

The Oversight of State Standards and Assessment Programs: Perspectives from a Former State Assessment Director

Pasquale J. DeVito, Ph.D.
Director, Massachusetts Comprehensive Assessment Program (MCAS)
Measured Progress

Common Education Standards: Tackling the Long-Term Questions
Thomas B. Fordham Institute
June 2010



The opinions, thoughts, and perspectives expressed in this paper are solely those of the author and do not necessarily represent the views of Measured Progress, the Fordham Institute, or the individual states highlighted in the document.

Thomas B. Fordham Institute

Introduction

States have spent much time, energy, and financial resources over numerous years developing and implementing their individual assessment programs—yet substantial challenges abound. The landscape across the country is rapidly changing for state assessment programs and the standards on which the assessments have been based. Most states are economically strapped due to limited revenues and are in a “holding pattern” as they wait to see what changes the reauthorization of the federal education legislation will mean for them. Also, states that are successful Race to the Top (RTT) applicants will be required to work in large consortia (which has never been done for a comprehensive assessment program) in order to share in the federal funding.

The purpose of this paper is to provide information and insight into how state assessment programs are governed, how individual state and state-consortium assessment programs actually operate, and how key policy and technical decisions on these programs are made.

To address these issues, this paper is divided into three sections: First, the New England Common Assessment Program (NECAP) is presented in some detail as an assessment consortium that has struggled with a variety of challenges and, for the most part, has successfully overcome the obstacles. In fact, there are currently no large consortia operating that are tackling comprehensive, multi-grade, multi-subject, high-stakes assessment programs. The only current example is the NECAP Consortium developed originally by three states: New Hampshire, Rhode Island, and Vermont. Second, the

Massachusetts Comprehensive Assessment System (MCAS) program is presented as an example of one state's journey in developing, operating, and maintaining a quality, high-stakes assessment program that is considered by many to be one of the best in the nation. Third, relevant features of the assessment programs in three other states (Kentucky, Michigan, and North Carolina) are briefly described to illustrate the different approaches and pressures that have helped to shape those programs.

New England Common Assessment Program (NECAP)

When the No Child Left Behind Act of 2001 (NCLB) became law, the smaller New England states faced a formidable challenge. While each state had developed state assessment programs that they thought were generally effective at measuring student achievement at three grade levels, individually, they lacked the resources to expand both reading and mathematics assessments to seven grades as required. Discussions with the six New England states ensued and, in 2004, three states (New Hampshire, Rhode Island, and Vermont) joined together to form NECAP and to work collaboratively. Currently these states test in reading and mathematics at grades 3-8 and 11 and in writing and science at grades 5, 8, and 11. Within the last year, Maine has joined the NECAP initiative for grades 3-8 reading and math and grades 5 and 8 in writing.

In some ways, the arrangement may seem natural. The initial three states have much in common. They are each small states with histories of local control in their educational systems. They are close geographically, so that meetings could be held within a few

hours drive for any participant. They had common needs, in that no individual state could muster enough financial and staff resources to meet upcoming NCLB requirements. They had some history of collaboration and common planning; Rhode Island and Vermont were the only two states administering the New Standards Reference Exams as part of their state assessment programs. Although they did not develop or own those assessments, they held some common meetings with the contractor and discussed issues together as they arose.

But despite the commonalities, there were substantial hurdles to overcome if the multi-state consortium was to work, including governance, procurement procedures, and determining common and unique components of the NECAP. Very strong support from all administrative levels within the states made it work.

The states spent extensive time together discussing how the collaboration might work and which agreements could be made in common. There were many issues to consider, such as ownership of the assessment, the organizational structure and decision-making, procurement procedures, management of the consortium, etc.

Ownership of the design, development, and operation of the common assessment was a key issue that had to be addressed early in the consortium's existence. The states came to agree that the custom assessments under NECAP would be jointly owned by the states. There was the agreement of joint ownership but no legal definition was posited. They decided to work with a single testing contractor to design, develop, and implement

NECAP but struggled with the procurement procedures. Could the contractor be hired under one umbrella contract or were separate contracts necessary? Because procurement practices can vary widely across states, they decided that it was not feasible to have a single contract but that there could be a single contractor operating under individual contracts with each state.

Staffing commitment was also a key discussion point early on. Initiating and developing a consortium takes a great deal of time. Each of the original three states had an assessment director who, as the lead for their state, would have to devote a substantial amount of time to the effort. Some of their assessment staff members, as well as state content specialists, needed to spend many hours working to build a NECAP team and to discuss the many details related to developing common content standards and quality assessment instruments. Fortunately, there was commitment in each state from the top administrators (e.g., State Commissioners and State Boards of Education) in support of the overall NECAP initiative and the resources that were needed to develop a successful consortium.

The consortium operates as an association of state departments of education, not a formal legal entity. The state assessment directors act as the management team for NECAP.

While the goal is to arrive at consensus across states, if state staff members cannot agree on an important issue, the management team decides on the course of action. Each state

carries equal weight in the decisions, regardless of the size of the student population or other factors.

The consortium uses external organizations to help support activities. To provide management services to assist coordination of functions and decision-making, the NECAP states decided to contract with the National Center for the Improvement of Educational Assessment (Center for Assessment). The Center for Assessment works closely with the NECAP to facilitate reaching consensus on key issues, offer advice on matters important to the NECAP, draft Requests for Proposals as needed, provide research findings to the group, and act as a consistent “critical friend” to the consortium members. Each state also has a Technical Advisory Committee (TAC) of assessment and measurement experts to review technical issues and to provide counsel on all aspects of the state’s assessment efforts, including NECAP. At times, the TACs from each of the NECAP states come together for a group meeting.

Even with all this effort and assistance, the states simply do not have the personnel resources or necessary equipment and systems to operate the NECAP themselves. Like virtually every state in the country, the NECAP states needed substantial assistance from a major testing company to actually operate the testing system. Through competitive bidding, each of the NECAP states contracts individually with Measured Progress, a testing company centrally located in Dover, New Hampshire, to provide comprehensive testing services in support of the program, e.g., test development, test form construction

and production, shipping, receiving, scoring, and reporting. Measured Progress is an active partner with the states, the Center for Assessment, and the various TACs. The testing contractor operates within the agreements set by the consortium, some of which are different than non-consortium programs. The NECAP states agreed that:

- assessment materials would be the same and that they would bear the program name rather than the individual state name;
- a single set of achievement standards would be adopted;
- common administration procedures would be employed;
- common allowable accommodations would be used;
- a single set of reports would be generated; and
- a common administration period would be employed.

It was also decided that only state-level report data would be compiled; results would not be combined or reported across all states; and the release of results would be handled by each state.

The NECAP states, through much hard work and thoughtful discussion, along with the able assistance of its external partners, were able to overcome substantial challenges.

Still, some hurdles remain and others loom on the horizon. Among them are the following:

- People in state education agencies (SEA) frequently change positions within state government or leave state service completely. Maintaining the level of

commitment and understanding of the original players in the consortium if substantial staff turnover should occur could be difficult in the future.

- The addition of another state, Maine, to the mix after six years may change the dynamic somewhat and cause the states to rethink some of the processes that have been adopted over the early years of the collaboration.
- The RTT initiatives, with the emphasis on larger sets of consortia (minimum of fifteen states up to fifty or more) may present a challenge to the NECAP states. The collaboration and agreements that can make a four-state assessment consortium work well might be quite different when trying to get a twenty-five-state consortium made up of more divergent states to work. It is likely that the NECAP states will feel a need to join one or more of the RTT consortium efforts.
- The upcoming reauthorization of the federal education act and RTT efforts may introduce changes in the standards and assessment environment that will cause NECAP to consider changes to its current operation. For instance, the member states may feel that the Common Core State Standards currently in development are not as stringent as the ones adopted for NECAP.

Massachusetts Comprehensive Assessment System (MCAS)

The 1993 Massachusetts Education Reform Law gave rise to the development of a new assessment program called the Massachusetts Comprehensive Assessment System (MCAS). The law specified that the testing program must:

- test *all* students who are educated with Massachusetts public funds, including students with disabilities and limited English proficient students;
- measure performance based on the Massachusetts Curriculum Frameworks learning standards; and
- report on the performance of individual students, schools, and districts.

MCAS test instruments are custom designed to provide inferences about the degree of student achievement of the Massachusetts Curriculum Frameworks. The frameworks documents were developed by the Department of Elementary and Secondary Education (ESE) through a review panel comprised of teachers, administrators, and state department staff. The panel was assisted by content experts. The draft frameworks were then approved by the Commissioner for public review and comment. After considering the comments from the field and the general public, the frameworks in each content area were approved and finalized by the Commissioner.

The official uses of MCAS results include:

- determining whether high school students have demonstrated the knowledge and skills required to earn a Competency Determination (CD);
- providing information to support program evaluation at the school and district levels;
- determining school and district Adequate Yearly Progress (AYP); and
- making decisions about scholarships.

The MCAS is a high-stakes assessment at the high school level so students must reach a Competency Determination passing standard in a variety of subject areas, along with completing local graduation requirements, to receive a high school diploma in Massachusetts. The Commissioner recommends the CD passing standards to the Board of Elementary and Secondary Education and the Board votes on the CD and amendments as needed. The class of 2003 was the first class required to earn a CD to graduate.

The scale and complexity of MCAS require a great deal of effort from many groups to maintain the high quality and success of the program. The development and implementation of the program is managed by the ESE assessment office with a director overseeing the staff. The assessment group is very actively involved in all aspects of the program operations and monitors its contractors closely. The assessment staff works with other groups within the ESE—curriculum, technology, and budget, for example—in managing MCAS.

The assessment staff also garners support and assistance from educators within the state as well as several external expert groups. Massachusetts educators play a key role by serving a variety of committees, including Assessment Development Committees, Test Bias Committees, and standard setting groups. For instance, the Assessment Development Committees, made up of local Massachusetts educators, review the test items that have been developed and piloted to determine item quality and statistical strength, and to suggest revisions. The Test Bias Committees, composed of different

groups of local educators, review test items and related statistical analyses to determine whether the test items are not biased for any groups, do not contain sensitive material, and are generally accessible content-wise for students in particular grade levels. When content standards undergo revisions, it is necessary to reset the achievement-level standards cut points. Standard setting committees of expert educators are established to work with the assessment contractor using common methods of standard setting, e.g., Bookmark methodology. The resulting recommendations are submitted to the assessment staff. The Commissioner and the Board make the decision on the cut-off scores to be used.

MCAS is also supported by the work of external expert groups. The ESE, through its assessment office, has consistently held high expectations for its assessment programs and aggressively manages the entire program. As such, ESE requests proposals from testing contractors to assist in the development, operation, scoring, and reporting of MCAS in response to very detailed specifications in the Request for Response (RFR). ESE selected Measured Progress as its testing contractor for MCAS. The testing contractor has continued to work for several years in close partnership with ESE assessment and curriculum staff and Massachusetts educators in all aspects of the MCAS program. The Measured Progress staff that works on MCAS is deeply involved with virtually every operational and psychometric aspect of the program except for policy issues.

In addition to the assistance provided by Measured Progress, the program also benefits greatly from consultation and advice from staff at the National Center for the Improvement of Educational Assessment through a separate contract. Center for Assessment staff members are often consulted for advice and guidance on key projects and regularly attend meetings with the ESE and the testing contractor. Further, the University of Massachusetts measurement and statistics department provides redundant psychometric analysis to verify the analysis performed by the Measured Progress psychometric staff. The combination of ESE, UMASS, and Measured Progress technical staff analyzing test data independently and reviewing the results together virtually ensures that any anomalies will be identified, fully discussed, and resolved before being used for reporting by ESE.

Massachusetts also utilizes an actively engaged Technical Advisory Committee (TAC) which meets three times a year (January, May, and October) for several days. During these meetings, the TAC provides technical advice on test design, program operation, statistical analysis, and reporting of results. In between meetings, TAC members are often consulted on time-sensitive issues that may have arisen and could not wait until the next scheduled TAC meeting. The TAC is comprised of five experienced professionals, including two university-level researchers, a local school district administrator, a former state assessment director, and an assessment expert from Center for Assessment. A representative from UMASS also attends TAC meetings as a consultant and actively engages in discussions as appropriate, as do Measured Progress senior staff. The meeting

agendas are planned by ESE and Measured Progress with complete briefing materials distributed prior to the meetings. The meetings are chaired by the MCAS director or designee.

MCAS is a solid, well-run state assessment program that has produced a record of success. Overall, students have shown considerable gains in academic achievement on MCAS over time. As with NECAP, there are challenges to the program and next steps that need to be considered. Some of these are:

- MCAS is a single-state effort. Like most state assessments, it was not developed as part of a multi-state consortium or collaboration. The recent federal solicitation for the RTT Assessment Program requires that at least fifteen states (with five states agreeing to be governing states) agree to work as a consortium in order to be eligible for the award. The consortium states must agree to a common set of standards. Massachusetts has put a great deal of effort into establishing and revising its rigorous content standards, so it may prove difficult to consider moving away from the current set of standards on which MCAS is based.
- If the Common State Standards that are now in development are widely adopted, there may be the incentive to develop large-scale assessments based on them. States like Massachusetts that have solid and mature testing programs may face political, economic, and other pressures to significantly change or even abandon their current assessment systems. MCAS has been a centrally controlled

assessment system run by the ESE. It may provide a unique challenge if the decision is made to move to a multi-state assessment initiative.

- MCAS is seen primarily as a strong, centrally controlled testing program managed and imposed by the state. ESE wants to add components to MCAS that emphasize and support interim and formative assessments at the school level. As a start, ESE has received some funding from the Nellie Mae Foundation to develop a test design and specifications for Curriculum Embedded Performance Tasks (CEPT) which would provide high-quality, instructionally sensitive assessment tasks that are an integral part of the instructional process and will support classroom instruction in a more direct way. Securing additional funding and having sufficient staff resources to manage the new initiatives could become a substantial challenge.

A few other examples: Kentucky, Michigan, and North Carolina

To explore a bit beyond NECAP and MCAS, three additional states with long-standing assessment programs were identified. Given the length constraints of this paper, no attempt will be made to provide detailed descriptions of their standards and assessment systems. Rather, selected observations will be made about the governance and operation of the assessment programs and the influences that have helped to guide the development and revision of the programs.

Kentucky School Testing System

The Kentucky School Testing System includes a variety of different measures including core content tests, on-demand writing prompts, and norm-referenced tests in reading and mathematics. The testing program in Kentucky has changed several times as a result of legislation. While the Kentucky Department of Education (KDE) administers the testing programs and makes use of external contractors to help implement them, the Kentucky legislature is very involved in setting the requirements for the standards and assessment programs to the point of identifying advisory groups and specific components of the assessment programs.

For instance, House Bill 53—related to the Commonwealth Accountability Testing System (CATS) and passed into law in 1998—established four advisory groups charged with helping KDE build a better assessment and accountability system. The Legislative Research Committee (LRC) makes appointments to the National Technical Advisory Committee, which advises the legislature and KDE on technical aspects of assessment and accountability. The LRC also appoints the Education Assessment and Accountability Review Subcommittee that advises KDE on the development of the assessment and accountability system. Further, the Governor appoints members to the School Curriculum, Assessment and Accountability Council, which advises KDE on the design of the testing and accountability system and the highly skilled educator program.

The bill also charged the Office of Education Accountability (an auditing agency which is separate from the assessment group within KDE) to advise the LRE and KDE on the testing programs. In short, the governance in Kentucky related to standards, assessments, and accountability seems to be strongly and directly influenced by the legislature.

SB1, a bill passed in 2009 that mandates a new testing system by 2011-2012, is another example of strong legislative influence. Among other provisions, SB1 calls for:

- eliminating the open-response questions requirement;
- requiring writing portfolios but eliminating them from the state assessment;
- changing the high school readiness exam from grade 8 to grade 9;
- eliminating arts and humanities testing from the assessment program; and
- requiring writing assessments consisting of multiple-choice items emphasizing mechanics and editing.

Michigan Educational Assessment Program (MEAP)

Michigan follows a fairly typical governance structure for state standards and assessment initiatives. The State Board of Education approves changes in these programs. Unless legislation is required, the State Superintendent of Public Instruction, who is appointed by and responsible to the State Board, recommends policy and technical changes to the State Board of Education. Prior to deciding on recommendations, the staff of the Department of Education will often consult with educator and technical expert panels and post documents for public review. The governance of the assessment program itself is carried out by the Department of Education's Office of Educational Assessment and

Accountability, headed by a director, with operational and technical assistance from testing contractors for each of the Michigan assessment programs.

For example, MEAP was established over four decades ago, in 1969, by the State Board of Education and was mandated and funded by the state legislature the following year. In 1971 content standards were developed by educator and citizens groups, and these draft standards were sent into the field for extensive review. The Department staff considered all the comments and assembled suggested revisions. As a result the standards were revised and formally approved by the State Board of Education. MEAP and the standards on which it is based have undergone many changes over the lengthy history of Michigan assessments, but the core of the assessment is similar to that of other states in the post-NCLB era.

North Carolina Testing Program

Like most states, North Carolina has operated a state assessment program for many years. Also, like most states, the State Board of Education takes an active role in standards and assessment issues, guided in large part by their representative, the State Superintendent of Public Instruction. The implementation and management of the assessment programs are the responsibility of the Accountability Services Division within the department. The staff members are assisted by testing contractors and a Technical Advisory Committee, along with technical consulting help from North Carolina university staff.

In May 2007, the State Board of Education convened a Blue Ribbon Commission on Testing and Accountability to begin the process of assisting the Board in charting a course toward the next generation of assessments and accountability. While the current North Carolina Testing Program primarily emphasizes multiple-choice item formats for end-of-grade testing and end-of-course testing in a variety of subjects, it was felt that the future necessitated substantial change. The Commission effort resulted in a document that laid out a suggested framework for such change. The North Carolina Department of Public Instruction responded to the “Framework for Change” document by outlining a vision and structure of the next generation of standards, assessment, and accountability. The document portrayed a 21st Century Balanced Assessment System that includes formative assessment, benchmark assessments, statewide summative assessments, and ongoing authentic assessments, which has resulted in part in a new accountability model to measure both absolute performance and growth.

Conclusion

States across the country are struggling with their strategies for accessing the large amount of federal funding that is becoming available under RTT, their own economic shortfalls due to the recent recession, and the funding and continuance of their current standards, assessments, and accountability systems. While the large majority of states have developed their standards and assessment programs individually, the pursuit of significant funding through RTT, the promulgation of Common Core Standards, and the premium being placed on large assessment consortia could result in substantial dilemmas

for states. For example, states that are happy with their current content standards and/or believe that their standards are more stringent than the Common Core Standards may need to make compromises to be part of the future initiatives.

Also, assuming that the initiatives result, at least in part, in common assessment instruments being used in reading and mathematics across sets of states, participating states will need to begin to scale down or eliminate their current assessment programs in favor of the new tests.

Finally, each of the requests for new consortia calls for large groups of states (minimum of fifteen and probably many more). This level of consortia has never been tried for comprehensive general assessment. The NECAP effort, now with four states participating, worked countless hours to make it work. Consortia of divergent states that are three or four times the size of NECAP will present many significant challenges. The next few years should be very interesting in the standards and assessment business.

References

In addition to website searches on state departments of education highlighted in the paper as well as personal communications with various individuals related to those state programs, the following documents were very useful.

Blue Ribbon Commission on Testing and Accountability. *Report from the Blue Ribbon Commission on Testing and Accountability to the North Carolina State Board of Education*. Raleigh, NC: North Carolina State Board of Education, January 2008.

DePascale, Charles A. *Establishing a State Consortium for Assessment: A Discussion of Factors to Consider*. Dover, NH: National Center for the Improvement of Educational Assessment, 2009.

DePascale, Charles A. *The New England Common Assessment Program: Notes on the Collaboration Among Four New England States*. Dover, NH: National Center for the Improvement of Educational Assessment, 2009.

Massachusetts Department of Elementary and Secondary Education. *Ensuring Technical Quality: Policies and Procedures Guiding the Development of the MCAS Tests*. Malden, MA: Massachusetts Department of Elementary and Secondary Education, September 2008.

North Carolina Department of Public Instruction. *Response to The Framework for Change: The Next Generation of School Standards, Assessments and Accountability*. Raleigh, NC: North Carolina Department of Public Instruction, October 2008.

U.S. Department of Education. *Race to the Top Assessment Program— Executive Summary*. Washington, DC: U.S. Department of Education with Recovery.gov., April 2010.