A Comparison of the MCAS and PARCC Assessment Systems

Prepared by Dana Ansel, Ph.D.

Commissioned by the Executive Office of Education
Presented to the Board of Elementary and Secondary Education

October 15, 2015

This page has been intentionally left blank for two-sided copying.

Acknowledgements

About the Author

Dana Ansel, Ph.D., is an independent education policy research and evaluation consultant. She works with public, private, and non-profit organizations. From 2000 to 2009, she was the Research Director at the Massachusetts Institute for a New Commonwealth (MassINC), a nonpartisan think tank whose mission is to promote the growth of a vibrant middle class. As Research Director, Dr. Ansel directed research on a wide variety of topics, including K-12 education, higher education, workforce development, immigration, the aging of the population, public safety, and the Massachusetts economy. During her tenure, The Boston Globe called MassINC research "the gold standard" in the public policy arena. She has also served as the Director of Research and Policy at ConnectEDU, a national education technology company.

This page has been intentionally left blank for two-sided copying.

Table of Contents

This page has been intentionally left blank for two-sided copying.

As we enter fall of 2015, the "test-drive" of PARCC has concluded, and the Board of Elementary and Secondary Education faces an important choice with respect to the statewide assessment for ELA and math: Should Massachusetts continue with the Massachusetts Comprehensive Assessment System (MCAS), adopt the Partnership for Assessment of Readiness for College and Careers (PARCC) as the Commonwealth's statewide assessment system, or consider another option?

This decision will have long-lasting implications for the direction of K-12 education in the Commonwealth and is particularly difficult in light of the large number of uncertainties concerning both assessments. These uncertainties include both assessments' impact on long-term outcomes and their cost. Uncertainties exist for PARCC, a new and still evolving assessment that by definition lacks long-term outcomes, and for MCAS, an 18-year old assessment that would need renewed attention and improvements. Despite the state's long history with MCAS, there is a striking paucity of research showing the relationship between MCAS and long-term outcomes such as college completion or labor market success.

As college and career readiness have become the focus of K-12 education, the lack of preparedness of high school graduates for the demands of college and the workplace has become a concern across the country and in Massachusetts. In 2012, 35 percent of students who were enrolled in a public college in Massachusetts took at least one remedial course. Because the 10th grade MCAS was designed as an assessment of students' level of mastery of the knowledge and skills in relationship to the Massachusetts Curriculum Frameworks, a different criterion than that of college and career readiness, questions have been raised about whether it is well-suited to meet the Commonwealth's needs of today.

According to a recent study by Mathematica Policy Research, both MCAS and PARCC predict college readiness, as measured by first-year college grades. Both assessments are comparable to SAT scores in predicting first-year college outcomes. Furthermore, both MCAS and PARCC scores provide similarly strong predictions about which students need remedial coursework in college. These research findings are consistent with Professor Andrew Ho's analysis of grade 10 MCAS performance levels and remediation in college, presented in this report. He finds that a student's MCAS performance level is a strong predictor of remediation. Students who score higher on grade 10 MCAS are less likely to need remediation in college.

While MCAS and PARCC scores appear to be equally predictive of college remediation and first-year college grades, there are still important differences between the assessment systems. First, differences in performance standards suggest that the current signaling of MCAS regarding college readiness is not as useful as PARCC's. In the case of PARCC, the five performance standards are set to signal whether a student is on track to meet the end goal of satisfactory

performance in the first year of college. In contrast, the four performance levels within MCAS are set in relation to the Curriculum Frameworks for each grade. In addition, because PARCC's performance standards are aligned across grades, it can signal whether students in lower grades are "on track" to meet the end goal. This is not currently the case with MCAS. The performance standards of MCAS, however, could be reset and aligned to enable better signaling, although this would have some implications for the ability to compare recent scores with scores from previous years. In addition, the report discusses other related issues that are important for the Board to consider regarding the appropriateness of the current grade 10 MCAS test as the state's graduation requirement.

Beyond performance standards, there are other potential consequential differences between the MCAS and PARCC assessment systems. Both MCAS and PARCC can be considered high-quality assessments with respect to issues of validity and reliability, and they both can be aligned to the Common Core content standards. Yet, because no realistic assessment can cover all the state's content standards, design choices determine which standards are assessed and at what depth. Choices about item types combined with differences in assessment systems, such as policies around item release and score reporting, create different incentives for educators and students. The design of the assessment system should support the educational goals of the system as a whole. The needs of the multiple audiences – including educators, parents, policy-makers, and the broader public – may differ and require careful balancing and trade-offs.

Academic research finds that the types of items on an assessment influence classroom instruction. Test developers must consider a number trade-offs, including balancing the amount of time taken away from instruction for testing with the need to fully represent the content standards in a way that reflects their relative importance. For instance, multiple-choice questions might allow for more breadth of coverage of the content standards but could also create incentives for teachers to place a lower priority on the deeper learning skills that are not typically assessed through multiple-choice questions. An apples-to-apples comparison of MCAS and PARCC's items is difficult, because they use different types of items and weigh them differently within the scoring. To the extent that MCAS contains more multiple-choice items, it is likely more focused on measuring breadth of knowledge, as compared with depth of knowledge, although MCAS also includes a long composition in several grades.   The PARCC Consortium spent a great deal of time and effort, aided by substantial financial support from the federal government, in developing its item types in an effort to assess students on the full-range of cognitive complexity and to create incentives for classroom instruction to focus on deeper thinking skills. PARCC appears to rely less on multiple-choice questions, indicating that it might focus more on assessing students' depth of knowledge.

In addition to item types, the release of items and types of reports also have implications for classroom instruction as well as for transparency to families, policy-makers, and the broader public. The quality and transparency of an assessment

system can contribute to the confidence in and the legitimacy of the standards-based educational system in the view of the broader public. There has been a lot of transparency around MCAS that has been facilitated by the history of releasing test items and the availability of different types of reports; there may be some limitations to allowing the same amount of transparency with PARCC, particularly as it pertains to the cost implications of releasing test items and reporting options.

There are clear differences between MCAS and PARCC with regard to the issue of governance. MCAS is a custom-designed assessment, which the Commonwealth owns in its entirety. In contrast, PARCC is a multi-state consortium, where Massachusetts serves in a leadership role on the Governing Committee but is still only one member of the consortium. Being a member of a consortium offers certain opportunities and advantages for Massachusetts. In addition to potential cost efficiencies, the existence of a common assessment across multiple states could help create a larger market for aligned instructional materials that could help elevate the quality of such products and reduce their average cost.

At the same time, there are risks and uncertainty that result from being part of a consortium. The membership of the PARCC Consortium is still in flux, and the loss of other member states reduces some of the anticipated advantages of scale and comparability of student achievement across states. In addition, Massachusetts must compromise with the other members to reach mutually agreed-upon choices. This includes decisions on item development, item release, report types, accommodation rules, and many other aspects of the system. Being a member of a consortium also has implications for future changes to the Curriculum Frameworks, Massachusetts's content standards. By their nature, standards evolve over time to reflect changing expectations for students, and as standards evolve, so too must assessments. If the Board selects PARCC, the stakes for any future modifications of content standards, particularly at the state level, become much higher.

While the focus is currently on the grade 3-8 assessments, the Board must consider and anticipate the consequences of future choices for high school assessments.   This is particularly relevant with regard to the 10th grade Competency Determination (CD) required by the Education Reform Act of 1993. If the Board chooses PARCC, it must subsequently decide what the high school graduation requirement will be for students after the class of 2019. The Board will have to define an appropriate requirement for high school graduation within an assessment system focused on college readiness. Because of the two different high school math tracks in PARCC - the integrated math track and the Algebra I, Geometry, Algebra II track - the Board will need to consider how to set the graduation requirements so as not to create unintended incentives and in a manner that maintains public confidence in the fairness and equity of the standard for all students.

There is not a simple answer to the question of "MCAS or PARCC?" Rather, the answer requires a balancing of priorities and judgment by the Board regarding the trade-offs, including weighing the significant uncertainties and potential risks. This

report aims to frame some of the key considerations for the Board based on the best research publicly available and information provided by the Department of Elementary and Secondary Education, informed by the Advisory Group to the Secretary of Education. The report is organized as follows: after an introduction (Section I), the report briefly reviews the context for the PARCC "test drive" (Section II). Section III discusses the purposes and quality of assessments in a standards-based education system. Sections IV and V offer a descriptive overview and comparison of various attributes of the MCAS and PARCC systems. The final section, "Other Policy Considerations," underscores the importance of considering issues related to governance and cost, although these issues are not the focus of this report. It is not the intent of this report to provide a recommendation to the Board but rather to help inform its decision.

I. Introduction and Purpose of the Report

In November 2013, the Board of Elementary and Secondary Education decided to allow Massachusetts educators and students the chance to "test drive" the Partnership for Assessment of Readiness for College and Careers (PARCC) assessment before making a decision about whether or not to adopt it statewide. The transition time was intended to give the state the opportunity for a robust comparison of the Massachusetts Comprehensive Assessment System (MCAS), the Commonwealth's longstanding statewide assessment, and PARCC, a new assessment that Massachusetts helped to develop.

As we enter fall of 2015, the "test-drive" of PARCC has concluded, and the Board faces an important choice: Should Massachusetts adopt PARCC as the Commonwealth's statewide assessment system for ELA and math or embark upon another course to ensure that the state's assessment system continues to advance the state's educational goals.

The choice requires considering the characteristics of the existing MCAS assessment system and the present and intended elements of the PARCC system. The Board's deliberations must also allow for and consider the implications of existing uncertainties inherent to each of the paths forward. The Board is being asked to compare a newly developed assessment system that is still evolving with an assessment system that has an 18-year track record but likely needs renewed attention and modifications in order to continue serving the needs of the Commonwealth well going forward.

As would be expected, there are both advantages and disadvantages associated with each assessment system. Throughout the five public hearings held in the spring and summer of 2015, educators, parents, students, and other interested parties articulated the pluses and minuses based on their different perspectives. There was no consensus; there is no simple answer. Rather, coming to an informed and appropriate decision will require balancing different priorities and considering the costs and implications of the alternatives.

The purpose of this report is to bring forward credible research-based information to help guide the Board's evaluation of MCAS and PARCC. Whenever possible, the report relies upon MCAS- and PARCC-specific research, and also includes more general research on a given topic, as relevant. There are, however, obvious gaps in knowledge. While the design of PARCC has relied extensively on research, 2015 was its first administration at scale. This report can speak to the intentions of PARCC, but there is currently little, or no, publicly available information about student performance or longer-term outcomes. It will be at least several years before a cohort of PARCC students graduate from high school and enroll in post-secondary education or embark upon other post-secondary endeavors. In addition, as a newly created assessment, PARCC is still in development and is evolving in response to knowledge gained from its recent administrations and from continuing efforts to

appropriately balance the various goals of any assessment system. For example, in 2016, the PARCC assessments will be shorter in length than in 2015. The membership of the assessment consortium is also still in flux with the current members including 7 states and the District of Columbia.[1] The evolving nature of PARCC presents a particularly challenging aspect of the Board's deliberations, as it will have to evaluate the risks and opportunities that arise from these uncertainties.

This report aims to frame some of the key considerations for the Board based on the best research publicly available, by information provided by the Department of Elementary and Secondary Education (ESE), and informed by experts in educational policy, assessment, and evaluation who were members of the report's Advisory Group.[2] The report is organized as follows: following this introduction (section I), the report begins by briefly reviewing the context for the PARCC "test drive" (section II), including an overview of the Massachusetts Curriculum Frameworks adopted in 2010 and the general role of assessments and accountability in Massachusetts's standards-based education system.   Section III discusses the purposes of assessments in a standards-based education system and identifies some core elements of a quality assessment system. Section IV offers a brief descriptive overview of MCAS and PARCC, and Section V compares various attributes of the MCAS and PARCC systems.   The final section, "Other Policy Considerations," underscores the importance of considering issues related to governance and cost, although these issues are not the focus of this report. It is not the intent of this report to provide a recommendation to the Board but rather to help inform its decision.

II. The Context for the PARCC "Test Drive"

Twenty-two years ago, with the passage of the 1993 Massachusetts Education Reform Act (MERA), Massachusetts made a bold commitment to all of its students. In exchange for additional state funding, the state would hold districts and schools accountable for high levels of student achievement. During the same period that Massachusetts more than doubled its investment in local aid to schools, it also created and adopted high standards and required more accountability for meeting those standards throughout the education system. As years of steady educational progress followed, Massachusetts's standards-based system became a national model of rigor and quality.

These high standards, quality assessments, and accountability provisions have established Massachusetts's reputation as a national leader in education.

---

[1] As of October 2015, PARCC members included: Colorado, Illinois, Maryland, Massachusetts, New Jersey, New Mexico, Rhode Island, and the District of Columbia.
[2] The Advisory Group to Secretary Peyser for the development of this report includes (in alphabetical order): Henry Braun (Boston College), Roland Fryer (Harvard University), Ronald Hambleton (University of Massachusetts, Amherst), Andrew Ho (Harvard University), Tom Kane (Harvard University), Kevin Lang (Boston University), and Martin West (Harvard University).

Massachusetts students have consistently earned top scores on national and international tests. The National Assessment of Educational Progress (NAEP) test, which is taken by a representative sample of students in grades 4, 8, and 12, is often referred to as the nation's report card. Because NAEP is the largest nationally representative assessment that has been administered continuously, it is often the measuring stick for comparisons of student achievement across states. In 2017, NAEP will begin administering digitally based assessments for mathematics, reading, and writing.[3]

On NAEP, Massachusetts students have earned the top scores in the country since 2005. In 2013, Massachusetts ranked first alone among the states in 8th grade mathematics, and tied for first in 4th grade reading, 4th grade mathematics, and 8th grade reading. Despite these top scores, recent trends have not been universally positive. The performance of students on three of the four tests has remained statistically the same since 2011, the previous administration of NAEP. In grade 4 reading, the average score for Massachusetts students in 2013 declined by five points, compared with results of the previous NAEP test, which took place two years earlier. Massachusetts was one of only three states in the nation to see a statistically significant decline in grade 4 reading scores between 2011 and 2013.[4]

In thinking about a standards-based system, the content standards set for students are, in many respects, the starting point. Content standards describe learning objectives for students. They set expectations for what a student should know and be able to do at the end of a grade or a course. Since 1993, the Massachusetts ELA and math standards – called the Massachusetts Curriculum Frameworks - have been updated or modified several times to reflect changes in expectations for students. They were significantly revised in 2000 (math) and 2001 (ELA).

In 2007, the Massachusetts Department of Elementary and Secondary Education convened a group of educators to review and revise the state's standards. Shortly thereafter, a similar process occurred at the national level, led by the National Governors Association (NGA) and the Council of Chief State School Officers (CCSSO). Efforts to establish common standards across the nation occurred within the context of the federal Race to the Top initiative that promoted adoption of the new standards and participation in consortia to develop new assessments aligned to the new standards. Over time, the focus and energy of the Massachusetts team shifted to the multi-state project that put forward the Common Core state standards, which were initially adopted by 46 states, including Massachusetts, and the District of Columbia. Over time some of the states that adopted the Common Core standards

---

[3] National Center for Education Statistics. (n.d.) *NAEP Overview*. Retrieved from http://nces.ed.gov/nationsreportcard/dba/

[4] Massachusetts Department of Elementary and Secondary Education. (2013). Massachusetts 4th and 8th Graders Lead the Nation in Reading and Mathematics Performance for the Fifth Consecutive Time [Press release]. Retrieved from: http://www.doe.mass.edu/news/news.aspx?id=7846

have voted to repeal or replace them, but 43 states and the District of Columbia continue to use them.[5]

College and career readiness is the focus of the Common Core standards. The shift to college and career readiness has been a response to the changing expectations for students and the increasing demand by employers for a higher level of skills. In a report published by the Stanford Center for Opportunity Policy in Education (SCOPE), the authors summarized the impetus behind the changes as follows: "The changing nature of work and society means that the premium in today's world is not merely on students' acquiring information, but on their ability to analyze, synthesize, and apply what they've learned to address new problems, design solutions, collaborate effectively, and communicate persuasively."[6] As the demands of the workplace and the world have changed, the new standards are intended to define learning expectations for students that will enable them to be successful in today's society and workplace.

> **PARCC Performance Standards**
>
> PARCC uses five performance levels that describe the knowledge, skills, and practices students are able to demonstrate:
>
> - Level 1: Did not yet meet expectations
> - Level 2: Partially met expectations
> - Level 3: Approached expectations
> - Level 4: Met expectations
> - Level 5: Exceeded expectations

In 2013, in Massachusetts, both the Board of Elementary and Secondary Education and the Board of Higher Education formally agreed - for the first time - upon a definition of college and career readiness. That definition describes Massachusetts students who are college and career ready as students who "will demonstrate the knowledge, skills and abilities that are necessary to successfully complete entry-level, credit-bearing college courses, participate in certificate or workplace training programs, and enter economically viable career pathways."[7] This definition is notable for its focus on the completion of entry-level, credit-bearing courses and not simply the entry into such courses.

The Common Core standards and the PARCC assessments aspire to promote and measure college and career readiness. The definition of college and career readiness has changed over time, and even today, it can be used to refer to different measures of success. The choice of performance standards for the PARCC – the descriptions of what students at different performance levels know and can do –

[5] Common Core or Something Else? A Map of State Academic Standards (2015, July 20). *Education Week*, Vol. 34, Issue 36. Retrieved from: http://www.edweek.org/ew/section/multimedia/map-states-academic-standards-common-core-or.html

[6] Darling-Hammond, L., Herman, J., Pellegrino, J., *et al*. (2013). *Criteria for high-quality assessment*. Stanford, CA: Stanford Center for Opportunity Policy in Education. Retrieved from: https://edpolicy.stanford.edu/sites/default/files/publications/criteria-higher-quality-assessment_2.pdf

[7] Massachusetts Department of Elementary and Secondary Education and Massachusetts Board of Higher Education. (2015). *Massachusetts Definition of College and Career Readiness*. Retrieved from: http://www.mass.edu/library/documents/2103College&CareerReadinessDefinition.pdf

were informed by 1) the content standards; 2) comparisons with definitions of college readiness and performance levels on other national and international assessments; and 3) expert judgment. They are intended to signal whether students' performance indicates that they are "on track" to meet the standard for satisfactory performance in the first year of college. The PARCC definition of college ready is that a student who achieves Level 4 or higher (on a 5-level scale) will have a 75% chance of earning at least a "C" grade in an entry-level, credit-bearing college course. In setting these college-relevant standards, the assessment should incent some changes in curriculum and instruction by educators.

2010 Massachusetts Curriculum Frameworks
The 2010 Massachusetts Curriculum Frameworks incorporate the Common Core State Standards and also include some additional standards, above and beyond the Common Core standards.[8] These additional standards are unique to Massachusetts. In both ELA and math, the Massachusetts Curriculum Frameworks include standards for pre-kindergartners, which are not part of the Common Core. In ELA, the Massachusetts Curriculum Frameworks expand the Common Core's glossary and bibliography and also suggest appropriate classic and contemporary authors for different grade-level ranges and include other additional standards. In math, in addition to the pre-K standards, Massachusetts added 19 math standards to the K-12 Common Core standards.[9]

A specific example of the unique Massachusetts standards is a standard for both preK-5 students and students in grades 6-12 within the anchor writing standards. The Massachusetts Standard (MA.3.A) is: "Write fiction, personal reflections, poetry, and scripts that demonstrate awareness of literary concepts and genres."[10] Thus, while the Massachusetts Curriculum Frameworks for English Language Arts and Literacy are primarily based on the Common Core standards, they are not identical.

Similarly, the Massachusetts Math Curriculum Frameworks include additional standards. For example, within Operations and Algebraic Thinking, Massachusetts first graders are expected to: "Write and solve number sentences from problem situations that express relationships involving addition and subtraction within 20." Within Measurement and Data, Massachusetts first graders are also expected to work with money in order to be able to: "Identify the values of all U.S. coins and know their comparative values (e.g., a dime is of greater value than a nickel). Find

---

[8] States were allowed to customize 15% of the standards.
[9] Massachusetts Department of Elementary and Secondary Education. (n.d.)*10 Frequently Asked Questions about the Common Core State Standards and PARCC*. Retrieved (October 5, 2015) from: http://www.does.mass.edu/parcc/
[10] Massachusetts Department of Elementary and Secondary Education. (2011). *Massachusetts Curriculum Frameworks for English Language Arts and Literacy: Grades Pre-Kindergarten to 12: Incorporating the Common Core State Standards for English Language Arts and Literacy in History/Social Studies, Science, and Technical Studies*. Retrieved from: http://www.doe.mass.edu/frameworks/ela/0311.pdf

equivalent values (e.g., a nickel is equivalent to 5 pennies). Use appropriate notation (e.g., 69¢). Use the values of coins in the solutions of problems." There are also additional Massachusetts math standards in fifth grade, seventh grade and in high school.[11] These standards can be assessed within MCAS but would not be tested in PARCC.

Although content standards do not directly determine curricula or prescribe particular instructional practices and the use of specific educational materials, the choice of standards and assessments aligned to them does influence the adoption of curriculum and instructional practice. Educators decide which curriculum to use and how to teach students so that they can best master the standards, with attention to how the standards will be assessed. Academic research, discussed in the sidebar "The Influence of the Format of Assessments on Instruction" on page 22, shows how the choice of an assessment system's item formats (e.g., multiple choice vs. performance-based) can influence classroom practice. Accountability provisions that rest heavily on student assessment results, including the publication of school and district achievement results, also create powerful incentives for educators to align curriculum and instruction closely with the standards and assessment system.

Assessments and Accountability
Assessments are a key element of a standards-based education system and were a central part of the Massachusetts Education Reform Act. Assessments measure students' learning progress with respect to the content standards and create incentives for classroom instruction. They also send important signals to parents about their children's level of achievement. It is noteworthy that two different assessments can both be high quality and both be aligned to content standards but still have consequential differences. No realistic assessment can cover all the state's content standards, so design choices determine which standards are assessed and to what depth. These choices lead to differences in the assessments that create different incentives for classroom instruction and practice.

State law created by MERA requires that the state's testing program must:

- Test all students who are educated with Massachusetts public funds, including students with disabilities and limited English proficient students;

- Measure students' knowledge of the material contained in the Massachusetts Curriculum Frameworks; and

- Report on the performance of individual students, schools, and districts.

---

[11] Massachusetts Department of Elementary and Secondary Education. (2011). *Massachusetts Curriculum Frameworks for Mathematics, Grades Pre-Kindergarten to 12: Incorporating the Common Core State Standards for Mathematics*. Retrieved from: http://www.doe.mass.edu/frameworks/math/0311.pdf

The 10th Grade Competency Determination

The Education Reform Act of 1993 established a new state standard called a competency determination or CD, which went into effect in 2001 for the graduating class of 2003. In addition to any local requirements, the CD required students to meet the 10th grade standards adopted by the Board through regulation and measured by the statewide student assessment (MCAS) in order to earn a high school diploma. Today, students must pass 10th grade MCAS tests in ELA and math and one of the high school science and technology/engineering (STE) tests. (More information about the competency determination can be found on page 45.)

Under current Board policy, the graduation requirements will remain in place for current students through the class of 2019. Although a decision on the competency determination is not required at this time, the Board must consider and anticipate the consequences of the choices it makes now on the future options that will be available.

As discussed in Section III, assessments serve a variety of policy objectives at the same time. In order to successfully serve each purpose, however, policy-makers must be clear about the objectives of the assessment, balancing the demands of breadth, depth, available time, and resources to effectively meet their goals. Moreover, each purpose should be assessed for its validity.

As a member of a consortium, the Commonwealth is more limited in its ability to alter the design of the assessment system to meet Massachusetts-specific needs and purposes. If the Board chooses to remain in PARCC, they will have to establish an appropriate requirement for high school graduation within an assessment system much more focused on college readiness than the MCAS system it may replace. Not all PARCC members will use the PARCC assessment to establish a high school graduation requirement. Because of this, it is unlikely that the high school tests have been or will be developed by the consortium in a way that focuses their precision on items and standards most appropriate for use as a CD. This question of how a high school competency determination standard fits into an overall college-and-career-readiness system will require careful attention by the Board as it moves forward.

If the Board chooses to use PARCC and does not choose to develop or maintain a separate and distinct assessment to determine the CD, it will need to decide: which forms and scores will be relevant for federal reporting purposes, high school graduation purposes, and, along with their colleagues on the Board of Higher Education, for college placement and remediation purposes. These may or may not be based on the same scores or the same tests. If they are different scores, a clear communication outreach effort will be necessary to avoid sending potentially confusing signals to families about their children's level of achievement and preparedness for their choices after high school. In addition, in PARCC, there are two different high school math tracks – the integrated math track and the Algebra I, Geometry, Algebra II track. The Department's early consideration of this matter has explored the possibility of using combinations of PARCC assessments. The Board will need to consider how to set equivalent graduation requirements to ensure that one track is not more difficult than another so as not to create unintended incentives for students. Creating different options to fulfill the high school requirement could open new opportunities and ways of thinking about the high school credential, but it could also raise difficult questions about the consistency of the high school graduation standard across the various options. Given the high-stakes nature of the high school tests, any questions about consistency and equity could diminish public confidence in the standard itself. Technical challenges around item release under the current PARCC design could also limit transparency and potentially compound this threat.

In the face of these complexities and unknowns, it may prove tempting to opt for the existing MCAS test. Yet, for a variety of reasons, the existing MCAS 10th grade test does not offer the Board an easy alternative. There are a number of concerns about the 10th grade test as it currently exists.

As discussed throughout this report, the state revised its ELA and math standards in 2010 to incorporate the Common Core standards plus some additional standards unique to Massachusetts. At that time, partly due to fairness concerns over limited exposure to the new standards by current high school students at the time, the

Department and Board decided to treat the transition of the 10th grade assessment differently from that in the lower grades. In grades 3-8, nearly all of the 2010 Curriculum Frameworks are assessed, according to the Department of Elementary and Secondary Education. In contrast, the grade 10 assessments are based on a narrower set of standards. The current grade 10 MCAS test only assesses standards that are both part of the 2010 standards and the 2000/2001 standards (the intersection of the two standards); the 10th grade MCAS does not assess students on all of the Curriculum Frameworks. The "test drive" of PARCC likely delayed the full transition of the 10th grade test as limited resources were stretched to maintain two assessment systems simultaneously. However, if allowed to continue, this narrowed focus of the 10th grade test would be expected to incent changes in curriculum and instructional practice in ways that the Board is not likely to regard as positive, if it has not already begun to do so.

Beyond the narrowing of the standards assessed, the 10th grade test has also experienced technical challenges due to the utilization of different equating practices to ensure year-to-year consistency in tested materials. This challenge is sometimes referred to as "drift" or "inflation." Although the Department has recently addressed these technical challenges, it is still the case that the current results of the 10th grade MCAS test likely overstate the progress made by students over the past decade. For more information about this issue see "Trends in Grade 10 MCAS Proficiency and College Remediation" on page 31.

If the Board chooses to continue using MCAS as the statewide assessment, it should pay particular attention to improving and updating the 10th grade test since it is likely that the current assessment may be sending inaccurate signals regarding student achievement. If the Board chooses to remain with the MCAS, there is much work to do to put the state's competency determination (CD), in particular, on a firm, transparent, and educationally appropriate footing as we move forward.

Massachusetts developed the Massachusetts Comprehensive Assessment System (MCAS) as its annual summative assessment pursuant to the requirements of MERA. MCAS is intended to measure student knowledge and skills near the conclusion of a school year. Massachusetts has administered MCAS tests since 1998. As content standards have changed, the assessments have occasionally changed in order to remain relevant and appropriate. There have also been new statewide assessments, such as science and technology/engineering, introduced during that time to assess how students are meeting other standards established by the state.

Historically, most states, including Massachusetts, have used their own assessments as they have moved towards a standards-based system, although there are limited examples of states working together, such as the New England Comprehensive Assessment Program (NECAP).[12] As part of the Race to the Top assessment program, the U.S. Department of Education awarded grants to help support two different multi-state consortia in developing a common set of K-12 assessments aligned to the Common Core. Massachusetts chose to participate in the Partnership for the Assessment of Readiness for College and Career (PARCC) consortium and has played a leadership role in the design and development of the PARCC assessment.[13]

---

[12] Before the Common Core state standards were adopted, Maine, New Hampshire, Rhode Island, and Vermont worked together to develop common grade-level expectations and test specifications. These states have transitioned from the NECAP to different assessments for the Common Core State Standards, but the NECAP science assessment still exists.

[13] The second consortium is the Smarter Balanced Assessment Consortium (SBAC). There are also two consortia for alternative assessments for special needs students who, even with

Given the state's long history with and investment in MCAS, the Board decided to approve a "test drive" of PARCC before making a decision in the fall of 2015 regarding which statewide assessment system to adopt for ELA and math. In 2014-15, Massachusetts offered districts a choice about whether to administer PARCC or MCAS in grades 3-8 with a commitment to "hold harmless" the accountability level of districts that chose to administer PARCC, meaning that a district's accountability level could only improve or remain the same based on student performance on PARCC; their accountability level could not decline. Schools administering MCAS would be classified into an accountability level as usual, and would not be held harmless.[14] In most cases, a district's decision applied to every school in the district, but in Boston, Springfield and Worcester, individual schools administered one or the other assessment, within the district.

In spring 2015, approximately 227,000 Massachusetts students (53%) in grades 3-8 took the PARCC tests; the remaining students took the MCAS tests. An additional 22,500 students (approximately 15% of the students enrolled in these grades) took the optional PARCC tests in 9th and/or 11th grades (Table 1). All 10th grade students had to take the MCAS test.

Table 1: Choice of Assessment, Spring 2015

Spring 2015	Number of public districts (including charters)	MCAS			PARCC		
		No. of districts	%	Approx. no. of students (percent)	No. of districts	%	Approx. no. of students (percent)
Grades 3-8	359	165	46%	207,500 (47%)	194	54%	227,000 (53%)
PARCC for Grade 9 and/or 11 (optional)	295	n/a	n/a	n/a	69	23%	22,500

Source: Massachusetts Department of Elementary and Secondary Education, retrieved from: http://www.doe.mass.edu/news/news.aspx?id=13541

With the conclusion of the 2014-15 school year, Massachusetts has completed its test-drive of PARCC, and the Board of Elementary and Secondary Education faces critically important choices regarding which statewide assessment systems for ELA

accommodations, are not able to take the standard assessment. Massachusetts chose not to participate in either of the alternative assessment consortia and instead opted to continue with MCAS-Alt, the state's existing alternative assessment system.

[14] According to *Education Week's* analysis, Massachusetts was the only state that did not decide upon a single assessment for the academic year of 2014-15. Gewertz, C. (2015, February 4). A Map of States' 2015 Testing Plans: The Dust Has Finally Settled. *Education Week Curriculum Matters blog.* Retrieved October 5, 2015 from: http://blogs.edweek.org/edweek/curriculum/2015/02/a_map_of_states_2015_testing_p.html

and math to use in Massachusetts in future years. Massachusetts will continue to need to maintain a separate assessment system, and associated contracts with separate vendor(s), for the science and technology/engineering (STE) assessments, which are administered in 5th and 8th grades and in high school. Students must pass the science and technology/engineering assessment as part of the state's high school graduation requirement. There are also continuing discussions about the possibility of adopting a statewide history test. In addition, Massachusetts will continue to need an alternative assessment for students who, even with accommodations, are not able to take the standard assessment.[15] The Board has also voted to continue using the 10th grade MCAS tests in ELA and math to determine the competency determination through the class of 2019. The November 2015 decision before the Board, then, concerns only the statewide ELA and math assessments in grades 3-8.

III. The Purpose and Quality of Assessments

It is common for state level, standardized tests to serve a variety of purposes. Resource constraints coupled with concerns about too much time spent on testing have led to a shared desire to minimize the amount of time devoted to testing. As a consequence, increasingly, a single test is used to meet a number of different goals. When a single test is used for multiple purposes, it is necessary to balance competing objectives.

In a presentation to a workshop on Best Practices for State Assessment Systems organized by the National Research Council in December 2009, Margaret Goertz categorized a number of common purposes of assessments (Table 2).[16] A test can be used to serve diagnostic and evaluative purposes. It can be intended to inform teaching and learning while also being used for accountability purposes. Assessments send signals to parents about their children's academic achievement and help create transparency around the education system. The same test might be used to evaluate the performance of students, teachers, and schools. The differences in purpose have a number of consequences for test design.

---

[15] Instead of participating in either of the multi-state consortia for alternative assessments, Massachusetts chose to continue with MCAS-Alt as its alternative assessment. According to *the 2013 MCAS Technical Report*, "a total of 9,111 students, or 1.7% of the assessed population, participated in the 2013 MCAS-Alt in grades 3–10. A slightly higher relative proportion of students in grades 3–8 took the MCAS-Alt compared with students in grade 10, and slightly more students were alternately assessed in mathematics than in ELA"(p.86).
[16] Beatty, A. (2010). *Best Practices for State Assessment Systems, Part 1: Summary of a Workshop*, Washington, D.C.: National Academy of Sciences.

Table 2: Purposes of Assessment

Use	Student	Teacher	School
Diagnostic	Instructional decisions, placement allocation of educational services	Professional development and support	Resource allocation, technical assistance
Inform Teaching and Learning		Focus, align, redirect content; instructional strategies	Instructional focus align curriculum to skills or content; school improvement and planning
Evaluation	Certification of individual achievement	Teacher preparation programs, teacher pay	Program evaluation
Public Reporting	Transcripts		Parent or community action
External Accountability	Promotion, high school graduation	Renewal, tenure, pay	Sanctions and rewards

Source: Goertz (2009) as cited in *Best Practices for State Assessment Systems, Part 1: Summary of a Workshop*.

The purpose of an assessment should drive the design of the test, and different purposes often have different design implications. Consider two central purposes of assessments: informing teaching and learning and providing external accountability to an educational system. To inform teaching and learning, ideally, the assessment should provide educators with timely information throughout the school year, with sufficient coverage of each standard of interest, in a format that would allow them to easily analyze their students' performance and make corresponding modifications to their classroom instruction. Timeliness is a key attribute for this goal. In addition, the content and formats of score reports available are also important to enable educators to easily use the information. In contrast, when assessments are used as part of an accountability system, such as high school graduation requirement or teacher evaluation system there must be an emphasis on the fairness, validity, and reliability of the scores for these purposes. These attributes typically require more analytic time to obtain results.[17] To effectively use a single test for multiple purposes requires deliberate consideration of each intended use and a balancing of priorities.

---

[17] Beatty (2010), p. 8.

The Influence of the Format of Assessments on Instruction

Research suggests that testing influences what and how students are taught. In a 2013 literature review "New Assessments, Better Instruction? Designing Assessment Systems to Promote Instructional Improvement," by Susannah Faxon-Mills, Laura Hamilton, Mollie Rudnick, and Brian Stetcher of RAND, the authors review research that has examined how instructional practices change in response to assessments. They caution that these studies are often qualitative with small sample sizes; however, they provide a useful framework for intended and unintended consequences of changing assessment programs and switching to assessments with different item formats. The information below is drawn from their report.

Researchers find that the format of the test and the relative emphasis of items within a test will likely impact instruction. Teachers change their instruction to emphasize the kinds of skills measured by the test, and teachers may even target their instruction based on differences in item format within the same test. The research finds that multiple-choice tests are often accompanied with an increased emphasis on basic skills, while performance-based assessments are associated with a greater focus on problem solving and inquiry. One of the strategies in newer assessments is to use performance-based assessments to help shift toward deeper thinking and problem solving. Performance-based assessments vary in their formats; they can include short answers, longer composition, multiple-select, or other formats but the core element is that the student is not constrained by a pre-determined set of options (e.g. multiple choice).

Research shows how performance assessments can influence classroom instruction and practices. In the 1990s, Vermont adopted a statewide portfolio assessment system, and in response, teachers reported increasing their emphasis on problem solving in math. Similarly, when math assessments require students to explain their answers, math teachers report increased emphasis on explanation in their class. And, if the ELA assessment includes writing, then teachers spend more time in class on writing. With performance assessments, some research is cautionary, suggesting that the effects on classroom practice are not always as significant as desired.

Instructional strategy is another choice that teachers make as they decide how to present the material – through lectures, discussions, small groups, or other methods. Overall, the research identifies many instances of teachers using more traditional teaching practices, such as lecturing, in response to high-stakes testing. The instructional activities tend to be highly structured and emphasize transmission of content. The ultimate implications of these practices is not clear, because there is not definitive evidence showing that one instructional approach is more effective than others in all contexts. Moreover, these findings may be related to the format of the assessments. The authors find that while "teachers may respond to high-stakes multiple choice testing by relying on traditional or teacher-centered instructional practices, the literature around performance assessments reveals a potential shift in a different direction." The evidence suggests that teachers might respond to performance assessments with "an expanded repertoire of teaching strategies and techniques."

As with all assessments, performance-based assessments involve a number of trade-offs. Performance-based assessments typically take more time to complete, potentially leading to or adding to existing concerns about too much time spent on testing. They also might focus more on assessing depth, which can include trade-offs regarding the breadth of assessment.

Although research finds that testing influences classroom practice, it is also the case that changes in test format are not sufficient on their own to enable teachers to change their practice. Other supports must also be offered to change instructional practice. These factors include both the purpose and quality of the test itself plus the context surrounding the test, including the accountability context and district and school policies. Moreover, tests that are explicitly intended to influence classroom instruction, such as classroom-based formative assessments, may be more likely than other types of assessments to influence instructional practices. In addition, the timeliness of the score reporting is central to its instructional impact.

As decisions about assessments are made, it is important to remember that the format of the tests signals educators and other stakeholders what kinds of skills should be emphasized in the classroom.

Another key element of a high quality assessment system designed to serve multiple purposes is the collection and evaluation of validity evidence for each use of a test. The *Standards for Educational and Psychological Testing*, a collaboration by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME), provide widely agreed-upon criteria for the development and evaluation of tests and testing practices as well as guidelines for assessing the validity of interpretations of test scores for the intended test uses. As the AERA/APA/NCME *Standards* explain, "It is incorrect to use the unqualified phrase 'the validity of the test.' Evidence of the given validity of a given interpretation of test scores for a specified use is a necessary condition for the justifiable use of the test."[18] For each purpose of the test, it is necessary to make certain that the test accurately measures what it is intended to measure. To the extent that the tests are high-stakes, these considerations become even more important.

Looking at the big picture, Margaret Goertz poses the following four questions to help guide planning and choices about assessments:

- What do we want to test and for what purpose?
- What kinds of information do we want to generate and for whom?
- What is the role of a state test in a comprehensive assessment system?
- What supports will educators need?[19]

Answering these questions could help delineate the priorities for the statewide assessment. As has been discussed, tests, and particularly those that are considered to be high-stakes, create powerful incentives for schools, teachers, and students.   It is important to consider what the state wants to measure and how can it best create incentives consistent with those goals. The design of the assessment and all the related aspects, such as score reporting, as well as accountability policies must be aligned to support the educational goals of the system as a whole.

The Quality of the Assessment

While there are many elements involved in creating a high-quality assessment, the *Standards* describe three core, foundational characteristics by which the quality of an assessment is judged: 1) Validity; 2) Reliability; and 3) Fairness.

While fully explicating the process of developing and administering a quality assessment is beyond the scope of this report, our purpose here is to briefly highlight some of the most critical aspects of quality assessments to help frame comparisons between MCAS and PARCC. As noted above, validity is at the core of a

---

[18] American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education (AERA/APA/NCME). (2014). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association, p. 11.
[19] Beatty (2010), p. 11.

quality assessment. According to the *Standards*, "Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests."[20] Validity must be established for each purpose of an assessment and also depends on technical aspects of the assessment, such as appropriate test administration and scoring and accurate score scaling, equating, and standard setting. To the extent that a test is considered high-stakes, the standards of evidence of validity should be higher.

To give an example – there can be the desire to use a single test to both assess a student's current level of knowledge and skills in high school and also to predict a student's performance in college as a measure of college readiness. Each of these purposes must be validated. Does the test accurately assess a student's knowledge and skills at the time of the test, and does that test also accurately predict a student's performance in college at a later point in time? Both can be true but each purpose must be considered and validated.

Reliability - often considered the second fundamental element of an assessment - refers to the "consistency of scores across replications of a testing procedure, regardless of how this consistency is estimated or reported."[21] Put simply, a test is reliable if it produces similar results for any given student with the same level of understanding of the material. There are different types of reliability to be considered. For instance, if there are multiple forms of a test, the forms should be parallel in content, have similar distributions of item difficulties, and equivalent psychometric properties. A student's score should not depend on which form of the test he or she takes. There are many aspects of reliability that must be considered during the development and administration of an assessment. Reliability is generally considered necessary but not sufficient for validity.

The fairness of the test is also fundamental, although there is no single technical definition of fairness. Fairness refers both to the development and use of the test. While there are many dimensions to and debates over fairness, accessibility is a key aspect of fairness. According to the *Standards*, accessibility is " the notion that all test takers should have an unobstructed opportunity to demonstrate their understanding on the construct(s) being measured."[22] The ability of a test to measure its target skills and knowledge may be affected by certain characteristics of a test-taker that is not related to the target. According to the *Standards*, "a fair test does not advantage or disadvantage some individuals because of characteristics irrelevant to the intended construct."[23]

Fairness is also important both as related to a student's opportunity to learn (OTL) and during the testing process. For instance, if technology is involved, it is important that the test-takers have had similar prior exposure to the technology and

---

[20] *Standards* (2014), p. 11.
[21] *Standards* (2014), p. 33.
[22] *Standards* (2014), p. 49.
[23] *Standards* (2014), p. 50.

that the technology used by the test-takers is comparable. It should have similar processing speed, size of images, and other relevant attributes.[24] All test takers should have the same opportunity to demonstrate their mastery of the materials.

Accommodations are made in testing to allow for equitable access to demonstrate the knowledge and skills that are being assessed. It is important, however, that accommodations do not interfere with the measurement of the target of the assessment. For instance, if speed is one of the skills being tested, then allowing extra time as an accommodation would not be appropriate. Similarly, if English comprehension is being assessed, then it would not be appropriate to offer translated materials as an accommodation for students with limited English proficiency. The types of accommodations offered should depend on what skill or knowledge is being measured as well as the purpose of the test.[25]

In addition to validity, reliability, and fairness, there are many considerations and specifications involved in the development of a test. As an example, one critical choice concerns the length of the test. Test developers must balance the amount of time taken away from instruction for testing with the need to fully represent the content standards in a way that reflects their relative importance.   As the test is developed, it is also imperative to consider where the greatest precision is desired, particularly where the tests are intended to serve multiple purposes. Is it important to be able to differentiate among students in the middle of the distribution? Alternatively, is it important to differentiate among students at the top of the distribution or the bottom of the distribution? The answers to these questions should be driven by the purposes for which the assessment is being used. The limit on the number of test questions, which varies depending on the different item types, limits the precision of measurement and is related to the amount of time allowed for the test, forcing these types of trade-offs. While a longer test will generally allow for more items and greater precision, that desire has to be balanced with concerns about too much time spent on testing. This balancing act frequently comes back to the questions of: What exactly do you want to measure, for what purpose, and what incentives do you hope to create with the assessment?

Finally, while this discussion has focused on the test itself, it is more appropriate to think of the assessment system in which the test is embedded. In addition to the summative assessment itself, there are other aspects of the system that also have important consequences. For instance, the types of reports that are generated and how scores are reported have implications for the response to and use of the assessment. There are multiple audiences for the reports. Educators use the reports to inform their practice, and thus the content of the reports, as well as their usability matters. Districts may use the reports to inform curricula and professional development efforts. In addition, families and the general public are also an audience for score reports. An individual family might want to know: "How did my child do and what types of questions did he or she miss?"   Beyond individual

---

[24] *Standards* (2014), p. 51.
[25] *Standards* (2014), p. 62.

families, reports are also important in helping to create greater transparency about the test and the skills and knowledge it is assessing. In this way, the quality and usability of reports can contribute to the confidence in and the legitimacy of the standards-based educational system in the view of the broader public.

IV. A Brief Overview of MCAS and PARCC

We begin with a brief overview of each assessment before comparing different attributes of MCAS and PARCC.

About MCAS

MCAS was developed in response to the statutory requirements of the 1993 Massachusetts Education Reform Act (MERA). According to the 2013 MCAS technical report, MCAS is designed to:

- Measure student, school, and district performance in meeting the state's learning standards as detailed in the Massachusetts curriculum frameworks;
- Provide measures of student achievement that will lead to improvements in student outcomes;
- Help determine ELA, mathematics, and STE competency for the awarding of high school diplomas.

MCAS results are used to fulfill federal requirements by contributing to school and district accountability determinations.[26]

There is substantial information publicly available about the development and design of MCAS. Each year, a comprehensive technical report is produced that describes test design, development, and specifications. It includes information about the test's reliability and validity. Validity evidence includes information on test design and development; administration; scoring; technical evidence of test quality (classical item statistics, differential item functioning, item response theory statistics, reliability, dimensionality, decision accuracy and consistency); and reporting. All of these technical reports are available online at the Department's website.[27] In addition to the technical reports, in the early years, between 2003 and 2009, the UMass Center for Educational Assessment completed 20 validity reports on different aspects of MCAS. These validity reports, which are available online at the Center's website, cover a wide range of topics from model fit analyses to cognitive complexity levels. [28] While similar validity reports completed by an

---

[26] Massachusetts Department of Elementary and Secondary Education. (2013). 2013 MCAS and MCAS-Alt Technical Report. Retrieved from:
http://www.mcasservicecenter.com/documents/MA/TechnicalReport/TechReport_2013.htm
[27] The technical reports are available online at http://www.doe.mass.edu/mcas/tech/
[28] The validity reports are available online at http://www.umass.edu/remp/news_MCASreports.html

outside research organization have not been conducted since the standards were revised in 2010, these early reports focused on core features of MCAS and established a precedent to produce a high-quality assessment with transparency into the test development process.[29]

MCAS was designed to measure student proficiency in relationship to the Massachusetts Curriculum Frameworks. The grade 10 test is used to determine competency as part of the requirement for high school graduation. For more information on the Competency Determination, see the sidebar on page 17. In their report comparing MCAS and PARCC, the Massachusetts Business Alliance for Education (MBAE) and the Center for Assessment describe the purpose of the 10th grade MCAS as follows: "The passing score on the tenth grade MCAS test represents the minimum level of proficiency that all students have to meet to be eligible for a high school diploma."[30] At the time of MCAS's development, there was not a focus on the goal of college and career readiness comparable to the current policy context; MCAS was not designed to indicate a student's college and career readiness. Grade 10 MCAS was designed as an assessment of students' level of mastery of the knowledge and skills in the curriculum frameworks for 10th grade, a different criterion than that of college and career readiness. In addition, the Mass. Education Reform Act established the CD requirement in 10th grade in order to allow students time to remediate while still in high school, have additional opportunities to achieve the standard, and keep them on track to earn their high school diploma and graduate with their peers.

The lack of preparedness of high school graduates for the demands of college is a concern across the country and in Massachusetts. In 2012, 35 percent of students who were enrolled in a public college in Massachusetts took at least one remedial course. The remediation rates vary depending on the type of institution, with the lowest rates in the UMass system and the highest rates in community colleges. Nearly two-thirds of students (65%) who enroll in community colleges must take at least one remedial course.[31] For more information, see the sidebar "The Relationship Between Grade 10 MCAS and Remediation" on page 28. Concern about the high rate of remediation was a key motivation for revising the state's standards to ensure that high school graduates would be college (and career) ready.

---

[29] A detailed comparison of the 2010 Curriculum Frameworks and the previous Frameworks is available on the Department of Elementary and Secondary Education's website and can be accessed at http://www.doe.mass.edu/candi/commoncore/

[30] Massachusetts Business Alliance for Education and Center for Assessment. (2015). *Educating Students for Success: A Comparison of the MCAS and PARCC Assessments as Indicators of College- and Career-Readiness*, Boston, MA: Massachusetts Business Alliance for Education. p. 3.

[31] Massachusetts Department of Elementary and Secondary Education. (n.d.) [Data file.]

The Relationship Between Grade 10 MCAS and College Developmental Coursework

Concerns about remediation and also about what signals are being sent to families about their children's preparedness for college are important matters for Board consideration. Discussions about MCAS and the high rate of college remediation, however, typically do not differentiate between the different performance standards of MCAS (advanced, proficient, needs improvement, and failing) and thus paint an incomplete picture. The Department of Elementary and Secondary Education provided data regarding the number of students who enrolled in at least one remediation course in college (public higher education institution in Mass.) sorted by their performance levels on the 10th grade MCAS tests from 2004 to 2013. Professor Andrew Ho, a member of the Advisory Group for this report, analyzed the data, and his results are revealing.

The figure below, G10 MCAS and College Developmental Coursework, shows the implied correlation between student 10th grade MCAS performance and the need for a remedial class during the first year of college. The correlations between grade 10 MCAS performance levels and the probability of being assigned to college developmental courses are quite high, between 0.6 and 0.7. A perfect correlation would be 1.0, and 0.0 would indicate no relationship. The correlations are higher in ELA than math. There is also some evidence that the correlation between grade 10 MCAS and remediation have increased over time, particularly in math. MCAS performance levels explain about 36% - 49% of the variation in who takes a remediation classes. For example, in 2013, in ELA, the following share of students took a remediation class: 0% of the students who scored "advanced" on grade 10 test; 10% of the students who scored "proficient"; and 47% of the students who scored "needs improvement." In math, the remediation numbers are higher but grade 10 MCAS is still highly predictive of remediation. In 2013, 8% of students who scored "advanced" took a remediation class; 39% of students who scored "proficient" took a remediation class; and 62% of students who scored "needs improvement" took a remediation class.

A student's MCAS performance level is a strong predictor of remediation. Students who score higher on grade 10 MCAS are less likely to need remediation in college.
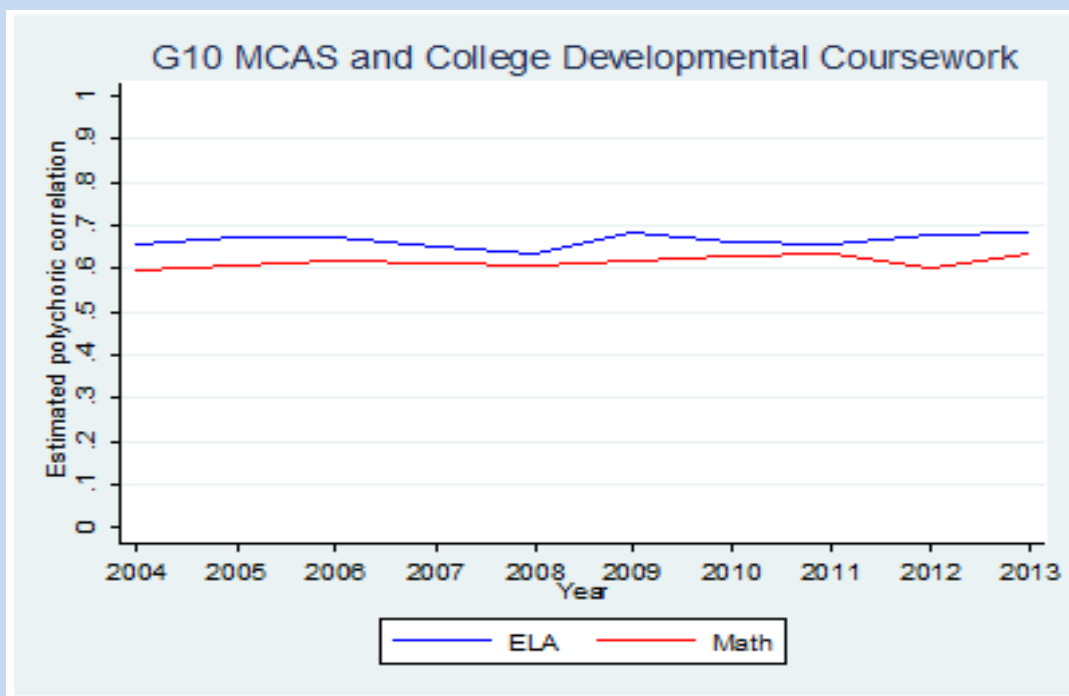
The 2015 report by Massachusetts Business Alliance for Education and Center for Assessment argues that MCAS has remained "virtually unchanged since the release of Version 1.0 in 1998" and that the changes to keep the assessment aligned with new content standards have only been incremental.[32] They conclude that the current MCAS tests do not identify students who are college and career ready.

The MBAE report also raises the concern that the elementary and middle school tests do not provide good information about student progress toward college and career readiness. According to the MBAE report, "Proficiency Standards were set independently for each grade level MCAS test when the test was introduced. That is, when the proficiency standards on the grades 4, 8, and 10 mathematics tests were established in 1998, they were set by separate panels and there was no discussion of the meaning or connection of those standards across grade levels."[33]

For a variety of reasons, including the passage of the federal No Child Left Behind law, the MCAS ELA and math tests in grades 3 – 8 were introduced for the first time at different points between 1998 and 2006. While the underlying content standards may reflect a general expectation of student progress over time, the performance standards set for each grade level do not convey consistent signals about a student's progress.

It is important to note that once performance standards are set for each grade level test, they remain consistent over time. Each subsequent version of the test includes items from the previous versions allowing the new test to be equated with the old tests, maintaining the same performance standards. Content standards define the knowledge and skills that a student should know and be able to do at each grade level. Performance standards set the level of performance or mastery that is expected of students to demonstrate a proficient or advanced level (or failing or needs improvement level) in relation to the content standards at each grade level. Resetting performance standards of a test within an assessment system is possible but requires careful consideration of its impact.

The differences in MCAS performance standards across grades make it difficult to evaluate a student's progress and to understand whether a student is "on track" to successfully reach the end goal. As an example, consider how the proficiency rates in MCAS math bounce around between different grades. In grade 3, nearly 70% of students score proficient. The share of students scoring proficient in grade 4 drops to 50%. Then, the share scoring proficient in grades 5 and 6 increases to 60%, and then it drops back down to 50% in grades 7 and 8.[34] These large changes in proficiency rates are counterintuitive. On the one hand, it could mean that there is a large drop in student achievement between 3rd grade and middle school. More likely, the drops and gains reflect a lack of consistency regarding how the different

---

[32] Massachusetts Business Alliance for Education and Center for Assessment (2015), p. 4.
[33] Massachusetts Business Alliance for Education and Center for Assessment (2015), pp. 22-23.
[34] Massachusetts Business Alliance for Education and Center for Assessment (2015), p. 23.

performance levels were set at each grade to represent the level of student mastery at each grade.

This latter explanation is supported by the Department of Elementary and Secondary Education's own internal analysis of raw MCAS scores, or the underlying scores of students before they have been "scaled" and classified within a particular performance standard. The bouncing around of the share of students who are placed in the "proficient" performance standard appears to reflect differences in how the performance standards have been set for each grade and not differences the underlying raw scores. Students are placed in one of each of the four performance standards (advanced, proficient, needs improvement, and failing) based on their scores. Because all student scores are represented by four standards, there is a range of raw MCAS scores within each standard. (On some score reports, the Department divides student scores into "high" and "low" for each performance standard.) According to the Department's analysis, despite the variation in the percentage of students placed in each performance level from year to year, there is a high correlation between a student's MCAS scores in lower grades and their scores in 10th grade. The correlation between 8th grade MCAS scores and 10th grade MCAS scores, for example, is high at 0.8.[35] These correlations show that students who do well on lower grades in MCAS are likely to do well in higher grades, and there is a very strong relationship between how students perform in lower grades and how they perform in subsequent grades, including high school. The commensurability of the performance levels is a separate question from the correlation of the underlying test scores.

The Department's analysis suggests that the underlying information about student mastery of the content standards provided by the MCAS test, as represented by raw scores, could potentially serve as a signal for students in lower grades in terms of being "on-track" for 10th grade, but that signaling has been obscured by the differences in performance standards established for the tests at different grades. Correcting this obstacle to enable better signaling would have implications for longitudinal student growth reporting and other comparisons across time, but the performance standards could, in fact, be aligned across grades.

To be clear, there exist other, and often related concerns, regarding MCAS, specifically concerning the 10th grade test. There are a number of issues that are important for the Board to consider regarding the appropriateness of the grade 10 test as the state's graduation requirement. We discuss some of the most pressing issues in three sidebars in this report. The relationship between grade 10 MCAS and college remediation is discussed in the sidebar on page 28. We also consider "Trends in MCAS Proficiency and College Remediation" on page 31. We discuss the complexity around determining the state's Competency Determination for students after the class of 2019 separately in a sidebar on page 17.

---

[35] Massachusetts Department of Elementary and Secondary Education. (n.d.) [Data file.]

Trends in Grade 10 MCAS Proficiency and College Remediation

The Department of Elementary and Secondary Education provided data regarding the number of students enrolled in at least one remediation course in college (public higher education institution in Mass.) sorted by their 10th grade MCAS proficiency levels from 2004-2013. Professor Andrew Ho, a member of this report's Advisory Group, analyzed the data, and his results shown in the figure below reveal two different trends regarding student proficiency and remediation rates.

In both ELA and math, the share of students scoring "proficient" on grade 10 MCAS has increased considerably. In 2004, 64% of students scored proficient or higher in ELA and 46% in math. By 2013, those numbers had increased dramatically. In ELA, 85% of students scored proficient or higher and 79% in math. During this same period, the remediation rates have stayed stable. In both 2004 and 2013, in ELA, 11% of students took a remediation class. In math, in 2004, 28% took a remediation class; in 2013, 27% took a remediation class.

There have been dramatic rises in MCAS proficiency, but remediation rates have not changed. Professor Ho finds that a student in 2013 within any MCAS performance level has almost 3 times the odds of remediation as a student with the same MCAS performance in 2004.

It is still true that within any given cohort, MCAS scores are a good predictor of whether a student will need remediation (as described in "The Relationship Between Grade 10 MCAS and College Remediation" on p. 28). This was true in 2004, and it is true in 2013. However, it is also true that a just-proficient student in 2013 is much more likely to need remediation than a just-proficient student in 2004.

Professor Ho's analysis raises a number of questions about the explanation for this change between Grade 10 MCAS scores and enrollment in developmental courses in college. There are several possible explanations and quite possibly the answer is a combination of explanations. It could be that the standards for assigning students to developmental coursework have risen, resulting in stable numbers of students recommended for and enrolling in developmental coursework, even though student proficiency rates have been higher than in years past. Alternatively, there could be increasing fadeout of MCAS knowledge that occurs between grade 10 and college, where students do not retain the material underlying their demonstrated proficiency in 10th grade.

A third alternative is that students have become able to raise their scores on MCAS without actually improving their underlying understanding of material relevant to place out of college developmental coursework (say, material assessed by the placement test, Accuplacer), and this lack of transferred mastery is reflected in the stable remediation rates. This is consistent with the finding that Massachusetts NAEP scores and MCAS scores in lower grades have remained relatively flat during this time. Further evidence is provided in the 2014 MCAS Technical Report. Psychometricians identified a technical issue particular to the Grade 10 test that may have led to the year-to-year underestimation of test item difficulties, resulting in systematic steady inflation of scores. As the report notes, steps were taken in 2014 to address this issue. If subsequent years of data show more stable trends commensurate with trends in enrollment in developmental coursework, this problem may be deemed historical.

Trends in MCAS proficiency and developmental coursework
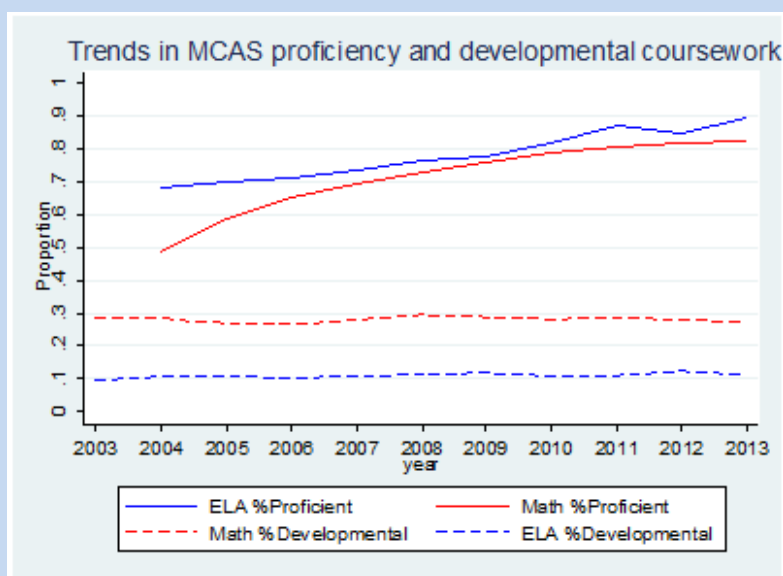
About PARCC

The explicit goal of the Common Core State Standards is to ensure that high school graduates are college and career ready.   In recent years, the goal for high school graduates has changed to more explicitly focus on preparing them for success after high school. The standards for each grade were developed across all grades with this end goal in mind. Developed as an assessment for the Common Core, PARCC was created to assess a student's college and career readiness or his or her progress toward that goal in the case of students in lower grades.

The design and development of PARCC has included the collection and analysis of a variety of data. PARCC has commissioned research that analyzes its 2014 field test, the item development, the test administration, accessibility, quality of items, comparability of the paper and computer-based assessments.   The findings of these reports have offered a lot of information about what has worked well and where improvements are needed.   The Department has indicated that it intends to summarize and discuss these reports with the Board. Based on the feedback, PARCC has made adjustments to the assessment. For instance, in response to concerns about too much time spent on testing, PARCC decided to reduce the amount of testing time by combining the performance-based assessments with the end-of-year assessments for one administration, starting in 2016.

PARCC includes three high school tests in grades 9, 10, and 11. For the 11th grade test, PARCC has set the threshold as college and career readiness. PARCC has 5 performance standards. By PARCC's standards, Levels 4 and 5 on the 11th grade or high school tests indicate that a student is college and career ready, meaning they are academically prepared to enter college and will be able take entry-level, credit-bearing courses without taking any remedial classes. In the lower grades, Levels 4 and 5 indicate "on-track" or "on the path" toward college and career readiness.

As part of its test development, PARCC conducted research with college faculty members to help determine its initial setting of performance standards so that they would be truly college and career ready. In fall 2014, a study, conducted with college professors who teach entry-level English and math courses, included 100 professors who reviewed the Algebra II items and 90 professors for the ELA/L items. After reviewing a test item, the professors were asked to answer the following question: "How many points would a borderline 'academically-ready' student likely earn if they answered this item (or set of items) on their first day in my class?" PARCC has defined borderline 'academically-ready' as a student who would have a 75 percent chance of earning 'just-barely' a C in a relevant course, without taking a remediation class. The professors' judgments were then used to determine an estimated total raw score for what PARCC would consider the borderline academically ready student.[36]

---

[36] Moyer, E. L., Miles, J., Davis, L. L. & You, W. (2015). *Postsecondary Educators' Judgment Study: Executive Summary*. Washington, DC: Partnership for Assessment of Readiness for College and Careers.

PARCC also did a benchmarking study to gather information from other international, national, and state assessments to help provide information about the percentage of students who are college and career ready. The expectations for Level 4 were defined in relationship to the results of the other assessments. For each external assessment, the researchers estimated the percentage of high school students graduating college- and career-ready as well as the percentage of students on track to readiness in earlier grades. Based on these findings, they then estimated a range of students who should be college and career ready on PARCC. The comparison assessments included: PISA, NAEP, ACT, SAT, the New York Regents exam, the Michigan Merit Exam, and the Virginia End-of-Course exams in high school. For grade 4, PARCC was compared with TIMSS, PIRLS, and NAEP.[37] These comparisons helped to inform the initial establishment of performance standards set by PARCC, including the definition of college and career readiness.

Finally, in the summer of 2015, as the final step in the standard setting process, the consortium used data gathered during the 2015 test administration to compare actual student performance with their earlier estimates and, based on that analysis, established the final cut scores to be used to distribute student performance across the five performance standards.

Over time, PARCC will need to continue using testing data and other analyses to ensure that the assessment is measuring and reporting student results in a manner consistent and aligned with its objectives. Empirically, over time, for example, PARCC will be able to investigate the claim that there is a 75 percent chance that students who earn a college-and-career readiness certification from PARCC will earn at least a grade of C in entry-level courses in college.

PARCC also continues to work with higher education institutions in the consortium to reach agreement that colleges will recognize and accept PARCC's college-and-career-ready determination. Students who earn a 4 or 5 would no longer have to take the Accuplacer or other placement exams.

PARCC also aspires to measure student progress in lower grades toward the end goal of college and career readiness. PARCC does not use a vertically integrated scale, but it has aimed to classify performance standards in each grade to indicate whether a student is on track to meet the goal of college and career readiness. PARCC will track progress in relationship to a common understanding of college and career readiness and the performance standards have been set with this end-goal in mind. The accuracy of PARCC's goal of measuring whether younger students are "on track" will be determined over time.

Based on the information gathered by the first wide-scale administration of PARCC, and administrations in future years, the consortium will need to continue

---

[37] McClarty, K., Korbin, J., Moyer, E. L. *et al.*, (2015). Draft: *PARCC Benchmarking Study*. Washington, DC: Partnership for Assessment of Readiness for College and Careers.

monitoring its performance standards and developing items to assess student performance in ways that return information across a broad range of performance. Given the concerns that were raised by the 2014 field test regarding the level of difficulty of items, this is an area that will require the dedication of appropriate additional resources and attention to both new item development and performance-standards. In order to ensure that the assessment system enjoys broad public confidence and support, that process of continuous evaluation and improvement by the consortium will need to be as public and transparent as possible. For additional information regarding the level of difficulty of test items and range of performance measured by those items see the discussion of "rigor" on page 51.

If the Board chooses to use PARCC as the statewide assessment system going forward, there are also a number of decisions that will need to be made regarding the high school graduation requirement for students after the class of 2019. It will have to decide whether college and career readiness is the appropriate requirement for high school graduation, and if not, what should be required, and how will that be communicated to students and their families. In addition, in PARCC, there are 2 different high school math tracks – the integrated math track or the Algebra I, Geometry, Algebra II track. The Board will need to consider how to set the graduation requirements equivalently for both tracks so that there is not one difficult and one easier track that create unintended incentives for students. How to use PARCC results will be decided by each state; it is not a consortium decision. At this time, there does not appear to be a consistent approach across consortium members on how graduation requirements will be set among the PARCC states that will use this statewide assessment as part of their graduation requirement.

The Predictive Validity of MCAS and PARCC

In the spring of 2015, the Executive Office of Education commissioned a study of nearly 850 first-year college students at 11 public colleges and universities throughout Massachusetts. The study was intended to provide objective evidence about the extent to which students' scores on the high school MCAS and PARCC math and ELA assessments accurately predict success in college. First-year college students at six community colleges, three state universities and two University of Massachusetts campuses chose to participate and were randomly assigned to take one component of the MCAS and PARCC math and ELA tests. Mathematica Policy Research was then contracted to analyze student scores and correlate them with the student performance in college (measured by grade point average), college readiness (measured by Accuplacer scores), and placement in remedial courses (measured by course enrollment data).

The central research question is: Do high-scoring students (on either MCAS or PARCC) perform better than low-scoring students? If the answer is "yes," then the scores on the assessment have validity in predicting college outcomes. The key findings include:

- Both MCAS and PARCC predict college readiness as measured by first-year college grades. Both MCAS and PARCC scores provide statistically significant predictions, and both are comparable to SAT scores in predicting first-year college outcomes.
- Scores on both MCAS and PARCC provide similarly strong predictions about which students need remedial coursework in college.
- In math, meeting the PARCC performance standard of college readiness predicts a higher college GPA than meeting the MCAS standard of proficiency. In ELA, there is not a statistically significant difference between the two performance standards.
- In math, students who achieve the college-ready standard on PARCC are less likely to need remediation than students who achieve the proficient standard on MCAS, while in ELA the two standards are not statistically distinguishable.

In sum, in ELA, both MCAS and PARCC provide equally useful information about college readiness. In math, the underlying scores of the two assessments are equally useful, but PARCC's performance standard of college readiness is better than MCAS's performance standard of proficiency at identifying students who do not need remediation and can earn "B" grades in college.

It is important to note that even though MCAS and PARCC are similar in their predictive validity regarding college grades and college remediation outcomes, they may differ in other ways. Differences in the content and structure of the two assessments could create different incentives, promoting different types of instructional practices within the classroom.

The full report by Mathematica Policy Research is available at https://www.mathematica-mpr.com/our-publications-and-findings/publications/predictive-validity-of-mcas-and-parcc-comparing-10th-grade-mcas-tests-to-parcc-integrated-math-ii

V. A Comparison of MCAS and PARCC

MCAS and PARCC are two different assessment systems. While they both might be aligned to the Common Core standards, there are still considerable differences between the two assessment systems. There are questions about depth and breadth of the assessments, beyond their measure of alignment. These differences will likely create different incentives for districts, schools, and students. The differences might also have consequences in terms of the broader discussion about standards-based education. As the Board considers the differences, it should come back to the questions of: what is the goal of the state's assessment system and which system is better suited to help meet that goal?

Relying on the most current information available, we have created a table (Table 3) that directly compares the MCAS and PARCC assessment systems on a number of different dimensions. It is important to note that the development and decisions around PARCC are still ongoing and not all issues are settled. The information in the table was compiled from publicly available sources and from information provided by members of the staff at the Department of Elementary and Secondary Education. For both MCAS and PARCC, this comparison is based upon the most current information available at the time that this report was prepared.

In addition, it is important to note that there have been ongoing considerations and discussions of how MCAS might be changed and improved. The detail and scope of such discussions, however, are not reflected in the table. The MCAS column in Table 3 is based on the current MCAS assessment system.

Finally, a note about reading the table: each row offers a snapshot of both assessments. For many rows, there is additional information to share. Below the table, we have included a narrative discussion for some of the rows. If additional information is included in the narrative section below the table, we have noted this fact in the right hand column of the table. The narrative discussion draws on relevant research and additional information ascertained by the author – to the extent that it is available.

Table 3: A Comparison of MCAS and PARCC Assessment Systems

		Current MCAS	PARCC	Additional Information
1.	**Who Oversees and directs Test Development**	Massachusetts Department of Elementary and Secondary Education (ESE)	The Partnership for Assessment of Readiness for College and Career (PARCC) Consortium	
2.	**Assessment Vendor**	Measured Progress (M.P.): Testing contractor	Pearson is the testing contractor; PARCC, Inc. is the project management partner	
3.	**Standards that are Assessed**	2010 Massachusetts Curriculum Frameworks (based primarily on the Common Core State Standards)	Common Core State Standards	page 44
4.	**Aligned to the Common Core Standards**	According to the Department of Elementary and Secondary Education, grades 3- 8, ELA and math are aligned. The grade 10 ELA and math MCAS are aligned to the ***overlapping standards*** that are in both the 2010 Mass. Frameworks and the 2000/2001 Mass. Frameworks, not the full Common Core.	Yes	page 44
5.	**Grades and Subjects Tested**	Grades 3–8 and grade 10 ELA Reading and math Grades 4, 7, 10 ELA Writing Grades 5 and 8 Science Technology and Engineering (STE) High school STE end-of-course tests in Biology, Chemistry, Introductory Physics, or Technology/Engineering	Grades 3- 8 ELA and math, High School Tests (Grade 9 - 11): ELA/Literacy I, II, and III, Algebra I, Geometry, Algebra II –OR– Integrated Math I, II and III (Students are assessed based on the courses taken.) Grades 5, 8, and high school STE (required but not offered by PARCC)	page 44
6.	**High School Competency Determination**	In addition to any local district requirements, students must either earn a scaled score of at least 240 (proficiency level) on the grade 10 MCAS ELA and Mathematics tests, or earn a scaled score between 220 and 238	TBD	Page 45

		Current MCAS	PARCC	Additional Information
		(needs improvement level) on these tests and fulfill the requirements of an Educational Proficiency Plan (EPP). Students must also earn a scaled score of at least 220 on one of the high school MCAS Science and Technology/Engineering (STE) tests: Biology, Chemistry, Introductory Physics, or Technology/Engineering		
7.	**Dates of Administration**	March–April: Grades 3–8 and 10 ELA May: Grades 3–8 and 10 math; Grades 5 and 8 STE June: High school (grades 9–12) end-of-course STE (Biology, Chemistry, Introductory Physics, and Technology/Engineering)	2015: Two administrations March–April: Performance-Based Assessments May–June: End-Of-Year Assessments 2016: One administration, including the Performance-Based Assessments April -June	

8.	**Item Types**	Multiple choice (all grades)
		Open responses (all grades)
		  o Requires students to generate, rather than recognize, a response. Students create a one-or two-paragraph response in writing or in the form of a narrative or a chart, table, diagram, illustration or graph, as appropriate.
		Short-response (grade 3 ELA only)
		  o Requires students to generate a brief response to a reading comprehension question
		Short-answer (math only)
		  o Requires students to generate a brief response, usually a numerical solution or a brief statement
			(All item types for all grades)
			Multiple choice
			Multiple select
			  o Requires students to select more than one correct answer to a question
			Short answer
			  o Requires students to generate a brief response to a reading comprehension or mathematics question. This may be written out or completed via the equation editor
			Prose constructed response
			  o Requires students to produce a written response to a text
			Technology-enhanced constructed response
				page 46

		Current MCAS	PARCC	Additional Information
		Writing Prompt (grades 4, 7, and 10) o Requires students to respond by creating a long written composition	o Requires students to respond to a test question by using technology, such as drag and drop, cut and paste, and highlight text features	
9.	**Rigor, as compared with other state standards as mapped onto NAEP***	4th grade reading = 3rd highest 4th grade math = 2nd highest 8th grade reading = 23rd highest 8th grade math = 4th highest This is a comparison across state standards.	TBD	page 51
10.	**Predicts College Readiness and Performance (as measured by placement into remediation and first-year college GPA)**	Yes	Yes	Page 35
11.	**Acceptance into credit bearing coursework by Massachusetts Public Higher Education Institutions without placement exams, based on state assessment performance**	No	Mass. Board of Higher Education plans to consider in coming months whether to accept the PARCC Consortium determination of CCR Level 4 or higher (of 5 performance levels) for placement directly into credit-bearing college entry-level English and math courses	page 54
12.	**Timed or Untimed**	Untimed: Recommended Time (2 hours per session): Grade 3: 8.0 hrs Grades 4&7: 12.0 hrs Grades 5,6,8: 8.0 hrs High School: 14.0 hrs Times shown are for ELA and Math	Timed: 2016 Administration Times for ELA and Math Grade 3: 8.25 hrs Grades 4-5: 8.5 hrs Grades 6-8: 9.2 hrs High School: 9.7 hrs	page 54
13.	**Computer-based or Paper-based**	Paper-based	Currently: Either computer-based or paper-based at district/school option; current contract until 2018 allows for paper-based option, but it is uncertain what will happen after the contract expires.	page 55

		Current MCAS	PARCC	Additional Information
14.	**Scoring: Machine or hand scored**	Multiple-choice items are machine scored. Open response, short-response, and short-answer items, and writing prompts (grades 4, 7, and 10) are hand scored. Cost for hand scoring is included in the fixed-cost contract with vendor.	Multiple-choice items, multiple-select items, and technology-enhanced constructed responses are machine scored; short answer and prose constructed response items are hand scored at an additional cost for Massachusetts.	
15.	**Accessibility**	Large menu of accommodations for students with an IEP or 504 plan, including changing setting, group size, timing/scheduling options, as well as test presentation in large print, Braille, read-aloud/electronic text reader CD; and response options, including use of scribe, approved graphic organizer, individualized Math reference sheet, calculators (for small number of students on non-calculator sections); tracking and monitoring placement of responses; word processor. 10th grade math is available in ASL on DVD.

Students with significant disabilities who cannot participate in the MCAS tests, even with accommodations, must take the MCAS Alternate Assessment (MCAS-Alt) (also under a contract with Measured Progress). | **Accessibility features** Available to *all* students, no IEP or 504 plan required, on the online and paper tests (including some that were previously available only as accommodations to students with disabilities on MCAS), such as zoom/screen enlargement, answer masking and place marker, text-to-speech/read-aloud (math only), and redirecting student to test. Additionally, each principal may determine administration considerations, such as test setting, group size, timing/scheduling;

Accommodations Available only for students with an IEP or 504 plan, including paper rather than online test, scribe, use of assistive technology, text-to-speech, embedded ASL video, hard-copy Braille, digital screen reader.

PARCC does not allow use of ELA graphic organizers or individualized math reference sheets for any students taking the assessment.

Students with significant disabilities | page 57 |

		Current MCAS	PARCC	Additional Information
			who cannot participate in the PARCC tests, even with accommodations, must take the MCAS-Alt developed and administered under existing, separate state contract with Measured Progress.	
16.	**Requirements and Accommodations for English Language Learner (ELL) Students**	All ELL students must participate in MCAS with the exception of first-year ELL students. The ELA test for first-year ELL students is optional. ELL students, including those who have been identified as Limited English Proficient (LEP) in the past, but are no longer reported as LEP, may use an authorized bilingual word-to-word dictionary and/or glossary. Grade 10 Spanish-speaking ELL students who have been enrolled in U.S. schools for fewer than three years may choose to take the English/Spanish edition of the grade 10 MCAS Mathematics test. English/Spanish editions of the Mathematics test are available for the grade 10 MCAS Mathematics test and retests only.	All ELL students must participate in PARCC with the exception of first-year ELL students. The ELA test for first-year ELL students is optional. ELL students, including those who have been identified as LEP in the past, but are no longer reported as LEP, may use an authorized bilingual word-to-word dictionary and/or glossary. ELL students are permitted extended time (up to the end of a school day) to complete a test session and general test directions read aloud, repeated, and/or clarified as needed in English or student's native language. They are also allowed approved Bilingual Word-to-Word dictionaries, scribing (or speech-to-text) responses on the math assessment, if dictated by the student in English.	
17.	**Retesting Options**	Prior to their scheduled graduation in 12th grade, students have at least five opportunities to earn a competency determination in ELA and math. For ELA and math, there are two regularly scheduled retest administrations each year in the fall and spring. Individuals may continue to participate in ELA and math retests and STE tests after leaving high school.	PARCC doesn't offer a specific retest form, but students may take the tests over again during the fall and spring administrations. This would be available for high school students only.	

		Current MCAS	PARCC	Additional Information
		MCAS Retests are specifically designed assessments and have separate administrations from spring MCAS administrations.		
18.	**Items Released**	In grades 3-8, approximately 50% of the common items are released. In high school, 100% of common items are released, with the exception of Chemistry. The 2015 items were released in July 2015.

In 2015, in grades 3-8, an average of 19 items were released in each grade in ELA and 21 items in math. In high school, 42 ELA and 42 math items were released. | Current PARCC policy is to release a model blueprint, which would represent a composite form of common items for all grades. In 2015, a more limited release in some grades is anticipated. It will include 40-50 items per grade, although it likely will be fewer released items in some grades.

In 2016 and subsequent years, PARCC plans to release an average of 45 items per grade, which will represent 33-50% of items per grade.

The 2015 item release is scheduled for the end of October 2015. In subsequent years, PARCC plans to release the items by the beginning of the school year. | page 58 |
| 19. | **When Results are Released** | Preliminary results made available to schools and districts in June and July; results provided to parents in September. | First operational year (school year 2014-15): results provided to parents in late fall/early winter. In future years, PARCC aspires to release results prior to the end of the school year, but it is unclear how Massachusetts's decision to continue hand scoring under separate contract will impact release schedule. | |
| 20. | **Types of Reports** | Annual reports include individual student results produced for parents/guardians, student rosters for schools and districts that include reports on individual student performance on specific released items, summary reports of school and district results (including posting to the ESE website), reports of | Annual reports include individual student results produced for parents/guardians. Electronic data will be transferred to ESE, where specific reports (the top six or so reports for MCAS will be modified for PARCC data in the first year for Edwin Analytics according to ESE and EOE IT staff) can | page 60 |

		Current MCAS	PARCC	Additional Information
		state results across MCAS tests, and reports of the performance of subgroups of students based on factors such as race/ethnicity and gender. All reports are generated by ESE except for the parent report, which is a Measured Progress product.	be developed and disseminated through Edwin Analytics (including aggregated data posted to the ESE website). Similar technical and cost challenges associated with the multiple forms of PARCC may pose limits on the production of reports related to item analysis.	
21.	**Diagnostic Tools and Formative Assessments**	None	Diagnostic assessments that are computer adaptive are planned for release by spring 2016. Beginning in school year 2016-17, PARCC diagnostic assessments will be available at cost to states or districts. PARCC will be releasing "downloadable pdfs" for K–2 formative assessments, scheduled to be During the first year (2015-16) there is no cost to districts. Beginning in school year 2016-17, PARCC formative assessments of some type and form are scheduled to be available at cost to states or districts.	
22.	**Technical Issues: test score reliability, validity evidence, approach to setting performance levels, approach to item analysis, assessment of content validity, DIF detection, and IRT model fit.**	MCAS Technical reports can be found here.	PARCC's technical report from the first operational year is not currently available.	

Source (for Row 9): Bandeira de Mello, V., Bohrnstedt, G., Blankenship, C., and Sherman, D. (2015). *Mapping State Proficiency Standards Onto NAEP Scales: Results From the 2013 NAEP Reading and Mathematics Assessments (NCES 2015-046)*. U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch.

Note: The information provided for PARCC is accurate as of 8/27/2015 but is subject to change. MCAS information is accurate with no changes anticipated.

MCAS assesses the 2010 Massachusetts Curriculum Frameworks, which include the Common Core standards plus additional standards that are unique to Massachusetts. According to the Department of Elementary and Secondary Education, one of the additional ELA standards and all nine of the additional math standards in grades 3-8 are currently assessed as part of MCAS.

The goal of MCAS is to be aligned with the 2010 Massachusetts Curriculum Frameworks (including the Common Core) in grades 3 – 8. The Department's efforts to bring the grades 3-8 MCAS into alignment with the new 2010 Frameworks was a multi-year process. While the updating and improvement of the assessments is always an ongoing process, the Department reports that the alignment with the new standards was substantially complete by the time grade 3-8 administrations took place in the spring of 2014.

The 10th grade MCAS test in ELA and math is different, however. For the 10th grade tests, MCAS is aligned to the **overlapping standards** in the 2010 Massachusetts Curriculum Frameworks and the 2000/2001 Massachusetts Curriculum Frameworks. That is, the current grade 10 MCAS test only assesses those standards that are part of both the 2010 and 2000/2001 Curriculum Frameworks (the intersection of the 2 standards), and as a consequence, students are not assessed across the full spectrum of the 10th grade standards.

The PARCC assessment has been developed, from its inception, to be aligned with the Common Core standards. It also aspires to assess and report on whether a student's annual performance on the assessments indicates that he or she is "on track" to meet the "college and career" standard that the high school assessments are designed to measure. The performance standards established by the consortium are designed to signal this "on track" status in a consistent manner across all grades assessed.

An independent evaluator is best positioned to determine whether an assessment is, in fact, aligned with a set of standards. There is currently research underway by the Thomas B. Fordham, Center for Assessment and the Human Resources Research Organization (HumRRo) to assess the alignment and quality of the PARCC, MCAS, SBAC, and ACT Aspire assessments. The full study is expected to be completed in early 2016, but preliminary results may be available at the end of October.

Grades and Subjects Tested

Both MCAS and PARCC include ELA and math tests for grades 3-8. MCAS has separate writing tests in grades 4, 7, and 10. PARCC incorporates writing into all of its tests but does not include a distinct, longer-form composition assessment similar to that of MCAS.

There are differences in the high school ELA and math tests, as well. Excluding retests for low performers, MCAS includes only one ELA and math test during the

high school grades; it is the grade 10 exam. PARCC includes tests in high school grades 9, 10, and 11. PARCC also allows two different math tracks – either integrated math or Algebra I, Geometry, and Algebra II.   Students are assessed at the end of course based on the specific courses taken. In PARCC math, students are expected to take a test in grade 9, 10, and 11 but the specific content of the math test will be dictated by their coursework.

In Massachusetts, students also take a Science and Technology/Engineering (STE) assessment in grades 5, 8, and high school. Currently, Measured Progress, the MCAS vendor to ESE for all ELA and Math assessments, is also the vendor for the STE assessments.

There are trade-offs involved in considering the most desirable approach to assessing high school students. Some people believe that more tests in high school are necessary to appropriately assess the breadth and depth of knowledge and skills and to track student growth over time more accurately. In addition, results from the 11th grade tests may be a better indicator of college and career readiness, since students will be one year closer to college. During public testimony at hearings conducted by the Board, many members of higher education community stated their belief that 11th grade performance will be a better indicator of a student's preparation for higher education. Although an 11th grade test will be a better indicator, the tradeoff is that there will be less opportunity for a remedial response for those students who are not on track. There are other consequences to assessing students in three years of high school. First, quite simply, it means more time spent on state tests. In addition, during some of the public hearings, some parents voiced concerns about testing in 11th grade, a year when many students already take the SAT or ACT and AP tests. The Board must balance a desire for more data, and data that are gathered closer to the end of high school, with concerns about too much time spent on testing.

High School Competency Determination

In addition to the local district requirements, the current state competency determination (CD) requires that students must either earn a scaled score of at least 240 (proficiency level) on the grade 10 MCAS ELA and math tests, or earn a scaled score between 220 and 238 (needs improvement level) and also fulfill the requirements of an Educational Proficiency Plan (EPP).

For students who score Needs Improvement, an Educational Proficiency Plan is required for each content area. According to the Department's website, the plan must include, *at a minimum*:

> (a) A review of a student's strengths and weaknesses in the content area.
> (b) The courses the student will be required to take and *successfully complete* in grades 11 and 12 that will move the student toward proficiency on the grade 10 curriculum framework standards as well as on grade 11 and 12 standards in English language arts or grade 11 and 12 grade span standards or Algebra II standards in mathematics.

(c) A description of the assessments the school will administer on at least an annual basis to determine whether the student is making progress toward proficiency, or has become proficient on the grade 10 standards. These assessments must include MCAS tests or other tests identified by the Department of Elementary and Secondary Education for this purpose.

Students must also earn a scaled score of at least 220 on one of the high school MCAS Science and Technology/Engineering (STE) tests: Biology, Chemistry, Introductory Physics, or Technology/Engineering.

The Board has voted to keep the current CD in place for students through the class of 2019 with retests available through at least 2020.

There is not a specific test designed to establish the Competency Determination within the PARCC system. Each member state will make its own determination about which, if any, of the PARCC tests will be utilized to determine eligibility for high school graduation, consistent with its own state policy and practice. Within the PARCC consortium, some states do not require that their students pass an assessment as a requirement for graduation. Among those that do, Maryland is planning to use assessments administered in the 10th grade to fulfill its graduation requirement, while New Mexico is planning to use assessments administered in the 11th grade.

Among the many complications that arise from the need to determine a high school competency standard is whether, and how, the college and career readiness standard should be related to the standard established for high school graduation. Leadership at ESE has begun to consider how various tests in the PARCC system of assessments could be utilized to establish a standard for high school graduation. Those preliminary considerations are not likely to result in a recommendation to the Board prior to the November decision on whether to adopt PARCC or pursue an alternative path for the state assessment system. For more information, see the "The 10th Grade Competency Determination" on page 17.

Item Types

The types of items – including the format of the stimulus and the format of the response - are critical elements in the ability of an assessment to accurately represent the breadth and depth of students' skills and knowledge in relation to the relevant content standards. Research finds that the types of items on an assessment can have an impact on classroom instruction. For instance, academic research finds that if an assessment includes writing, then there is a shift in the classroom to include more writing. (See sidebar on "The Influence of the Format of Assessments on Instruction" on page 22).

Because of limits in the amount of time available for a test, choices must be made about which standards to assess and which item types to use. A multiple-choice item takes less time for the test-taker and can be machine scored. At the same time,

a multiple-choice question is less likely to assess deeper thinking. An apples-to-apples comparison of test items on MCAS and PARCC is difficult because of the different types of items and different terminology. Tables 4 and 5 are intended to provide an overview of item types. It is important to note that because of the differences in the item types, they are not precisely comparable. In addition, because of the differences in scoring and weighting, the percent of total points does not mean that an assessment contains more of those items. For instance, in MCAS ELA, most tests include four open response items, compared with three Prose Constructed Response (PCR) items in PARCC, but the PCR items in PARCC account for a higher portion of the total points.   Tables 4 and 5 are intended to describe the item types on each assessment, but comparisons should be made with caution.

Table 4: ELA Item Types: MCAS and PARCC

MCAS ELA	Percent of Total Points	PARCC ELA	Percent of Total Points
Grade 3		**Grade 3**	
Multiple Choice (36 items @ 1 pt each)	75%	EBSR and TECR* (26 items @ 2 pts each)	55%
Short Response (4 items @ 2 pts)	17%		
Open Response (1 item at 4 pts)	8%	Prose Constructed Response (3 items @ 12-15 points each)	45%
Grades 4, 7, 10		**Grades 4, 5**	
Multiple Choice (36 items @ 1 pt)	50%	EBSR and TECR * (28 items @ 2 pts)	53%
Open Response (4 items @ 4 pts)	22%		
Writing Prompts (1 item @ 20 pts)	28%	Prose Constructed Response (3 items @ 12-19 points each)	47%
Grades 5,6, 8		**Grades 6,7,8,9,10,11**	
Multiple Choice (36 items @ 1 pt)	69%	EBSR and TECR* (34 items @ 2 pts)	56%
Open Response (4 items @ 4 pts)	31%		
		Prose Constructed Response (3 items @ 15-19 points each)	44%

Sources: MCAS ELA Blueprint.
http://www.doe.mass.edu/mcas/tdd/ela.html?section=testdesign, PARCC ELA High Level Blueprint Grades 3-5.
http://www.parcconline.org/files/83/Spring%202016/97/PARCC%20Grades%203-5%20ELA%20Literacy%20Combined%20Common%20Form%20Specifications.pdf.pdf, and PARCC ELA High Level Blueprint Grades 6-11 http://www.parcconline.org/files/83/Spring 2016/388/Grades 6-11 High Level Blueprint (Updated).pdf.

* Refers to Evidence-Based-Selected Response and Technology-Enhanced-Constructed Response, which is only for the computer-based tests.

Table 5: Math Item Types: MCAS and PARCC

MCAS Math	Percent of Total Points	PARCC Math	Percent of Total Points
Grade 3		**Grades 3-8**	
Multiple Choice (26 items @1 pt)	65%	Type I (29-36 items @ 1 pt or 2pts)*	60%
Short Answer (6 items @ 1 pt)	15%	Type II (includes multiple-choice) (4 items @ 3 pts or 4 pts)	21%
Open Response (4 items @ 2 pts)	20%	Type III (3 items @ 3 pts or 6 pts)	20%
Grades 4-8			
Multiple Choice (32 items @1 pt)	59%		
Short Answer (6 items @ 1 pt)	11%		
Open Response (4 items @ 4 pts)	30%		
Grade 10		**High School**	
Multiple Choice (32 items @ 1 pt)	53%	Type I (32 -42 items @ 1 pt, 2 pts or 4 pts)	60%
Short Answer (4 items @ 1 pt)	7%	Type II (4 items @ 3 pts or 4 pts)	17%
Open Response (6 items @ 4 pts)	40%	Type III (4 items @ 3 pts or 6 pts)	22%

Sources: MCAS Mathematics Test Blueprints.
http://www.doe.mass.edu/mcas/tdd/math.html?section=testdesign and PARCC High Level
Blueprint - Mathematics. http://www.parcconline.org/files/96/Spring
2016/143/PARCCHighLevelBlueprints-Mathematics_08.25.15.pdf

*Grades 6-8 assessments include Type I items that are worth 4 pts.

Both MCAS and PARCC use multiple-choice items. Multiple-choice questions have
the advantages of providing information across a breadth of learning, are relatively
quick to administer, and are inexpensive to score, but they are also susceptible to
test-taking tricks such as strategic guessing, and they may encourage undesirable
pedagogy.[38] Although there can be different levels of cognitive complexity within
multiple-choice items, research finds that multiple-choice questions are less likely
than open-ended items to assess more complex thinking. A heavy reliance on
multiple-choice items also makes it more likely that the depth of alignment to
content standards is weak while the breadth of alignment may be strong.[39]

There is also concern that when multiple choice tests are used in high-stakes
environments, teachers might have an incentive to emphasize the skills and
knowledge tested by those questions and may place a lower priority on the deeper

---

[38] Marzano, R. J., Pickering, D, McTighe, J, (1993). *Assessing Student Outcomes: Performance
Assessment Using the Dimensions of Learning Model*. Mid-Continent Regional Educational Lab.
(McREL). Aurora, CO.
[39] Braun, H. personal correspondence, September 27, 2015.

learning skills that are not typically assessed through multiple-choice questions.[40]

In ELA, PARCC uses a particular type of multiple-choice question called Evidence-Based Selected-Response (EBSR) items that are two-part multiple-choice questions. The goal of this format is to delve deeper into student understanding and to reduce the likelihood of guessing. Part A of the EBSR question is designed to measure reading comprehension. As in a traditional multiple-choice question, the student reads the question, considers the choices, reviews the text, eliminates incorrect choices, and marks the correct response. For some EBSR questions, students are asked to identify more than one correct answer. For Part B, students are asked to support their Part A response with evidence from the text. Each EBSR item is then worth 2 points. Students can earn full credit by getting both questions on Part A and Part B correct. Or, they can earn partial credit (one point) by getting Part A correct and Part B incorrect. However, if students have an incorrect answer on Part A but a correct answer on Part B, they earn no credit. Some researchers have cautioned that these items require more complex scoring procedures due to the dependent nature of these item components.

The online PARCC ELA test also uses an item type called Technology Enhanced Constructed Response (TECR), which is designed to capture student comprehension by using drag and drop, cut and paste, and highlight text features to move items from a text or about a text to construct a response that supports students' prior responses.[41] These responses are often paired with the EBSR items as a follow-up question (Part B) in which students are asked to provide evidence for their response in Part A in the form of a short answer. These items are only on the online PARCC test. For the paper-based assessment, PARCC has tried to develop EBSR items that closely match the content of the TECR items.

In addition to multiple-choice questions, both MCAS and PARCC include short-answer and short-response items, which require students to generate a brief response in writing. Like multiple-choice questions, these item types require a relatively small amount of time for test takers and are simple and quick to score. However they encourage demonstration of a slightly deeper level of learning than can be assessed with multiple-choice responses because students must generate their own responses, rather than choosing the correct answer from a provided list.

Open response and constructed response items are used to assess skills and knowledge that are too complex for either multiple-choice or short-answer/short-response questions, and both PARCC and MCAS include these items. Open response and constructed response items are designed to assess more complex learning such as students' ability to synthesize and analyze information and make it much easier to determine what students know as compared to when they are guessing. However,

---

[40] Livingston, S.A. (2009). *Constructed-Response Test Questions: Why We Use Them; How We Score Them*. Educational Testing Service (ETS). Retrieved from:
https://www.ets.org/Media/Research/pdf/RD_Connections11.pdf
[41] PARCC Glossary of Terms. Retrieved October 5, 2015 from:
http://www.parcconline.org/resources/parent-resources/glossary-of-terms

these types of items require more time for the test takers, which involves trade-offs within the constraints of the total amount of time for the assessment; they are also more expensive to score and are vulnerable to interrater variability in scoring.[42]

The MCAS ELA assessment includes long composition (writing prompts) and open-response questions. According to the Department of Elementary and Secondary Education, the composition item is designed to assess writing, so Standard English conventions, style, and organization are scored. These items are intended to elicit a student's thoughts on a topic, supported by a student's experience or ideas. Open response questions, in contrast, are intended to assess reading comprehension, so a student must answer the questions asked using answers found in or inferred from the reading passage provided and support those answers with details from the text.[43]

For the PARCC ELA assessment, the constructed response item, called the Prose Constructed Response (PCR), is designed to assess students' knowledge through each of three assessment tasks: the literacy analysis task, the narrative task, and the research simulation task. The literacy task requires students "to carefully consider literature worthy of close study and compose an analytic essay." The narrative task requires students to write a story (or the next part of a story), detail a scientific process, write a historical account of important figures, or to describe an account of events, scenes or objects, for example. In the research simulation task, students analyze an informational topic presented through articles or multimedia. Students engage with the texts by answering a series of questions and synthesizing information from multiple sources in order to write an analytic essay.[44] In PARCC, there is no item comparable to the MCAS writing prompt in which students generate their own writing and are scored on their writing skills.

In the PARCC Math assessment, item types are categorized as Types I, II and III. According to PARCC, Type I tasks are designed to assess concepts, skills and procedures, a balance of conceptual understanding, fluency, and application and are machine scorable. Type II items assess students' ability to express mathematical reasoning. These items call for written arguments/justifications, critiques of reasoning, or precision in mathematical statements and may include a mix of machine scored and hand scored responses. Type II items can include multiple-choice questions. Type III items assess modeling and application. These items require test takers to model/apply mathematical concepts in a real world context or scenario and may include a mix of machine scored and hand scored responses. [45]

---

[42] Hollingworth, L. Beard, J.J. & Proctor, T.P. (2007). An Investigation of Item Type in a Standards-Based Assessment, *Practical Assessment, Research and Evaluation*. Vol 12, No 18.

[43] Massachusetts Department of Elementary and Secondary Education. ELA Composition vs. Open-response Questions, Retrieved from: http://www.doe.mass.edu/mcas/tdd/longvopen.doc

[44] PARCC ELA/Literacy. Retrieved from: www.parcconline.org

[45] PARCC High Level Blueprints – Mathematics. Retrieved from: http://www.parcconline.org/files/96/Spring%202016/143/PARCCHighLevelBlueprints-Mathematics_08.25.15.pdf.

Rigor

There is no single definition of rigor. The "rigor" of an assessment can refer to the difficulty of its items, the cognitive complexity of its items, the stringency of its performance standards, the relevance of the construct it measures, and the appropriateness of the pedagogical responses the assessment inspires. It typically refers to both difficulty and cognitive complexity. Rigor, however, should not be confused with quality. It is possible to design a "bad" test that sets a high standard such that only a few students pass. In this case, "bad" might mean that the test does not do a good job of measuring what it is intending to measure or that it provides no useful information to students, teachers, administrators, and policymakers. While rigor is a not a substitute for quality, it is hoped that a rigorous assessment will lead to an increase of academically challenging material and instruction in the classroom.

For the purpose of this report, we consider rigor in two different ways – first in terms of the performance standards mapped to the NAEP test. NAEP is used as the point of comparison because it is a national- and state- representative assessment that enables comparisons across states of student performance and, indirectly, of state standards. The second manner by which we discuss rigor is by reviewing a Rand study that utilized Webb's Depth-of-Knowledge (DOK) Framework.[46]

The National Assessment of Educational Progress (NAEP), often referred to as the nation's report card, is an assessment given to a sample of students nationwide in grades 4, 8, and 12. The American Institutes for Research (AIR), under contract with the National Center for Education Statistics, recently mapped different states' standards onto NAEP scales. AIR matched the percentage of students who have met the proficiency level on the state assessment (MCAS) in math and ELA in grades 4 and 8 in each NAEP school to the point on the NAEP achievement scale corresponding to that percentage. "For example, if the state reports that 70 percent of the students in fourth grade in a school are meeting the state's reading achievement standards and 70 percent of the students in the NAEP achievement distribution in that school are at or above 229 on the NAEP school, then the best estimate from that school's results is that the state's standard is equivalent to 229 on the NAEP scale."[47]

This mapping allows an understanding of both how an individual state's standards compare with the 3 achievement levels of NAEP and, perhaps more important, allows a comparison across different states in terms of the rigor of their standards, using NAEP as the measuring stick.

---

[46] Yuan, K. and Le, V. (2012). *Estimating the Percentage of Students Who Were Tested on Cognitively Demanding Items Through the State Achievement Tests*. RAND Working Paper. Santa Monica, CA: RAND Corporation. WR-967-WFHF. Retrieved from: http://www.rand.org/content/dam/rand/pubs/working_paper/2012/RAND-WR967.pdf

[47] Bandeira de Mello, V., Bohrnstedt, G., Blankenship, C., and Sherman, D. (2015). *Mapping State Proficiency Standards Onto NAEP Scales: Results From the 2013 NAEP Reading and Mathematics Assessments* (NCES 2015-046). U.S. Department of Education, Washington, DC: National Center for Education Statistics. Retrieved on September 25, 2015 from http://nces.ed.gov/pubsearch, p. 3.

NAEP reports results with respect to three performance standards: Proficient, Basic, and Below Basic. In Massachusetts, in 2013, the grade 4 reading MCAS proficiency standard mapped onto the NAEP Basic range, and the grade 4 MCAS math proficiency standard mapped above the NAEP Proficient score. In grade 8, both MCAS reading and math proficiency standards mapped onto the NAEP Basic range.

In three out of the four tests, the Massachusetts MCAS standards for proficiency are among the highest in the nation. In grade 4 reading, the Massachusetts proficiency standards are the 3rd highest; the standards in New York and Wisconsin were higher. In grade 4 math, the Massachusetts proficiency standards are the 2nd highest in the nation; Texas had the highest standards. In grade 8 math, the Massachusetts proficiency standards are the 4th highest: the standards in New York, Texas, and North Carolina were higher. The exception is 8th grade reading. In 8th grade reading, Massachusetts proficiency standards are the 23rd highest in the nation; New York and Wisconsin had the highest standards in 8th grade reading.

Compared with the same mapping process that was done two years earlier, however, the rank of Massachusetts standards has declined compared with other states' standards. It appears that other states have adopted higher proficiency standards. In 2011, Massachusetts had the highest proficiency standards in the nation for grade 4 reading and math. For grade 8 math, Massachusetts had the 2nd highest proficiency standards. For grade 8 reading, Massachusetts had the 17th highest proficiency standards.[48]

At this point, we don't have comparable information that would map PARCC performance standards onto NAEP. NAEP is only one benchmark but is often used to facilitate comparisons across states because it is a common assessment.

We also reviewed a 2012 RAND analysis of MCAS and 16 other state assessments using Webb's Depth-of-Knowledge (DOK) Framework. Norman Webb defined four levels of cognitive rigor.[49] His four levels of cognitive rigor are:

- Level 1 represents recall;
- Level 2 represents demonstration of skill/concept;
- Level 3 represents strategic thinking; and
- Level 4 represents extended thinking.

In their analysis of 5,100 test items across 17 states' assessments, they found the overall rigor of math and ELA was low, especially for math. In terms of item types, Yuan and Le found that open-ended items had a greater likelihood than multiple choice of reaching DOK Level 3 or 4, representing deeper thinking. By the nature of the format, multiple choice questions typically do not require extended thinking, although the questions can still be challenging and sophisticated.

---

[48] Bandeira de Mello, V., Bohrnstedt, G., Blankenship, C., and Sherman, D. (2015), pp. 8-23.
[49] Yuan and Le (2012), pp.14-15

The purpose of their research was to look at the 17 states collectively, and as a consequence, they do not provide any ranking of individual states, but they do report state-by-state findings in the appendices. Their analysis of MCAS items in Massachusetts was based on items used in 2010-11 academic year. According to their analysis, the math items in all grades of MCAS – including both multiple choice and open-ended questions - were almost entirely classified as Level 1 or 2.   The majority of the multiple-choice questions in math in all grades were Level 1. With the exception of Grade 3, the majority of the open-ended questions were Level 2. In grade 7 math, only 1 of the 10 open-ended questions was Level 3; the others were Levels 1 and 2.

In reading, both the multiple-choice and open-ended MCAS items were more cognitively rigorous than their counterparts in math. For the multiple-choice questions, there were differences in rigor between the different grades. For instance, in grade 7, 13% of the questions were Level 1, 44% were Level 2, and 44% were Level 3, while in grade 10, 27% were Level 1, 57% were Level 2, and 16% were Level 3. In grade 6 reading, there was a Level 4 multiple-choice question.

The open-ended questions in reading were more cognitively rigorous than the multiple-choice reading questions, although only a very limited number of items were released.   None of the released open-ended questions were Level 1. There was one Level 4 open-ended question in grade 4. The rest of the open-ended questions were a mix of Level 2 and Level 3.[50]

Although these findings offer some insight into the cognitive rigor of MCAS, they should be interpreted with caution because they are based on the MCAS four years ago. MCAS incrementally changes over time, and we do not know if this finding is accurate for the most recent MCAS. In addition, we do not have a comparable analysis of PARCC items. The intention of PARCC is to assess every student on the full range of cognitive complexity, meaning Levels 1 – 4 by Webb's Depth of Knowledge (DOK).

While we don't have a depth-of-knowledge or other cognitive complexity analysis of PARCC items, there was an analysis of the difficulty of items on PARCC's field test. Based on this analysis, there is concern that the assessment might be too difficult. In the 2014 field test, PARCC PBA and EOY assessments, in both ELA and math, had very few items flagged as easy (fewer than 2% for all assessments). This is based on the proportion of students answering the items correctly. Many more items were flagged as difficult, especially for the math assessments. In grade 10 ELA, between 21% and 47% of items were flagged as difficult (fewer than 30% of students answered correctly). More than 50% of items were flagged as difficult for all math grades higher than grade 6 for both the performance-based and end-of-year assessments. In fact, some subject area tests had as many as 76% of items flagged as

---

[50] See Appendix B in Yuan and Le (2012) for Massachusetts specific results.

difficult.[51] According to a report by PARCC, "These data are concerning because such a difficult test may not differentiate well among students in the center and lower half of the distribution, limiting the potential interpretations of scores for many students."[52] We do not have the comparable information for the actual administration in 2015, but it should be an area of consideration.

Acceptance by Public Higher Education Institutions

High school students who matriculate at public higher education institutions typically must take the Accuplacer exam or some other college placement exam to determine if they are prepared to enter but not necessarily succeed in credit-bearing courses in college. There have been some recent changes, however. In fall 2013, the Board of Higher Education (BHE) authorized the use of new criteria for placement in developmental education and college-level courses. In fall 2014, public higher education institutions began using student GPA and other placement methods and will continue to do so until spring 2016. At a future date, the Department of Higher Education will make a recommendation to the Board of Higher Education about modifying existing policies, depending on the outcome of the pilot efforts.

The development of PARCC included the input of faculty members from higher education institutions with the goal of creating better alignment between the K-12 system and the expectations of higher education institutions. The PARCC Consortium has stated that their definition of success will be: students who earn Level 4 or higher will have about a 75 percent chance of earning at least a grade of C in entry-level courses in college.

The Department and Board of Higher Education have been in discussion with the Department of Elementary and Secondary Education and their colleagues in PARCC consortium regarding whether the Board of Higher Education will accept the PARCC determination of college and career readiness (Levels 4 or 5 on the 11th grade or high school assessments) for placement directly into credit-bearing college entry-level English and math courses.

Timed or Untimed and Total Time for Testing

The MCAS assessment is an untimed test, meaning that students can spend as much time as they need to complete the test. It is recommended that schools allow 2 hours per session, and there are different numbers of sessions depending on the grade level. The recommended amount of time for both annual MCAS ELA and math assessments ranges from 8.0 to 14.0 hours, depending on the grade level.   However, because of variation in actual test performance time, it is reported that many

---

[51] Thacker, A., Dickinson, E., Bynum, B. *et al.* (2015). *Findings from the Quality of Items/Tasks/Stimuli Investigations: PARCC Field Tests*. Washington, DC: Partnership for the Assessment of Readiness for College and Careers.

[52] Thacker, Arthur *et al.* (2015), p. 74.

schools schedule much longer blocks of time impacting the school day more than the recommended time might suggest.

In contrast, PARCC is a timed assessment. English Language Learners and students with disabilities are allowed to take an untimed PARCC until the end of school day for each session. Based on the feedback from the testing in spring 2015, PARCC is planning to shorten the amount of time spent on testing by about 90 minutes overall (60 minutes in math and 30 minutes in ELA) by shortening the assessments and also consolidating the performance-based assessments and the end-of-year assessments into a single assessment. In 2016, the time allocated for PARCC will range from 8.25 to 9.7 hours depending on the grade level. There has been no publicly available analysis of what the impact of this consolidation and shortening of the test will be regarding the precision of the assessment and its ability to assess students at all levels.

After the first administration of PARCC in 2015, there was a student survey to get feedback from those students who took the computer online. One of the questions asked: "Did you have enough time to finish the test?" More than 90% of students in Massachusetts responded that they had enough time. About 8% said that either they were rushed or that they did not complete the test.[53] Based on these results, it appears that almost all students had adequate time to finish the computer-based PARCC assessment in the unconsolidated form in which it was administered in 2015.

Computer-based or Paper-based

The computer-based format of PARCC as compared with the more traditional paper-based format of the current MCAS provides some advantages and disadvantages that are important to explore. PARCC's computer-based format is intended to provide several advantages to paper-based tests: improving the precision of measuring difficult-to-measure constructs, better aligning assessments to instruction, engaging and motivating students, improving accessibility of assessments for English language learners and students with disabilities, expediting the return of test results, and improving the ease of interpreting test results.[54] There are also potential disadvantages; most prominent among them are a lack of access to technology and the cost of expanding this access. With a computer-based test, students often cannot all be tested at one time on the same day, as with a paper-and-pencil test.[55] In order to efficiently administer the computer-based version of PARCC, many districts in Massachusetts will need to incur significant costs.

---

[53] Results from the 2015 PARCC student surveys, 2015, Table 6.
[54] Educational Testing Service. *Technology-enhanced Assessments*. Retrieved from: https://www.ets.org/c/22116/
[55] Rennie Center for Education Research & Policy (2015). *Testing the Test: A Study of PARCC Field Trials in Two School Districts*. MA: Rennie Center for Education Research & Policy. Retrieved from: http://renniecenter.org/topics/PARCC_case_study.html.

It is also important to note that PARCC is computer-based, not computer adaptive, in contrast to the assessments developed by the Smarter Balanced Assessment Consortium. Computer adaptive tests hold some advantages when compared with fixed computer based tests. Computer adaptive tests can target items, eliminating the items that are too easy or too hard for each individual student. This can then result in either shortening the testing time or allowing for more precision by using the time to administer additional targeted items. The greater efficiency can also be used to provide more diagnostic information about students' areas of difficulty.[56] However, computer adaptive tests can require a much bigger pool or bank of questions than fixed assessments to keep test security high, but this would increase cost (at least until a desirable item bank size has been achieved) and limit PARCC's ability to publicly release items after each test administration. So again, trade-offs must be considered—shorten test time with the use of an adaptive test design versus retaining the fixed test design (all candidates in the same subject and grade level taking the same test) and likely lowering the cost of testing. Actual costs and comparisons would not be known until specifics are developed for each test.

While not adaptive, the computer-based PARCC assessment does include several technological enhancements not possible with paper-based tests. One of the advantages of computer-based assessments is the opportunity to assess more complex cognitive skills while still being machine scored, thus, allowing a more cost effective method of assessing such skills. Some of the technological enhancements include: 1) drag and drop - used to provide supporting evidence, sequence steps or match equations with words; 2) multiple select - used to choose multiple themes or central ideas, identify synonyms or antonyms, or select equivalent equations; 3) text select/highlighting - used to support an idea, or provide context; 4) equation builder - used in math to build and solve equations related to word problems, to justify answers, or to prove answers; 5) drop down menus - used in math to build and solve equations; 6) constructed response – used to revise and rewrite given passages, write conclusions, or explain numerical results; and 7) multiple part questions – used to probe students' deeper knowledge and to justify answers in math.[57]

Because PARCC currently provides a paper-based form of their test, some researchers have cautioned that PARCC's provision of a computer-based and paper-based option presents challenges in comparability. For example, the technological enhancements described above cannot be offered to paper-based test takers resulting in groups of students who take two different assessments but are held accountable to the same performance standards. Attention to these concerns has

[56] Policy Analysis for California Education and Rennie Center for Education Research & Policy. (2011). *The Road Ahead for State Assessments.* MA: Rennie Center for Education Research & Policy. Retrieved from: http://renniecenter.issuelab.org/resource/road_ahead_for_state_assessments
[57] Ronan, A. (2015, February 7). Seven Technology Enhanced Item Types You'll See on Common Core Tests This Spring. Edudemic. Retrieved from: http://www.edudemic.com/7-tech-enhanced-items-on-common-core-tests/

resulted in less extensive use of technology-enhanced items in PARCC than might have been the case if PARCC were solely computer-based.

Accessibility

There are differences between MCAS and PARCC in terms of their approach to accessibility and accommodations for students. MCAS allows a large menu of accommodations for students with an IEP or 504 plan. In contrast, in PARCC, there are a number of accessibility features that are available to all students. Some of these features, such as the ability to zoom in or enlarge the screen are related to fact that PARCC is commonly computer-based, and these features are available to any student who is using a computer.

But, PARCC's approach to accessibility goes beyond the fact it is designed as a computer-based assessment. In PARCC, "each principal can determine administration considerations, such as test setting, group size, and timing/scheduling" for any student – not just students with an IEP or 504 plan. In addition a text-to-speech/human reader for the PARCC math test is an accessibility feature available to all students, whereas in MCAS a test administrator may read aloud the math test only to students, as specified in an IEP or 504 plan.

At the same time, not all of the accommodations allowed under MCAS are allowed under PARCC. Specifically, PARCC does not allow students to use individualized graphic organizers, checklists or reference sheets, whereas these accommodations are allowed under MCAS.

While PARCC tests are timed, students with disabilities and English Language Learners are allowed extended time until the end of the school day.

For more information about PARCC accessibility, please see an overview prepared by the Department of Elementary and Secondary Education: PARCC Accessibility and Accommodations - Overview for MA Educators.

While thinking about accessibility, it is important to note that many experts believe that computer-based tests offer the opportunity to improve the accessibility of test items for all students. Tests can be developed so that they are accessible for all students in the design process, rather than making accommodations after the "regular" test has already been designed. Michael Russell, a Professor at the Lynch School of Education at Boston College, argues that "By applying principles of accessible test design, the next-generation assessment systems will deliver more valid inferences about student learning based on test scores for all students."[58]

According to Russell, computer-based tests allow the opportunity to embed additional accessibility information into the digital content files as an item is

---

[58] Russell, M. (2011). *Digital Test Delivery: Empowering Accessible Test Design to Increase Test Validity for All Students*. Bill and Melinda Gates Foundation.

developed and the ability of a computer-based test to selectively allow subsets of that information to individual students based on their specific need.

As with any accommodations, the types of accommodations offered should depend on what skill or knowledge is being measured as well as the purpose of the test. Russell explains, "Perhaps the most important challenge to address focuses on developing a clear definition of the intended construct measured by an item. Without a clear definition, it is difficult to determine whether supplemental and alternate item content alters the construct measured by the item."[59] While accommodations allow for equitable access to demonstrate the knowledge and skills that are being assessed, it is important that accommodations do not interfere with the measurement of the target of the assessment. Nonetheless, it appears that computer-based assessments, because of the nature of the format, have advantages beyond paper-based assessments in terms of their potential accessibility for all students.

Release of Test Items

Until 2009, all common MCAS test questions were released. Student scores are based only on common items; the other items on MCAS, called matrix items, are used for field testing and equating.[60] Concerns about too much classroom time being spent on testing led to changes in this policy. Reducing the number of items released decreases the number of items that need to be developed and field-tested, which allows for a shorter test and also reduces test development costs. Today, about half of the common test items are released in grades 3-8, and 100% of the common items on the 10th grade test are released. In 2015, in grades 3-8 ELA, half of the items translated into an average of 19 items in each grade that were released, and in math, an average of 21 items in grade were released. In grade 10, 42 items were released in ELA and 42 items in math.

The Department of Elementary and Secondary Education maintains a publicly available database of 3,238 MCAS items from the past five years of tests on its website. In addition, the Department provides access to each released short-response question, open-response question and writing prompt that was included on the MCAS tests from the last five years, including the scoring guide that accompanies it. There are also samples of student work at each score point for that question. All of this publicly available information is intended to offer a clear

---

[59] Russell (2011), p. 10.

[60] On each test, there are common and matrix items. The 2013 MCAS Technical Report explains: "Common items are administered to all students in a given grade level. Student scores are based only on common items. Matrix items are either new items included on the test for field-test purposes or equating items used to link one year's results to those of previous years. In addition, equating and field-test items are divided among the multiple forms of the test for each grade and content area. The number of test forms varies by grade and content area but ranges between 5 and 32 forms. Each student takes only one form of the test and therefore answers a subset of equating and field-test items" (p. 8.).

picture of the expectations for student performance. This approach has facilitated a broad transparency of the MCAS to educators, families, and the general public regarding what types of knowledge and skills the test assesses and what are the standards for performance.

Massachusetts educators value the release of items. Of all of the Department's reports that include MCAS data, the "MCAS Student Item Analysis Roster" was downloaded 142,057 times between July 2014 and July 2015. It was the third most popular MCAS report. In addition, the report "MCAS School Test Item Analysis Summary" was downloaded 47,315 times and was the 11th most popular report downloaded. "MCAS Classroom Item Analysis Roster" was downloaded 28,954 times during last year.

Choices about the number and share of items released involve trade-offs. There are cost implications, because if items are released into the public domain, they cannot be used on future tests. Thus, the more items that are released, the more new test items that must be developed and go through the entire vetting process, which involves significant time and expense.

PARCC is taking a different approach to item release. Currently, PARCC is planning to release the equivalent of a full test. It will not be an actual test but a composite form of common items from multiple forms of the test administered in each grade. MCAS is a single-blueprint or test, while PARCC develops multiple forms from several underlying blueprints and does not have the same set of items across the multiple forms administered to students. Within the same classroom, students assessed on different forms of PARCC might share few common items across all forms. All forms are equated through a linking process so that the scoring is consistent across forms, but most of the test items will be different across forms.

Because of the multiple forms in PARCC, there are many more overall test items in a given grade. In 2015, PARCC will release 25 – 35% of test items per grade, although it will be less in some grades. This translates into an average of 40 to 50 items per grade, although again it will likely be fewer items in some grades. PARCC will also be releasing samples of student work that illustrate each score level.

In future years, PARCC plans to release an average of 45 items per grade per year, which will represent between 30 – 50% of the items per grade. Thus, even though PARCC plans to release a lower share of the test items, it will be releasing a larger number of test items.

This will represent a change in practice around items release in Massachusetts and will have consequences for how test results are reported to families and educators. The composite form released by PARCC may not include many actual questions that any individual student answered. Unlike with MCAS, students and their families will not be able to see the specific items that were on their specific test and then examine the corresponding exemplary answers to the specific items.   Educators will

similarly be less able to analyze the specific, individual items that their students answered.

Although this would be a change from current practice and expectations and would decrease the level of transparency around the assessment, the consequences may also open new possibilities and opportunities. It could reinforce a focus on the underlying content standards and diminish the focus on specific questions or types of questions. PARCC intends to report how students fared in relationship to each standard and provide a range of sample questions that illustrate how each standard was assessed. This approach could emphasize the content standards, shifting the focus from "which questions" did my students have trouble with to "which standards" did my students have trouble with and what various types of questions could help me assess whether the student is continuing to struggle with the standard? If there are adequate examples to demonstrate what is expected for each standard, this approach could help teachers improve their practice, even without a more detailed item analysis. This is all somewhat speculative at this point, because we do not currently have a comprehensive understanding of how teachers have used item release in the past. Nor do we know what types of PARCC teacher reports will ultimately be developed and made available to educators.

Types of Reports

The types of reports developed and made available are integral to the impact of an assessment system. There are multiple audiences for the reports. Classroom teachers and other educators rely on the reports to inform their practice. Families use the reports as a way to understand their own child's level of achievement and relative standing among their peers. Policy-makers and members of the broader public use reports as a window into understanding what is happening in K-12 education, including which schools, districts, or programs are most effectively supporting student learning.

Over the years, many reports that incorporate MCAS data have been developed. Developing and fine-tuning these reports has taken many years, perhaps even a decade by some estimates. The reports are available at the district, school, and student level. The reports includes some that are publicly available, some that are accessed by local educators and state education officials through Edwin Analytics, and some that are sent directly to individual families.

Governance, technical, and economical challenges all make it nearly impossible to simply state that ESE will use PARCC data to produce the same reports as they did under MCAS. Because the PARCC assessment system is based on multiple forms, there are additional layers of complexity to understanding the opportunities for, and cost of, various types of reporting to different stakeholders. While these complexities may prove to open new possibilities and opportunities, at this time, before any of the actual data has been aggregated, analyzed, and reported upon by the Department, it is enormously difficult to predict what a reporting regime will

look like under PARCC and what the state-specific costs and challenges are likely to be.

At this time, the nature of the technology platform ultimately to be used by PARCC to report assessment results is not yet settled. The PARCC consortium is moving towards developing a common data analysis and reporting platform for use throughout the consortium but its functionality and usability is not yet known. The consortium began the process of procuring a more extensive data platform for the generation and dissemination of reports but those efforts did not result in procuring a vendor and current plans now specify a more limited platform.

Because these efforts will not be completed in time to report on the spring 2015 PARCC results, the Department has begun to use its internal technical capacity from the Edwin Analytics system to fill the void for Massachusetts educators and families. Determining how this critical component of any assessment system will be fulfilled going forward under a PARCC system, however, will be an important matter for consideration by the Board. How these matters are settled will have significant educational, budgetary, and public confidence consequences going forward.

VI. Other Policy Considerations

There are many considerations for the Board as it decides which ELA and math assessments will best advance the educational goals of the Commonwealth as well as the educational attainment and future life prospects of Massachusetts students. The existing research base on the use and purpose of assessments and descriptions of various aspects of the two assessment systems - the main focus of this report – are only part of those considerations. Beyond the scope of this report, there are other critical considerations for the Board. We conclude this report by briefly discuss two important policy issues: 1) Governance and 2) Cost.

Governance
With regard to the issue of governance, there are clear differences between the two assessments. MCAS is a custom-designed assessment, which the Commonwealth owns in its entirety. In contrast, the Partnership for Assessment of Readiness for College and Careers (PARCC) is a multi-state consortium, where Massachusetts serves in a leadership role on the Governing Committee but is still only one member of the consortium. A number of consequences follow from being a member of a consortium.

Being a member of the consortium offers Massachusetts some important advantages and opportunities. The cost of developing PARCC was heavily subsidized by the federal government, with a grant of approximately $170 million as part of the Race

to the Top Assessment Program. In addition, there are potential economic efficiencies gained by states pooling resources. According to Commissioner Chester's Board report in spring of 2014, "While fiscal constraints forced Massachusetts to abandon our independent effort to modernize the MCAS in 2009 (including computerized administration opportunities), the PARCC consortium holds out the promise of being able to deliver a significant upgrade of the system within the existing cost structure of our legacy MCAS."[61]   The existence of a common assessment across multiple states could also help create a larger market for aligned instructional materials that could help elevate the quality of such products and help reduce their average cost, as well. In addition, the collaboration allows access to a broader set of experts and experienced educators in test design and development.

Using a common assessment across multiple states also makes state-to-state comparisons easier and more transparent. Currently, the most common way to compare student achievement across states uses NAEP. PARCC will potentially augment the more limited comparisons that are now made with the NAEP in grades 4, 8, and 12. A common assessment will allow comparisons of different sets of items within the assessment and will also facilitate more direct comparisons of sub-groups for states that are using the same assessment.   The current members of the PARCC consortium are 7 states, including Massachusetts, and the District of Columbia.

At the same time, there are also clear constraints to being part of a consortium. As in any collaboration, Massachusetts must work with the other members to make many decisions that will impact the quality, rigor, and utility of the assessment. If the "push and pull" of the collaboration leads to better-informed decisions, this is an advantage. If, on the other hand, that process results in decisions that change or compromise elements of the state's assessment system that have made Massachusetts a national leader in standards-based education, the state will run the risk of weakening its standards and may be faced with a costly decision to discontinue its relationship with the consortium. Massachusetts also faces the potential risk of needing to reestablish its own assessment system if the consortium dissolves or disbands itself.

Massachusetts has a long track record of developing and administering high-quality assessments. That experience and expertise has helped inform the early development of the new PARCC system. The state's experience with MCAS has also established certain expectations with regards to the assessment system among educators, students, and the general public here in the Commonwealth. These expectations extend beyond the general quality and rigor of the test items and have implications for how information from the assessments is utilized and the level of

---

[61] Chester, M. (2014). *Building on 20 Years of Massachusetts Education Reform*. Report prepared for the Massachusetts Board of Elementary and Secondary Education, November, p. 14. Retrieved from: http://www.doe.mass.edu/commissioner/default.html

transparency that is expected of the system. All of these experiences and consequent expectations have bearing on the level of public support for the education accountability system as a whole.

One example involves the release of test items, which has been discussed in previous parts of the report. The release of test items has become an expected practice and is highly valued by Massachusetts educators. Because of the cost implications of releasing items, PARCC is currently planning to release a lower percentage of items in each grade, but because of the multiple forms, it will result in a larger number of items released. However, the items will be released as a composite form, with items drawn from the various PARCC test forms. Therefore, while some students will have answered some questions that appear on the released form, the test form that is released will be one that no actual student has taken. Partial release of test items will mean that it will not be possible to view question-by-question results of any individual student or of a classroom of students. This difference will represent a change to the existing policies and practice of how information ascertained through assessments is used by educators and parents in Massachusetts going forward. The type of items and reports currently and potentially available under each of the two assessment systems is discussed in more detail in Section V of this report on pages 46 and 60, respectively.

Another challenge arising from the multi-state nature of the consortium relates to how to appropriately assess the Massachusetts-specific standards. As already discussed, the Massachusetts Curriculum Frameworks include supplemental standards beyond the Common Core, particularly in math. An important question for the Board is whether it is a priority to assess these additional Massachusetts standards, and if so, how?   The Department of Elementary and Secondary Education provided information indicating that one of the additional ELA standards and all nine of the additional math standards in grades 3-8 are currently assessed as part of MCAS.[62]   Currently, it is not an option for these state-specific standards to be assessed under PARCC. If Massachusetts wanted to assess additional standards, it would have to be through a different contract with a vendor.

Going forward, being a member of a consortium also has implications for future changes to the Massachusetts Curriculum Frameworks. As discussed, the Massachusetts standards have evolved over the past two decades, and it is reasonable to anticipate that they will continue to evolve in the future. As standards evolve, so too must assessments. Revisions to standards (and assessments) are complicated endeavors. As Commissioner Chester explains in his 2014 Board Report, "As would be expected, and as is appropriate given the centrality and importance of frameworks in a standards-based system, each adoption and revision of standards is characterized by vigorous debate and some measure of

---

[62] Information provided as part of interview with Department of Elementary and Secondary Education staff, July 2015.

controversy."[63] Will the Massachusetts Curriculum Frameworks and the Common Core evolve in similar ways? Will the collaboration and joint development of underlying standards continue in the future?

If the Board selects PARCC, the stakes in any future modifications of standards, particularly at the state level, become much higher. If Massachusetts chooses to diverge from the Common Core to any significant degree it would necessitate an abandonment of the PARCC system unless the consortium chose to diverge from the standards in a consistent fashion and align the assessment accordingly. Therefore, such a divergence from Common Core could easily leave Massachusetts in the position of needing to develop an entirely new assessment system on its own to maintain alignment with the new standards. Conversely, if Massachusetts chooses to continue developing and maintaining its own assessment, it would be able to modify the assessment to the degree the standards themselves are modified. It would, however, have to bear the costs of any such modifications alone without the efficiency benefits of a consortium.

Even this brief summary of the issues arising from questions of governance suggests that the Board place a particular focus on these and other questions that arise from the nature of a consortium as the governance structure for the PARCC assessment system.

Cost
The issue of the cost will also likely be an important consideration for the Board, but a thorough and comprehensive cost comparison is beyond the scope of this report. Cost can be broadly divided into administration costs and costs associated with test development. The following brief discussion again offers a few illustrative examples of the complexities that make an apples-to-apples comparison of the two assessment systems difficult at this time. It does not purport to be a comprehensive listing of the various issues involved.

Comparing the MCAS per-pupil cost with the PARCC per-pupil cost is not sufficient for a number of reasons. There are differences regarding what is included in the per-pupil costs in each contract. For instance, in MCAS, all open responses, short responses, short-answer items, and writing prompts are hand scored, and the cost for hand scoring is included in the fixed-cost contract with the vendor. In contrast, with PARCC, Massachusetts has chosen to pay an additional cost to have the short answer and prose-constructed items hand scored. In addition, there also are important differences between the assessments regarding how the data will be gathered, aggregated, and reported, related to both development and administration, which have cost implications. There are also the technical challenges of reporting under PARCC and the reporting platform is still being developed.   The Department of Elementary and Secondary Education is currently in

---

[63] Chester (2014), p. 8.

the process of building reports for PARCC data files using Edwin Analytics. These differences and the unknown nature of their potential cost implications are discussed elsewhere in this report (see pages 58 and 60 under "Release of Items" and "Types of Reports).

Not surprisingly, there is an upfront transition cost for PARCC, which would be expected with any new assessment. One such transition cost would arise from the need to change existing analytic tools and reports that incorporate or rely on student assessment data to reflect the new scoring and performance standards under the new assessment system.

Beyond transition issues such as this, however, there are other obstacles to easily comparing the costs that arise from differences in the form and administration needs of the two systems themselves. For example, PARCC is intended to be a computer-based assessment. Given the existing limited technology capacity of some schools in the Commonwealth and in some other areas across the country, it was not feasible for the assessment initially to be solely computer-based, so PARCC created a paper-based assessment. The paper-based PARCC will be offered through 2018, but it is not clear what will happen after that date when the initial contract with Pearson ends. While this accommodation of paper administrations helped avoid the need for an immediate, substantial investment in the technological capacity of schools and districts of the Commonwealth, it may have also had implications regarding the test items (discussed in more detail on page 46).

In Massachusetts, of the 54% of districts that administered PARCC in grades 3-8: 31% of the districts used a paper test; 50% used a computer-based test; and 19% of districts used a mix of the two (Table 6). According to information provided by the Department, among the states that administered PARCC this year, Massachusetts had the lowest percentage of students who took PARCC on a computer during the recent test administration. [64] The Department's 2014 analysis found that 24% of schools (388 schools) in Massachusetts are not ready to administer PARCC online or are ready for digital learning.[65] Presumably, this number has gone down since 2014 as more schools ready themselves for online tests, but no updated information was available.

---

[64] Information provided during interview with staff from Department of Elementary and Secondary Education, July 2015.
[65] Massachusetts Board of Elementary and Secondary Education. (2015, September 21). Technology Readiness Update. [PowerPoint presentation].

Table 6: Paper-Based and Computer-Based PARCC Tests, 2015

	Number of Districts Administering PARCC	PARCC Paper-Based Tests		PARCC Computer-Based Tests		PARCC Mix of Paper and Computer	
		No. of districts	%	No. of districts	%	No. of districts	%
Grades 3-8	194	61	31%	97	50%	36	19%
Grade 9 and/or 11 (optional)	69	19	27%	44	64%	6	9%

Source: Massachusetts Department of Elementary and Secondary Education, retrieved from: **http://www.doe.mass.edu/news/news.aspx?id=13541**

 As part of the transition costs to PARCC, it seems reasonable that the Board should consider the likelihood and potential cost of supporting schools and districts in developing their technological capacity to be able to administer the computer-based assessment. Of course, if the Commonwealth were to forego administering PARCC, an updated MCAS may also involve computer-based assessments. Given the Board's reliance on other state entities such as the Governor and Legislature to secure funding for such initiatives, the Board will need to consider the relative importance and prioritization of these matters carefully. To the extent that these technology upgrades are already a priority for the Board, PARCC might help to create the impetus for improving the technology capacity within schools.

Another significant issue with cost implications that the Board must consider is the expected scale of the PARCC consortium going forward. The initial planning for PARCC began with a consortium that had many more participating members than currently belong. The loss of these members has had to have significant impact on the ability of the consortium to deliver and maintain the assessment system it envisioned a year or two ago. The end of generous federal support for the initial development costs of PARCC and the loss of membership comes at a critical time for the PARCC consortium. Based on the information returned from the 2015 administration of the assessment, the consortium will likely need to make significant investments and critical policy decisions relative to the future of the PARCC assessment system, such as future performance standard setting and item development. The Board should consider carefully the cost and quality implications of these challenges facing the consortium as it weighs its various options before it.

The Board and other state policy leaders are already familiar with the costs of administering the MCAS as it currently exists. As the Board works to analyze the relative costs of the two systems, however, the Board must also consider whether they regard the current form of MCAS as adequate to allow the Commonwealth to advance its educational goals. As the Board considers which assessment best meets the present and future needs of the Commonwealth, it might consider what improvements or adaptations to MCAS are currently needed and likely necessary in

coming years, and what the accompanying costs are likely to be to implement those changes.   Considering revisions to the current MCAS obviously entails a number of uncertainties, relative to costs, for the Board to evaluate.

The relative uncertainties related to costs are not confined to a future under PARCC. Rather, those uncertainties also exist under a scenario of an updated and improved MCAS, as well. Either path will require significant financial resources, the dedication of significant staff time at the Department of Elementary and Secondary Education, and the attention and vigilance of the Board to ensure that the state's assessment system will advance the Commonwealth's educational goals.

References

American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education (AERA/APA/NCME). (2014). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.

Bandeira de Mello, V., Bohrnstedt, G., Blankenship, C., and Sherman, D. (2015). *Mapping State Proficiency Standards Onto NAEP Scales: Results From the 2013 NAEP Reading and Mathematics Assessments (NCES 2015-046)*. U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch.

Beatty, A. (2010). *Best Practices for State Assessment Systems, Part 1: Summary of a Workshop*, Washington, D.C.: National Academy of Sciences.

Chester, M. (2014). *Building on 20 Years of Massachusetts Education Reform. Report prepared for the Massachusetts Board of Elementary and Secondary Education*, November, p. 14. Retrieved from: http://www.doe.mass.edu/commissioner/default.html

Common Core or Something Else? A Map of State Academic Standards (2015, July 20). *Education Week*, Vol. 34, Issue 36. Retrieved from: http://www.edweek.org/ew/section/multimedia/map-states-academic-standards-common-core-or.html

Darling-Hammond, L., Herman, J., Pellegrino, J., et al. (2013). *Criteria for high-quality assessment*. Stanford, CA: Stanford Center for Opportunity Policy in Education. Retrieved from: https://edpolicy.stanford.edu/sites/default/files/publications/criteria-higher-quality-assessment_2.pdf

Educational Testing Service. *Technology-enhanced Assessments*. Retrieved from: https://www.ets.org/c/22116/

Faxon-Mills, S., Hamilton, L. S., Rudnick, M., & Stetcher, B. M. (2013) *New Assessments, Better Instruction? Designing Assessment Systems to Promote Instructional Improvement*. Santa Monica, Calif.: RAND Corporation. Retrieved from: http://www.rand.org/pubs/research_reports/RR354

Gewertz, C. (2015, February 4). A Map of States' 2015 Testing Plans: The Dust Has Finally Settled. *Education Week Curriculum Matters blog*. Retrieved October 5, 2015 from: http://blogs.edweek.org/edweek/curriculum/2015/02/a_map_of_states_2015_testing_p.html

Hollingworth, L. Beard, J.J. & Proctor, T.P. (2007). An Investigation of Item Type in a Standards-Based Assessment *Practical Assessment, Research & Evaluation*, . Vol 12, No 18.

Livingston, S.A. (2009). *Constructed-Response Test Questions: Why We Use Them; How We Score Them*. Educational Testing Service (ETS). Retrieved from: https://www.ets.org/Media/Research/pdf/RD_Connections11.pdf

Marzano, R. J., Pickering, D, McTighe, J, (1993). *Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model*. Aurora, CO: Mid-Continent Regional Educational Lab (McREL).

Massachusetts Business Alliance for Education and Center for Assessment. (2015). *Educating Students for Success: A Comparison of the MCAS and PARCC Assessments as Indicators of College- and Career-Readiness*, Boston, MA: Massachusetts Business Alliance for Education.

Massachusetts Board of Elementary and Secondary Education. (2015, September 21). Technology Readiness Update. [PowerPoint presentation].

Massachusetts Department of Elementary and Secondary Education. (2013). *Massachusetts 4th and 8th Graders Lead the Nation in Reading and Mathematics Performance for the Fifth Consecutive Time* [Press release]. Retrieved from: http://www.doe.mass.edu/news/news.aspx?id=7846

Massachusetts Department of Elementary and Secondary Education and Massachusetts Board of Higher Education. (2015). Massachusetts Definition of College and Career Readiness. Retrieved from: http://www.mass.edu/library/documents/2103College&CareerReadinessDefinition.pdf

Massachusetts Department of Elementary and Secondary Education. (n.d.) 10 Frequently Asked Questions about the Common Core State Standards and PARCC. Retrieved (October 5, 2015) from: http://www.does.mass.edu/parcc/

Massachusetts Department of Elementary and Secondary Education. (2011). *Massachusetts Curriculum Frameworks for English Language Arts and Literacy: Grades Pre-Kindergarten to 12: Incorporating the Common Core State Standards for English Language Arts and Literacy in History/Social Studies, Science, and Technical Studies*. Retrieved from: http://www.doe.mass.edu/frameworks/ela/0311.pdf

Massachusetts Department of Elementary and Secondary Education. (2011). Massachusetts Curriculum Frameworks for Mathematics, Grades Pre-Kindergarten to 12: Incorporating the Common Core State Standards for Mathematics. Retrieved from: http://www.doe.mass.edu/frameworks/math/0311.pdf

Massachusetts Department of Elementary and Secondary Education. (2013). 2013 MCAS and MCAS-Alt Technical Report. Retrieved from: http://www.mcasservicecenter.com/documents/MA/TechnicalReport/TechReport_2013.htm

McClarty, K., Korbin, J., Moyer, E. L. et al., (2015). Draft: PARCC Benchmarking Study.

Washington, DC: Partnership for Assessment of Readiness for College and Careers.

Moyer, E. L., Miles, J. Davis, L. L. & You, W. (2015). *Postsecondary Educators' Judgment Study: Executive Summary*. Washington, DC: Partnership for Assessment of Readiness for College and Careers.

National Center for Education Statistics. (n.d.) NAEP Overview. Retrieved from http://nces.ed.gov/nationsreportcard/dba/

PARCC Glossary of Terms. Retrieved October 5, 2015 from: http://www.parcconline.org/resources/parent-resources/glossary-of-terms

Policy Analysis for California Education and Rennie Center for Education Research & Policy. (2011). *The Road Ahead for State Assessments*. Cambridge, MA: Rennie Center for Education Research & Policy. Retrieved from: http://renniecenter.issuelab.org/resource/road_ahead_for_state_assessments

Rennie Center for Education Research & Policy (2015). *Testing the Test: A Study of PARCC Field Trials in Two School Districts*. Boston, MA: Rennie Center for Education Research & Policy. Retrieved from: http://renniecenter.org/topics/PARCC_case_study.html.

Ronan, A. (2015, February 7). Seven Technology Enhanced Item Types You'll See on Common Core Tests This Spring. *Edudemic*. Retrieved from: http://www.edudemic.com/7-tech-enhanced-items-on-common-core-tests/

Russell, M. (2011). *Digital Test Delivery: Empowering Accessible Test Design to Increase Test Validity for All Students*. Seattle, WA: Bill and Melinda Gates Foundation.

Thacker, A., Dickinson, E., Bynum, B. et al. (2015). Findings from the Quality of Items/Tasks/Stimuli Investigations: PARCC Field Tests. Washington, DC: Partnership for Assessment of Readiness for College and Careers.

Yuan, K. and Le, V. (2012). *Estimating the Percentage of Students Who Were Tested on Cognitively Demanding Items Through the State Achievement Tests*. RAND Working Paper. Santa Monica, CA: RAND Corporation. WR-967-WFHF. Retrieved from: http://www.rand.org/content/dam/rand/pubs/working_paper/2012/RAND-WR967.pdf