts Board of Education has not reconstituted any school due to student performance on MCAS.

As stated earlier, the state assuming the role of doctor is new to the field of school accountability. MERA dramatically altered the state's role in public education. Since 1993, the Massachusetts Department of Education has taken on many new roles including that of developer of state standards and assessments as well as monitor, enforcer, and supporter of those districts and schools unable to perform at the levels established by the Massachusetts Board of Education. Required to make determinations about whether a school is meeting the standards set forth in MERA, the state has grappled with the conflicting roles of supporting struggling schools and regulating them.

Following in the footsteps of other states, one might look to those who had gone before to make some determinations about what helps and improves schools. Unfortunately, the cure is still not clear. School improvement, as determined by student academic achievement, is complex. It takes years to accomplish, and each scenario of school improvement is attributed to unique local circumstances. For example, a charismatic principal or teacher from inside the school with exemplary leadership skill can move a school into transition and improvement, so researchers cite leadership as an important component of school improvement. This and other recommendations about school improvement seem obvious from a distance, but the prescription for any school treatment must take into account the existing resources within the school. Customized prescriptions based on expert analysis are needed; consequently, broad ideas and suggestions for improvement are not always helpful and are more likely situational. So, the Commonwealth continues to search for ways to improve schools where students are not achieving at state-mandated levels.

We already know the Massachusetts public education

system is failing students by the standards developed by the state. Only 69 percent of the state's tenth graders (Class of 2003) passed both the English language arts and mathematics MCAS exams in spring 2001. This is significant because members of the Class of 2003 will be the first required to pass the English and mathematics MCAS in order to receive a diploma. After the spring 2002 retest results were released about 80 percent of the Class of 2003 had passed, meaning that nearly 12,000 students had not passed both tests and were not eligible for a diploma. Yet, as of the spring 2002, only two high schools, out of 45 rated "critically low" and "failing to meet" improvement goals, have been reviewed by a state team for under-performance. This is clear evidence that the adult accountability system in Massachusetts is too far behind the student accountability system and lacking the resources to be effective.

Since 1993, Massachusetts' education reform has revealed the gaps in student achievement that exist among schools in the Commonwealth. Now that the doctors have identified the test to determine illness and dysfunction in schools, it is time to diagnose and treat the deeper causes of this problem. Comprehensive reform, legislated in MERA, has created a situation where the students have become victims of an accountability system that is using them as symptoms to diagnose larger problems within public education. We are, essentially, treating the symptoms of a cold, when pneumonia may be lurking deeper within the system. Faced with a limited supply of resources for doctors and high demand for them from the undiagnosed patients with symptoms of illness, the state faces a dilemma.

The legislature and the governor have kept the promise of providing additional funds to schools, but to keep the promise of holding schools accountable, a significant investment in state-level accountability is vital. Otherwise, we will continue to treat those schools that are in the most need, while those that may need less, but are still failing by state measures, go untreated. On their own, these schools may improve, or they may not, but without a state system available to support them, Massachusetts leaders are gambling with what may become an epidemic of chronic under-performance. Challenges abound. There is no clear or universal remedy for under-performing schools. It may take years to improve those schools with the most egregious problems, and researchers and state leaders may have difficulty putting a price on the endeavor of healing schools. Adequate funding of our schools is essential, but to truly transform education through the existing legislation, we must also be willing to invest fully in the doctors who will work to support and to improve the education of children in all Massachusetts schools.

The Federal No Child Left Behind Act Expands the State Sanctions to Under-performing Schools

Recent federal legislation reauthorizing the Elementary and Secondary Education Act passed in late 2001, better known as the **No Child Left Behind Act,** increases the state's powers to intervene within a school so that after four years of chronic underperformance, or not making "adequate yearly progress," a school may be reconstituted, taken over by the state, placed under private management or replaced by a charter school. This in effect increases the threat of the state usurping local control of schools in all states wishing to continue receiving federal education funds.

ENDNOTES

1 These figures are based on data from the Massachusetts Board of Education "Budget Analysis" of K-12 education found within the *2001 Board of Education Annual Report* and the state budget for the Office of Educational Quality and Accountability, part of the Governor's office and outside of the Department of Education.

2 Figures calculated from the Massachusetts Department of Education's *School Performance Ratings 2000* (September 2002).

3 Referring to a superintendent's group comprised of superintendents from larger urban districts that met regularly with the Director of Accountability and Assessment at the Massachusetts Dept. of Education in the winter and spring of 2001.

Susan Bowles (M.Ed. 2002) is currently pursuing her doctorate at the University of Massachusetts Amherst School of Education. She is a research assistant at the University of Massachusetts Amherst Center for Education Policy and works closely with staff and faculty at the Center on a variety of Massachusetts K-12 education policy projects.

Managing to Lead in the Decade of Education Reform

Francis L. Gougeon

Early in his presidency, Franklin D. Roosevelt commented on his effort to reform the federal government. He said,

> The Treasury is so large and far flung and ingrained in its practices that I find it almost impossible to get the actions and results I want…But the Treasury is not to be compared to the State Department. You should go through the experience of trying to get any changes in thinking, policy, and action of the career diplomats, and then you'd know what a real problem is. But the Treasury and State Department put together are nothing as compared to the Navy….To change anything in the Navy is like punching a feather bed. You punch it with your right and you punch it with your left until you are finally exhausted, and then you find the damn bed just as it was before you started punching. [1]

I would venture to say that more than a few Massachusetts school superintendents and principals felt much like President Roosevelt when confronted with the education reform mandate of 1993. To say that the public education system is analogous to the Navy would be an understatement. Reform of the system has been an agenda item for decades, always, it seems, with the same result: an adjustment here or an adjustment there, the appearance of some change, with the "steady state" prevailing in the end. The feather bed emerges slightly rumpled, but otherwise unscathed.

STRUGGLE TO REFORM

The reason our reform efforts struggle to produce significant or lasting results is because they are legislated and commanded by the state with great specificity and detail. Oversight and regulation, with accompanying paperwork, become the order of the day and the time committed to such things is taken from other endeavors relevant to the education of children.

Our current effort started out broad-based, with a promise to deliver systemic change, and then quickly seemed to focus on the things government can actually control and afford, such as curriculum frameworks and high stakes testing. By 2003, any child who has not passed the MCAS test in mathematics and English will not graduate from a public high school. Not to be outdone, the federal government authorized its own version of command reform in the No Child Left Behind (NCLB) Act. Like many other federal initiatives, it is long on regulation (accountability) and short on financial support.

CURRENT REFORM SETS A PRECEDENT

The state-level education reform of 1993, it should be pointed out, is different from previous manifestations because it sets a precedent by moving important aspects of public education from local to state control. Because this was accomplished by shifting from local curriculum standards to state standards (curriculum frameworks and MCAS testing), many are still unaware of what has happened. Similarly, the federal government has become more seriously involved in public education by demanding

adherence to certain regulations in order for local schools to continue to receive federal funds. This new degree of involvement by the federal government is, as many are aware, out of proportion to the overall financial contribution it makes to local school systems. This shift to greater state and federal control of public education has been made in the name of standard setting and accountability, which is something many would agree is a worthwhile goal, while reserving the right to disagree with the method employed to achieve it.

The tension all this has created surfaces periodically in the form of a resolution or pronouncement by local school committees stating that they will award diplomas to their students whether they have passed the MCAS test or not. This is usually followed by an advisory from the Department of Education reiterating that assessment testing is required by state law and any local decision to the contrary will be in violation of that law.

SHARPENING THE FOCUS

When reform reaches the point where we shift from looking at the system as a whole, usually because of declining resources, to focusing exclusively on specifics like curriculum frameworks and testing, we seem to lose sight of the impact this action is having on all aspects of the school organization.

If we take a minute to look at curriculum revision (frameworks) and mandated high stakes testing (MCAS), we see ripples throughout the entire system, bumping up against time constraints (planning and revising), personnel (adequate number of staff and staff development), students (readiness and remediation), budget (adequate resources), and administration (staff development opportunities for retraining).

IMPACT OF REFORM ON ADMINISTRATORS

If we focus on administration in particular, we can see how curriculum revision and testing might be problematic. At the outset it appears to bode well, because the latest reform enhances the authority of superintendents and principals while decreasing the authority of school committees. The idea, which many agree with, is to have less micro-management by elected officials while placing increased responsibility in the hands of on-site appointed officials. The reform superintendent then becomes a CEO and the school committee something akin to a board of directors. The building principal, similarly transformed, becomes a unit or division manager with a corresponding increase in authority. Both superintendent and principal are empowered to make important personnel decisions, as well as decisions regarding budget, curriculum, staff development, and goal setting. It would appear that with this new authority administrators should be in an ideal position to implement education reform, i.e., revise the curriculum, institute standards-based education, and move students along a continuum of improvement from failing to advanced standing.

Unfortunately, the doing turns out to be more complicated than simply expanding administrative authority. The post-1993 administrator, in addition to having greater authority, has also assumed much more responsibility. In the new era, he or she is judged, to a great degree, on how well students do on the state's assessment tests. Success or failure at this point hinges on student scores, as does the designation of a school as exemplary or on probation.

All of this has changed the dynamic within schools and school systems, making the role of administrator much more demanding and complicated. Some even say that schools are dangerously close to obsessing over testing, which can make the administrator's relationship with the staff more difficult. Superintendents may pressure principals, who in turn pressure teachers and staff.

THE BIG PICTURE: CAPACITY OF ADMINISTRATORS

In retrospect, if a serious analysis of the educational system was undertaken before launching reform (some would argue there was), policymakers might have become more aware of the actual readiness of school administrators to undertake the challenge of reform. In other words, since we have a worthwhile goal (improving student achievement), what is our administrative capacity to lead this initiative? Do we have the training and resources in place that are necessary to do the job?

If such a thoughtful analysis had been done, we would have discovered that prior to 1993 administrators, in general, and principals, in particular, were focused primarily on managerial tasks such as budget/finance, interviewing/hiring, supervision/evaluation, organization of the school day and year, as well as the establishment of a variety of schedules. Many school administrators, by their own admission, were comfortable with this well-defined role, and longtime service was the rule in this era and not the exception.

On closer analysis it would have been determined that placing greater emphasis on curriculum revision and mandatory testing would, in effect, substantially change the role of the administrator. It would have been clear that with the advent of standards-based education, much more would be expected from the administrator as the position shifted (out of necessity) from manger to manager/leader—which is no small distinction. Even systems that realized this on their own had problems finding the time and financial resources to do the necessary administrative retraining.

In order to function effectively in a reform environment, administrators would have to acquire an array of leadership skills, such as articulating a vision and developing shared decision making, systems analysis, staff development, performance standards for teachers, mentoring teachers, curriculum alignment, and addressing student failure. These are skills that many administrators tell us they do not have. With the additional issues of time and increased paperwork, it is easy to see why many principals have expressed concern about trying to learn a new role while simultaneously being held accountable for the implementation of education reform.

SUCCESSFUL REFORM

In efforts at education reform, we are drawn to the machine approach because it promises what we want—the illusion of a quick fix. Bolt on this part (the curriculum frameworks) or that part (the MCAS), and the machine (the school system) will run smoothly. However, if we are going to be successful with education reform, we have to stop treating the school system like a machine. Education reform must be based on seeing the system as a whole. We must understand that before we implement standards and testing, we need the time and resources to provide the necessary prerequisites for success.

Finally, President Franklin Roosevelt's view of organizations like the Navy was probably no more comprehensive than our view of public education. Like President Roosevelt, if we do not make the effort to base reform on a complete understanding of the organization, we will continue to "find the damn bed just as it was before [we] started punching."

ENDNOTE

[1]Edwards, III, G. C., & Sharkansky, I. *The Policy Predicament: Making and Implementing Public Policy.* San Francisco: W. H. Freeman, 1978.

Francis L. Gougeon is a lecturer in the Department of Educational Policy, Research and Administration. His extensive teaching and administrative experience includes serving as Superintendent of Schools in the Hatfield and the Granby, Massachusetts, public school systems. He also has served as President of the Connecticut Valley Superintendents' Roundtable and is currently pursuing doctoral studies with a concentration in Policy and Leadership at the University of Massachusetts Amherst.

Capacity to Implement Education Reform

Kathryn A. McDermott

Although passage of legislation grabs all the headlines, it is really only the first step of making policy. Public agencies then begin the far less visible process of policy implementation—translating abstract legislative language into concrete regulations and programs. Implementing policy is basically a matter of ensuring that the legislation has its intended consequences while also heading off negative unintended consequences. When the policy is ambitious, it may be necessary to make fundamental changes in how the affected government agencies function.

The Massachusetts Education Reform Act (MERA) of 1993 fits in this category. Historically, local control has been "the 'Battle Hymn of the Republic' for New England educators." [1] MERA challenged the local districts' autonomy by empowering the state to produce curriculum frameworks, test students' achievement, and hold districts accountable for student performance. Ultimately, the state might even take over management of a failing school or district. In order to carry out these major tasks, the state Department of Education (DOE) had to expand its activities in related areas, such as collection and use of data, professional development for educators, and technical assistance for schools and districts. Although MERA was fairly specific about what the state would be doing, it provided little guidance as to how, and with what resources.

Since 1993, many observers inside and outside the state government have questioned the state's capacity to implement the MERA. Three issues must be considered in analyzing state capacity to implement a law: the staff and financial resources available within the relevant state agencies, the staff and financial resources outside the agency, on which the agency might reasonably expect to draw, and the organizational links among all these resources.

INTERNAL RESOURCES

To a great extent, the seeds of DOE's capacity problem were sown by the same economic circumstances that produced the pressure for MERA. In 1980, DOE had 990 full-time equivalent (FTE) staff. Staffing decreased throughout the 1980s, particularly during the recession of the late 1980s and early 1990s. When MERA passed in 1993, DOE's staff totaled only 325 FTE. As state funds evaporated, the proportion of DOE staff who were in federally funded positions increased from 37 percent in 1980 to over half in the late 1990s. DOE has less flexibility in how it uses federally funded staff because their positions are directly tied to specific federal programs such as Title I.

As economic conditions improved in the mid-1990s, DOE did not recover many of its lost positions. Governor William Weld and his appointees in the other executive branch agencies were committed to curbing growth in the state workforce. Members of the Democratic-controlled legislature might have supported more staff for DOE, but did not want to make the first move and risk being portrayed as big spenders by the Republicans. DOE filled some of its staffing gaps with contract workers (also known as consultants or "03s"), but it continued to face increased

responsibilities with a decreased workforce until only a few years ago. Between 1998 and 2002, 108 new positions were added to DOE. However, the current economic downturn and state budget cuts threaten these gains.[2]

DOE acts as a conduit of state funds for local districts, but does not receive much funding for program administration and evaluation. In fiscal year 2002, for example, DOE received $3.9 billion in state funds, $3.3 billion of which it paid back out as aid to districts.[3]

EXTERNAL RESOURCES

For many early advocates of education reform, DOE's lack of staff did not seem especially problematic. They were committed to a theory of reform as a "tight-loose" process. According to this theory, schools were likeliest to improve when the state was "tight" in specifying goals for student achievement but "loose" in the sense of not closely regulating how the goals were attained. DOE's role would essentially be to establish a framework of incentives for reform, leaving superintendents, principals, teachers, and other providers of educational services to plan and carry out the actual work. This minimal role could be carried out without a large staff.

In a book called *Democratic Governance*, James G. March and Johan P. Olson point out that the capacity of any given governing institution depends in large part upon the capacities of others.[4] Viewed in these terms, the "tight-loose" vision of education reform would have allowed the relatively weak DOE to harness the energy and capacities of local educators, who were also closest to the actual teaching and learning process.

Early in the process, it looked like the "tight-loose" model might work. The DOE went out of its way to involve local educators and laypeople in developing a common core of learning that would unify and provide context for the subject-area curriculum frameworks. By the DOE's estimate, about 50,000 people participated in the process in some way. Educators saw the common core and the curriculum frameworks as part of a larger project of reforming curriculum and pedagogy along constructivist lines, which they supported.[5]

However, the tight-loose assumptions proved inaccurate in two respects. First, many local school districts faced capacity challenges at least as great as the DOE's. Aligning local curricula with state frameworks required skills that had not been part of most educators' training or experiences. Some of the larger school districts had administrators who could lead this process, but the smaller ones did not. Given all these challenges, many educators resisted the process and criticized the state for failing to provide them with adequate support and direction.

The second challenge to the "tight-loose" model was that rather than being "tight," state policymakers' goals changed course several years into education reform. John Silber, who became chair of the state Board of Education in 1996, was an outspoken critic of the public schools, of the "education establishment," and of the education reform assumption that all children can learn. In a 1998 *Boston Herald* interview, Silber said, "When I came on [the board], I found that the board had endorsed the statement that all children can learn at high levels. That happens to be false, and it is irresponsible for anybody to make a claim of that sort."[6]

Under Silber, the board revised the curriculum frameworks, eliminating their constructivist tendencies. The educators who had been involved in producing the common core and the frameworks, and in aligning their schools' curricula with the state requirements, opposed the new direction and criticized the state for changing its goals to suit leaders' whims. Problems with the first administration of the qualifying test for teachers, and state policymakers' charge that the prospective teachers who failed the test were "idiots," reinforced the

impression that the state was deliberately making schools and teachers look bad. Local school committees, administrators, and teachers lost their desire to cooperate with the education reform process. In the absence of voluntary cooperation, the DOE lacks the capacity to make education reform happen.

ORGANIZATIONAL LINKS

Communication and inter-organizational relations have also been problematic in three key sets of relationships: between state and local educational authorities, within DOE itself, and between DOE and other offices within the state government.

Many tasks related to education reform, such as altering curricula in response to students' MCAS scores and determining which schools the state should review because they are potentially "under-performing," require timely sharing of data between state and local education authorities. Local administrators resent the heavy reporting requirements of state and federal education policies. Districts must provide data for 15 different state reports, generate four of their own state-mandated reports, and produce five required annual plans.[7] DOE's staffing shortages have hindered its ability to communicate data back to districts. In a 2001 survey, a majority of local educators indicated that they do not believe DOE makes good use of the data it collects or feeds useful information back to districts and schools.[8]

Within the state, communication and allocation of responsibilities pose problems. When interviewed in 2001, many DOE senior staff identified lack of communication among staff clusters as a problem.[9] Competition between DOE and other state offices over components of MERA have slowed the law's implementation. This problem has been most acute in the area of school and district accountability. Late in 1999, the Board of Education approved regulations on how under-performing schools and districts would be identified. The Accountability and Targeted Assistance cluster within DOE began working on site visits to schools where the state might intervene, and on planning for what interventions would be carried out. However, in the summer of 2000, the state legislature approved a proposal by then-Governor Cellucci to shift the school and accountability function out of DOE and into the Office of the Governor. It took about a year for the governor (by this point, Acting Governor Jane Swift) to appoint members to the Education Management Audit Council and several more months for the Office of Educational Quality and Accountability to begin its work.

WHY THE STATE CAPACITY ISSUE MATTERS

Ten years after MERA, the law's different components have been unevenly implemented. This unevenness reflects differences in the level of administrative or technical challenge that various parts of the law posed for state government. The most crucial parts of the law—reforming the technical core of teaching and learning and holding adults accountable for student performance—have made the greatest demands on capacity and thus have been the least completely implemented. Whatever the outcome of the pending lawsuit claiming that students have not had a sufficient opportunity to learn the material included on the tenth-grade MCAS, the state's shortcomings in capacity and the reasons for those shortcomings are likely to loom large in the trial.

ENDNOTES

[1]Jerome T. Murphy, "Title I of ESEA: The Politics of Implementing Federal Education Reform," in Allan R. Odden, ed., *Education Policy Implementation.* Albany: State University of New York Press, 1991, p. 34.

[2] DOE staffing information is drawn from Kathryn A. McDermott, et al., *An Analysis of State Capacity to Implement the Massachusetts Education Reform Act of 1993* (Massachusetts Education Reform Review Commission, 2001) and Andrew M. Churchill, et al., 2002 *Annual Report on the Progress of Education Reform in Massachusetts* (Massachusetts Education Reform Review Commission, 2002).

[3] Churchill, et al., 2002.

[4] James G. March and Johan P. Olsen, *Democratic Governance*. New York: The Free Press, 1995, p. 107.

[5] Dan French, "The State's Role in Shaping a Progressive Vision of Public Education." *Phi Delta Kappan,* November 1998, pp. 185-194.

[6] J. Sullivan, "True Test: Silber Gives Ed-Reform Answers." *Boston Herald,* May 3, 1998, p. 18.

[7]Churchill, et al., 2002.

[8] McDermott, et al., p. 66.

[9] McDermott, et al., p. 15.

Kathryn A. McDermott is an assistant professor in the Department of Educational Policy, Research and Administration and is an Associate Director of the Center for Education Policy at the University of Massachusetts Amherst.

Conclusions: The Impact of Education Reform after Ten Years

Andrew Churchill

With support from the Massachusetts Education Reform Review Commission, the Center for Education Policy at the University of Massachusetts Amherst conducted a comprehensive study of the implementation status of the Massachusetts Education Reform Act (MERA), resulting in the recently released *2002 Annual Report on the Progress of Education Reform in Massachusetts*. The report presents primarily quantitative data on the Commonwealth's status in five areas: the context for education reform, finance aspects of education reform, standards-based reforms, student achievement and gaps, and capacity for education reform at state and local levels. Based upon this research, we offer the following conclusions.

AREAS OF SUCCESS

1. The legislature has brought virtually all districts to a common level of foundation funding. When the legislature committed itself in1993 to supporting an unparalleled increase in state support for K–12 education, few thought they would actually do it. They did it, and a greater level of district equity is the result.

2. The Department of Education (DOE) has developed high-quality, well-aligned standards and assessments. There is growing recognition, supported by external evaluations, that Massachusetts has some of the best standards and assessments in the country. The use of quality standards to align everything from teacher preparation to classroom instruction to school and district accountability is a truly systemic reform effort.

3. The state is paying more attention to the neediest students. While controversial in their implementation, the system of standards and assessments has served to illuminate the persistent inadequacy of education for many of our young people who might previously have fallen through the cracks. The foundation formula and subsequent appropriations for Massachusetts Comprehensive Assessment System (MCAS) remediation have significantly boosted resources for those students who historically have been the least well-served.

4. Districts are paying more attention to curriculum and instruction. With relatively clear standards, and as the standards stabilize and assessment results come in, schools have been able to re-examine their offerings and align them both internally, from grade to grade, and externally with the curriculum frameworks. The majority of districts appear to have realigned their curricula, teacher preparation programs have examined their standards and alignment, and professional development has been supported as a matter of state policy.

5. Students, overall, are inproving their performance. The passing threshold is still low (Needs Improvement), and the achievement gaps are cause for real concern (see below), but 87 percent of regular-education

students in the Class of 2003 have been able to pass the MCAS to date. English language arts performance is improving across the board. Improved NAEP results and relatively high SAT and TIMMS results further indicate that, on average, Massachusetts students are making significant progress.

AREAS FOR FURTHER WORK

1. Achievement gaps. While regular-education students are for the most part meeting the minimum graduation standard, special education students and participating LEP students, on average, are not. Students from poor districts, in general, are not. And African-American and Hispanic students, on average, are not. Further research is needed to delve into the relationships between these variables to identify key areas of opportunity. At the least, attention to instruction, resources, and a realistic timeframe for success will be important.

2. Mathematics. The state has made progress in English language arts performance, but in mathematics, particularly at the 4th and 8th-grade levels, progress has been less satisfactory. One-third of 8th-graders are still scoring at the Warning level in mathematics, and the percentage in the Advanced or Proficient categories has increased only slightly, from 31 percent to 34 percent over the past five years. Educators have successfully placed increased emphasis on literacy and writing across the curriculum in response to the writing components of MCAS;

new approaches to mathematics now seem warranted as well.

3. Proper implementation for accountability. Effective implementation is as important as high standards and systematic assessments. A number of stakeholders have noted a significant gap between the implementation of student accountability and the implementation of adult accountability. Others have countered that high stakes for students were necessary if students and educators were to take the goal of meeting standards seriously. Having reached this point, it will be important to ensure that students and educators have sufficient "opportunity to learn" resources to enable them to meet the standards (including curriculum aligned to standards, teacher professional development, current textbooks, and time for these to take effect). If the resources and timeline are insufficient, this could have dire consequences for both individual students and the collective reform effort.

4. Post-12th grade options. If Massachusetts is to maintain the expectation that all students can reach common standards, post-12th grade pathways will be an important vehicle for ensuring that the system can accommodate student needs, in terms of both time and resources. As of the publication of our report, no programs, policies, or resources had been finalized for meeting this looming challenge. Post-12th grade pathways to competency determination would represent significant progress toward a system in which standards are held constant and time spent on learning varies with student needs—as opposed to the traditional system of time being held constant and learning varying among students.

5. From Needs Improvement to Proficient. Massachusetts has set an ambitious target of getting all students to meet the minimum standards required for competency determination status. However, it should not be forgotten that Needs Improvement is an intermediate stage, and that the standard of Proficient is the level that more closely approximates the skill levels required for college-level work and/or skilled careers.

6. Educator supply, quality, and distribution. As standards have become more ingrained, attention has begun to shift to the critical importance of highly qualified teachers in helping students meet the standards. Looming retirements of

experienced teachers, insufficient retention of new teachers, teacher shortages in urban schools, teacher shortages in certain content areas (especially mathematics, sciences, special education, and languages), higher expectations, and a dwindling supply of principals present formidable challenges. New methods of preparation, induction/mentoring, professional development, scheduling, and pay will all require consideration and evaluation.

7. Updating the funding formula. Our research appears to indicate that the current formula inadequately estimates costs in some areas. Now that we have a number of years' experience with the demands of education reform, it is time to revisit the formula to ensure that it reflects the real demands of the new environment. In addition to accurately funding the costs of regular education, attention should be paid to ensuring that the special costs of bringing all students to competency are supported equitably.

8. Improved systems for data-based decision-making. MERA has generated a great quantity of potentially useful data, and the implementation of the No Child Left Behind Act will generate much more. Districts are already overburdened with information requests, and they feel that they have little opportunity or capacity to make use of the information that is collected. The DOE has recently made TestWiz software available to all districts to help them analyze their MCAS results. This is a very positive step, but issues of local and state data capacity require continued effort. More attention will need to be paid to making the overall data gathering and distribution system more efficient, connected, and effective—both at DOE and at the local level.

9. Re-connection of policy-makers and educators in the field. A great deal has been accomplished to date, but many front-line educators have come to feel disengaged from the process. As the state thinks about fine-tuning MERA implementation, it will be important for the governor, the legislature, the Board of Education, the Department of Education, and the various education associations to agree that education reform is a worthy goal, that we have much to celebrate, and that we will move forward together.

10. Continued monitoring and evaluation of the reform effort. CEP's Annual Report is a useful step in the collection and analysis of information about the successes and challenges of MERA, but there are a number of areas for future research. These include analyses of the following areas:

- various achievement gaps and how they interact, using student-level data;
- foundation budget formula and the real costs of ensuring that all students have the opportunity to learn in a standards-based environment;
- exemplary programs for helping special education, Limited English Proficient (LEP), and high-poverty students to reach achievement goals;
- roles of postsecondary and community institutions in expanding opportunities for students to meet standards;
- supply and demand "pipeline" for teachers and administrators;
- effectiveness of various teacher preparation, induction, and professional development approaches; and
- data needs and capacity of the state and local districts and how a data system might be developed to accommodate them.

The Center for Education Policy

Through the Center for Education Policy (CEP), University of Massachusetts Amherst School of Education faculty and graduate students conduct studies, convene conferences, and evaluate programs on policy topics relating to K–12 education reform and K–16 educational alignment and transitions. Research topics have included topics such as teacher recruitment and retention programs, state testing and graduation requirements, state intervention options for substandard schools and districts, state capacity to implement comprehensive reforms, and the impact of various policies on college access and attainment.

For more information on this study or the Center for Education Policy, contact Andrew Churchill at achurchill@educ.umass.edu.

Andrew Churchill is Assistant Director of the Center for Education Policy. He has a master's degree in public policy from the University of California Berkeley and has worked in local school districts, state and federal agencies, and national research organizations.