

HOW PARCC'S FALSE RIGOR STUNTS THE ACADEMIC GROWTH OF ALL STUDENTS

by Mark McQuillan, Richard P. Phelps, and Sandra Stotsky



PIONEER INSTITUTE
PUBLIC POLICY RESEARCH

White Paper No. 135
October 2015

PIONEER'S MISSION

Pioneer Institute is an independent, non-partisan, privately funded research organization that seeks to improve the quality of life in Massachusetts through civic discourse and intellectually rigorous, data-driven public policy solutions based on free market principles, individual liberty and responsibility, and the ideal of effective, limited and accountable government.



This paper is a publication of the Center for School Reform, which seeks to increase the education options available to parents and students, drive system-wide reform, and ensure accountability in public education. The Center's work builds on Pioneer's legacy as a recognized leader in the charter public school movement, and as a champion of greater academic rigor in Massachusetts' elementary and secondary schools. Current initiatives promote choice and competition, school-based management, and enhanced academic performance in public schools.



The Center for Better Government seeks limited, accountable government by promoting competitive delivery of public services, elimination of unnecessary regulation, and a focus on core government functions. Current initiatives promote reform of how the state builds, manages, repairs and finances its transportation assets as well as public employee benefit reform.



The Center for Economic Opportunity seeks to keep Massachusetts competitive by promoting a healthy business climate, transparent regulation, small business creation in urban areas and sound environmental and development policy. Current initiatives promote market reforms to increase the supply of affordable housing, reduce the cost of doing business, and revitalize urban areas.



The Center for Health Care Solutions seeks to refocus the Massachusetts conversation about health care costs away from government-imposed interventions, toward market-based reforms. Current initiatives include driving public discourse on Medicaid; presenting a strong consumer perspective as the state considers a dramatic overhaul of the health care payment process; and supporting thoughtful tort reforms.

TABLE OF CONTENTS

Executive Summary	5
1. Purpose of Paper	7
2. International Comparisons	10
3. Differences between pre-Common Core MCAS tests and 2015 PARCC Practice Tests	15
4. College Readiness in MCAS and PARCC	21
5. How Vocabulary Knowledge is Assessed across Tests	30
6. How Writing is Addressed across Tests	37
7. Why PARCC Does Not Prepare Bay State Students for an Academically Meaningful High School Diploma or College	44
Recommendations	49
About the Authors	51
Endnotes	52
Appendix A. : Critique of Criteria for Evaluating Common Core-Aligned Assessments	62
Appendix B. Links to Public Hearings and Other Sources of Public Comment on MCAS or PARCC	69
Appendix C. A Randomly-Chosen Test Item Used on the 2014 Grade 10 MCAS Math Test	71

EXECUTIVE SUMMARY

In July 2010, the Massachusetts Board of Elementary and Secondary Education (BESE) voted to adopt Common Core's standards in English language arts (ELA) and mathematics in place of the state's own standards in these two subjects. The vote was based largely on recommendations by Commissioner of Education Mitchell Chester and then Secretary of Education Paul Reville, and on the conclusions in three studies comparing the state's standards with Common Core's, all financed directly or indirectly by the Bill & Melinda Gates Foundation, and all issued by organizations that are among the primary boosters of Common Core (Achieve, Inc., Thomas B. Fordham Institute, and Massachusetts Business Alliance for Education).

Nevertheless, annual state testing for school and district accountability continued as part of the Massachusetts Comprehensive Assessment System (MCAS) mandated by the 1993 Massachusetts Education Reform Act (MERA). To accommodate the adoption of Common Core's standards, tests were based on both the old standards and an annually increasing number of Common Core's standards until 2015, when all of the pre-Common Core standards in ELA and mathematics were archived, and the MCAS tests were presumably only Common Core-based.

After the vote to adopt Common Core's standards in 2010, the state joined the testing consortium called Partnership for Assessment of Readiness for College and Careers (PARCC), funded by the United States Department of Education (USED) to develop common tests for its member states (about 25 initially), but with the costs for administering the tests to be borne by the states and local school districts. Since 2011, PARCC has been developing tests that BESE is expected to vote to adopt in the late fall of 2015 as the state's official Common Core-based tests in place of Common Core-based MCAS tests. (Indeed, the commissioner of education and his staff at the Department of

Elementary and Secondary Education (DESE) have been implementing a transition to PARCC tests for several years.) BESE's official vote will be guided, again, by the recommendations of the same commissioner of education (who now also chairs PARCC's Governing Board), the current Secretary of Education James Peysers, and the conclusions reached in "external" studies comparing PARCC and MCAS tests as well as in about 20 studies directly authorized by PARCC.

Two of the external studies are listed in the state's 2015 application to the USED for a waiver from No Child Left Behind Act requirements and are by organizations that had originally recommended adoption of Common Core. One, issued by the Massachusetts Business Alliance for Education in February 2015, claims that PARCC tests predict college readiness better than MCAS tests did. A second, to be completed by the Fordham Institute and a partner, is to be issued in time for BESE's vote. A third, issued in mid-October 2015 by Mathematica Policy Research (and requested by the state's Executive Office of Education) claims both tests are equally predictive of college readiness, although its report has major shortcomings.

This White Paper will be a fourth external report on the question BESE's vote will address; it is motivated by our interest in providing an analysis of how MCAS and PARCC assess reading and writing. Much less national attention has been paid to Common Core-based assessments of reading and writing than of mathematics, yet reading and writing skills are just as important to readiness for college and career as is mathematics.

At the order of Governor Charles Baker, BESE held five public hearings across the state in 2015 to enable the public to testify on whether it wants BESE to adopt the Common Core-based PARCC tests as the state's official tests. The purpose for the hearings remains unclear; over two years ago, the commissioner of education told local superintendents that the state would be transitioning to PARCC anyway.

If BESE officially votes to adopt PARCC as the state's testing system, it will automatically abandon the use of Common Core-based MCAS tests for K-12. (It is not clear if non-Common Core-based MCAS tests, such as those in science and history, would be prohibited as well.) It would also tie the state to joint decisions by the member states in the PARCC consortium (fewer than 10 at this writing) and to policies established by USED for new Elementary and Secondary Education Act (ESEA) grants to the states.

Congress rewrote ESEA in the summer of 2015, putting control of a state's standards and tests, which are required for receipt of ESEA funds, in the hands of state commissioners, boards, and staffs of education, with no approval required by state legislatures, higher education teaching faculty in the arts and sciences, local school boards, or parents. (A reconciliation bill remains to be approved by Congress and signed by the president.) Approval of a state's standards and tests is to be granted by USED based on the recommendations of those whom it chooses to review applications. In other words, federal control will remain intact, simply more indirect and hidden.

In a comparison of Common Core-based PARCC tests and pre-Common Core MCAS tests, this study identifies six major flaws in PARCC tests:

1. Most PARCC writing prompts do not elicit the kind of writing done in college or the real world of work.
2. PARCC uses a format for assessing word knowledge that is almost completely unsupported by research and seriously misleads teachers.
3. PARCC's computerized testing system has not shown more effectiveness than a paper-and-pencil-based testing system or a return of useful information to the teachers of the students who took PARCC tests.
4. PARCC uses "innovative" item-types for which no evidence exists to support claims

that they tap deeper thinking and reasoning in understanding a text.

5. PARCC tests require too many instructional hours to administer and prepare for. They also do not give enough information back to teachers or schools to justify the extra hours and costs
6. PARCC test-items do not use student-friendly language and its ELA reading selections do not look as if they were selected by secondary English teachers.

Central Recommendation

This White Paper's central recommendation is that Massachusetts use a testing system for K-12 that is much less costly, more rigorous academically, and much more informative about individual student performance, and with much less instructional time spent on test preparation and administration, than the current PARCC tests. Both the PARCC tests and the current MCAS tests in grade 10 are weak, albeit for different reasons, and neither indicates eligibility for a high school diploma, college readiness, or career readiness. In essence, the authors recommend that BESE reject the PARCC assessment system and vote for the MCAS system but on the condition that the responsibility for developing and administering K-12 standards and tests be assigned to an organization in Massachusetts independent of DESE and the state's education schools. This organization must focus squarely on providing the best possible content standards from disciplinary experts in the arts, sciences, and engineering throughout the state and be capable of providing oversight of high school standards and tests. If carried out, these recommendations would ensure the legacy and future promise of MERA.

1. PURPOSE OF PAPER

CONTEXT AND BACKGROUND

In July 2010, the Massachusetts Board of Elementary and Secondary Education (BESE) voted to adopt Common Core's standards in English language arts (ELA) and mathematics in place of the state's own standards in these two subjects. The vote was based largely on recommendations by Commissioner of Education Mitchell Chester and then Secretary of Education Paul Reville, and on the conclusions in three studies comparing the state's standards with Common Core's, all financed directly or indirectly by the Bill & Melinda Gates Foundation, and all issued by organizations that are among the primary boosters of Common Core (the Thomas B. Fordham Institute, Achieve, Inc., and Massachusetts Business Alliance for Education).

Nevertheless, state testing for school and district accountability continued as part of the Massachusetts Comprehensive Assessment System (MCAS) mandated by the 1993 Massachusetts Education Reform Act (MERA). To accommodate the adoption of Common Core's standards, tests were based on both the old standards and an annually increasing number of Common Core's standards until 2015, when all of the pre-Common Core standards in ELA and mathematics were archived, and the MCAS tests were presumably only Common Core-based.¹

After the vote to adopt Common Core's standards in 2010, the state joined the testing consortium called Partnership for Assessment of Readiness for College and Careers (PARCC), funded by the United States Department of Education (USED) to develop common tests for its member states (about 25 initially), but with the costs for administering the tests to be borne by the states and local school districts. The other testing consortium funded by the USED at the same time was Smarter Balanced Assessment Consortium (SBAC), with over 25 member states initially. Since 2011, PARCC has been developing tests that BESE is expected to vote

to adopt in the late fall of 2015 as the state's official Common Core-based tests in place of Common Core-based MCAS tests. BESE's vote will be guided, again, by the recommendations of Commissioner of Education Mitchell Chester (who now also chairs PARCC's Governing Board) and current Secretary of Education James Peyser, and by the conclusions of "external" studies comparing PARCC and MCAS tests as well as over 20 studies most of which were directly authorized by PARCC. These studies are listed in DESE's revision of its Elementary and Secondary Education Act (ESEA) Flexibility Request dated June 15, 2015. Of these over 20 studies, only six were completed as of that date; many are still works in progress.

Two of the external studies are listed in the waiver application and are by organizations that had originally recommended adoption of Common Core. One, issued by the Massachusetts Business Alliance for Education (MBAE) in February 2015, claims that PARCC tests predict college readiness better than MCAS tests did.² A second, to be completed by the Thomas B. Fordham Institute and a partner, is to be issued in time for BESE's vote. A third, issued in mid-October 2015 by Mathematica Policy Research (and requested by the state's Executive Office of Education) claims both tests are equally predictive of college readiness, although its report has major shortcomings.³

This White Paper will be a fourth external report on the question BESE's vote will address; it is motivated by our interest in providing academic analyses of how MCAS and PARCC assess reading and writing. Much less national attention has been paid to Common Core-based assessments of reading and writing than of mathematics, yet reading and writing skills are just as important to readiness for college and career as is mathematics.

At the order of Governor Charles Baker, BESE held five public hearings across the state in 2015 to enable the public to testify on whether it wants BESE to adopt the Common Core-based

PARCC tests as the state's official tests (see Appendix B for links to those hearings), although the commissioner had, two years earlier, already told all local superintendents the state would be transitioning to PARCC anyway. If BESE officially votes to adopt PARCC as the state's testing system, it will automatically abandon the use of Common Core-based MCAS tests for K-12. However, several important wrinkles remain to be straightened out.

First, although MCAS tests are in 2015 based only on Common Core's standards, MERA requires all students to pass state tests in English language arts, mathematics, science and technology, history/social science, and three other subjects in the school curriculum in grade 10 (or tests and retests based on grade 10 standards in these subjects) for a high school diploma. The statute does not specify grade 11 for these tests, and passing them does not entitle students to by-pass college placement tests and take only credit-bearing college coursework in their freshman year, as will passing PARCC's college readiness test in grade 11 entitle students to do in states whose institutions have agreed to this entitlement. A vote by BESE to use the PARCC tests for grades 3-11 as the state's official tests does not change the statute requiring students to pass state tests and retests based on grade 10 standards for a high school diploma. (Nor does it change earlier BESE decisions to award scholarships to high-scoring students on grade 10 MCAS tests if they stay in school through grade 12.) A change in the statute will require a vote by the legislature.

Second, MERA requires, as mentioned above, all students to pass state tests in seven school subjects; BESE has already voted to require students to pass state tests by a specific date in four of those subjects for a high school diploma. But PARCC provides tests only for mathematics and what it calls English language arts/literacy. So it is unclear how BESE can vote to replace MCAS tests with PARCC's when PARCC tests do not address the content of all four subject tests that BESE has so far voted to require for a high

school diploma. Nor does PARCC indicate any plans to provide tests for the other five subject areas required by MERA for a high school diploma. At most, BESE can vote to replace the MCAS tests in ELA and math with PARCC tests in math and ELA/literacy. But it has no PARCC tests to replace the current MCAS tests in science and history/social science (the grade 10 history test scheduled in 2009 was shelved indefinitely).

However, a vote by BESE in the fall of 2015 to make PARCC the state's official tests for grades 3 to 11 would tie the state to joint decisions by the small number of member states now left in the PARCC consortium (fewer than 10 at this writing) and to policies established by the USED for new ESEA grants to the states.

Congress rewrote ESEA in the summer of 2015, putting control of a state's standards and tests, which are required for receipt of ESEA funds, in the hands of state commissioners, boards, and staffs of education, with no approval required by state legislatures, higher education teaching faculty in the arts and sciences, local school boards, or parents. Reconciliation of the re-authorization bills passed separately by the House and Senate, plus the president's signature, still remains. Approval of a state's standards and tests will be granted by the USED based on the recommendations of those whom it chooses to review applications. In other words, federal control will remain intact, simply more indirect and hidden.

ORGANIZATION AND CONTENTS OF THIS PAPER

The purpose of this paper is to provide a comparison of Common Core-based PARCC tests and pre-Common Core MCAS tests along several critical dimensions: how well they address their respective goals so far as we can tell; how they model for teachers the pedagogy for teaching the basic element in reading comprehension; and how they choose to assess growth in writing skills. In Chapter 2, we compare large-scale assessment systems for K-12 across countries. We describe major

differences between national testing systems for high-achieving countries, pre-Common Core state testing systems, and Common Core-based testing systems. In Chapter 3, we explore specific differences between pre-Common Core MCAS and PARCC tests. These chapters constitute the foundation and background for the analyses that follow. In Chapter 4, we discuss the meaning of college readiness in MCAS and PARCC in both math and ELA in the Bay State. In Chapter 5, we describe the treatment of vocabulary knowledge across testing systems. We explain why its relationship to reading comprehension and the clarity of written expression makes it crucial for readers to understand vocabulary assessment in the context of the research on the nature of vocabulary acquisition and effective pedagogical approaches. In Chapter 6, we describe how writing is addressed through the grades in used MCAS test items and in PARCC practice test items. Finally, in Chapter 7, we summarize earlier chapters and offer recommendations on ways to make state tests more useful, acceptable, and transparent to the state's parents, K-12 teachers, and post-secondary teaching faculty.

The appendices of this paper provide additional evidence for the criticisms we offer in this paper. In Appendix A, we describe the criteria Fordham and its partner are using to assess both PARCC and MCAS despite the fact that these criteria were developed in 2014 by the Council for Chief State School Officers (CCSSO) for assessing testing systems based only on college and career readiness standards. In Appendix B, we provide links to the recent public hearings on PARCC or MCAS testing and to other sources of public comment in the Bay State today towards Common Core and PARCC. In Appendix C, we provide an example of a below-grade-level test item used in the 2014 MCAS math test for grade 10, one of the many used in recent years.

PRIMARY FOCUS OF THIS PAPER

The discussion in this White Paper of the content of the MCAS test items for both reading and writing used in the Bay State from 1998 to 2007

is a major distinction between this report and the report to be released in fall 2015 by the Fordham Institute and its partner. We focus on test items in reading and writing, not mathematics, because little attention has been paid to these two areas of the curriculum, although we do address other issues in assessing a mathematics curriculum.

Moreover, because the researchers hired by Fordham for its report have been given permission to examine the contents of the 2015 MCAS and PARCC tests under confidentiality conditions, they cannot identify the test items and discuss their contents publicly. In contrast, we focus on publicly available items so that we can discuss their content and format.

It is important to note that we had to rely on online practice tests for our analysis of PARCC test items because most items will not be released to the public. How the content and format of the actual test items differ, if at all, from the content and format of the practice test items made available online by PARCC will remain unknown. In contrast, all used pre-Common Core test items in Massachusetts were released annually through 2007, by statute. After 2007, the Department of Elementary and Secondary Education (DESE) stopped releasing all test items used every year chiefly because of the cost, it explained, in replacing them.⁴ The complete set of actual test items provided much more useful information to parents about the content of the school curriculum (and more pedagogically useful information to teachers) than can a few sample test items.

Commissioner Chester has already committed the state to participate in the Programme for International Students Assessment (PISA) instead of the Trends in International Mathematics and Science Study (TIMSS) in 2015. But whatever Bay State student scores are on PISA in the coming years, the scores will tell us little about the K-12 curriculum because the aptitude-test-like PISA does not reflect the school curriculum, as do TIMSS tests.⁵ PISA's scores will simply indicate some skills

and aptitudes the state's 15-year olds have in comparison with students in other participating countries.

2. INTERNATIONAL COMPARISONS

The directors of the Common Core State Standards (CCSS) project gave the members of the Validation Committee (VC) they had chosen in July 2009 seven criteria for judging the quality of the CCSS in February 2010, several months before release of the final version of these standards. One criterion was whether the standards were: “Comparable to the expectations of other leading nations.”⁶ Failure to meet this criterion was one reason given by those who refused to sign a statement agreeing that the CCSS met all required quality goals.⁷ In a letter to the Committee Chair, Dylan Wiliam wrote:

“The standards ... as Jim Milgram has pointed out, are in important respects less demanding than the standards of the leading nations.”

Nevertheless, the report on the VC, posted in June 2010, did not include comments from the five members of the 29-member committee who did not sign off on the standards. Instead, it asserted that the VC examined “...evidence that the standards are comparable with other leading countries’ expectations,” even though, as Sandra Stotsky pointed out: “No material was ever provided to the VC or to the public on the specific college readiness expectations of other leading nations in mathematics or language and literature.”⁸

Reporting on her own, self-initiated investigation into the matter, Stotsky wrote: “The two English-speaking areas for which I could find assessment material (British Columbia and Ireland) have far more demanding requirements for college readiness. The British Commonwealth examinations I have seen in the past were far more demanding in reading and literature in terms of the knowledge base students needed for taking and passing them.”

Indeed, the testing program aligned to CCSS—PARCC—bears little resemblance to testing

programs our overseas competitors use. Many observers from Europe and Asia would recognize essential features of the original MCAS because it shares those features with their own programs. Few would recognize PARCC’s major features. We detail here seven major differences among international testing programs, pre-Common Core state testing programs, and Common Core-based testing programs.

(1) The testing programs of most of the highest-achieving countries—our economic competitors—address multiple sets of secondary standards, or “pathways.” In contrast, PARCC addresses just one—Common Core’s college and career readiness standards. While MCAS also has addressed just one set of standards in each discipline tested, the standards themselves offered schools some options for grade level placement, especially at the secondary level (e.g., for the sciences, Algebra I, and U.S. History). Moreover, having one set of standards and tests in grade 10 is less problematic than having one set of standards and tests in grade 11 because there are fewer differences in what grade 10 students study compared with what grade 11 students study. Despite generally much smaller disparities in income and narrower “achievement gaps” across demographic groups, most European and East Asian countries provide for differences in curriculum preferences, in academic achievement, and in long-term goals.⁹

In the United States, the philosophy dominant in our schools of education and among some education policy makers discourages different academic pathways because, so the thinking goes, if students do not all experience the exact same curriculum at the same pace they will not all be exposed to the same opportunities.¹⁰ Allowing academically advanced students to learn at a faster own pace would be unfair to their peers, it is implied. And encouraging secondary students to choose among different occupational training programs or apprenticeships in high school (grades 9-12) would also be unfair, they believe, because it would reduce access to higher-status occupations requiring college degrees.

Such indifference to the energizing attraction of student curriculum choice was not always the case.

Lip service was paid in the early days of selling CCSS to a sequence of standards for career-technical programs. “To help improve outcomes in career and technical education, we are also establishing a second competitive preference priority for applications that include a high-quality plan to develop, within the grant period and with relevant business community participation and support, assessments for high school courses that comprise a rigorous course of study in career and technical education that is designed to prepare high school students for success on technical certification examinations or for postsecondary education or employment.”¹¹

In the final CCSS documents, however, the word “career” appears only with “college,” as in “college and career readiness,” allowing scores on a Common Core-based high school test to be used to determine both college and career readiness as the test defines it.¹² This merger of two different concepts ignores differences in student achievement, motivation, and interest. It also remains unclear how a so-called career-readiness test will serve in lieu of a college placement exam.¹³

European and East Asian testing systems reflect their instructional programs. Students are differentiated by curricular emphasis and achievement level, and so are their high-stakes examinations.¹⁴ Differentiation starts at the lower secondary or middle school level in many countries (e.g., Germany) and exists in virtually all of them by the upper secondary level (e.g., Japan).¹⁵ Students attend secondary schools with vastly different orientations: academic schools to prepare for a university, general schools to prepare for the working world or advanced technical training, and vocational-technical schools to prepare for direct entry into the skilled trades. Typically, all three types of school require exit tests for a high school diploma. Higher education institutions determine their own entrance requirements. Colleges define college

readiness, industry defines career readiness, and the two bear little resemblance to each other.

Figure 1 shows three different models for K-12. Why a comparison with Germany and Japan? Although they are no longer the top performers on international assessments, Japan usually ranks near the top in mathematics, and Germany’s achievement level is similar to ours. But as education systems that prepare their youth well not only for university but also for other desirable pathways to adulthood, the German and Japanese systems excel. Both countries have consistently low unemployment rates and strong export industries, even as their skilled laborers rank among the highest paid in the world.

As the figure shows, German and Japanese students can begin useful skills training early in high school. Most U.S. students who desire focused skills training must endure a four-year general high school curriculum before they can choose the kind of curriculum they prefer—in a post-secondary context. Many drop out instead. Not only do they not end up in a post-secondary institution, but they also end up with few skills or meaningful employment prospects.

*(2) A Common Core-based testing program differs from our competitors’ testing programs in the absence of student “stakes”—real consequences attached to test performance.*¹⁶ The two Common Core-based testing consortia, PARCC and SBAC, are mostly building replacements for the No Child Left Behind (NCLB)-mandated state tests, with no stakes for students other than the weak “medium” stakes of a “college readiness” determination built into the high school examination and the right to take credit-bearing courses in their freshman year if they choose to enroll in college and, then, to opt out of “developmental” or remedial courses (even if that is all they are ready for).

However, when one adds student stakes to tests, they appear to lead to a full grade-level increase in student achievement over the course of a K-12 education.¹⁷ Stakes are particularly important at the secondary level, where students

FIGURE 1. THREE MODELS OF PUBLIC EDUCATION SYSTEMS*

Primary Level	Secondary Level	Upper Secondary Level	Higher Education Level
Model 1: First major curricular split occurs at higher education level: UNITED STATES*			
elementary school (5-6 years, 100%)	middle/junior high (3-4 years, 100%)	comprehensive high school (3-4 years, 89%) career/technical high school (2-4 years, 10%) advanced, specialized high schools (3-4 years, <1%)	four-year college (4 years, 64%) community college (2 years, 34%) vocational-technical institute (1-2 years, 3%)
Model 2: First major curricular split occurs at upper secondary level: JAPAN			
	6-year secondary school (6 years, 8%)		
elementary school (6 years, 100%)	lower secondary school (3 years, 92%)	comprehensive high school (3-4 years, 50%) part-time & correspondence schools (3 years, 16%)	university (4 years, 40%) junior college (2 years, 20%)
		polytechnics (3-5 years, 8-13%) specialized training colleges (1-5 years, 8-13%) miscellaneous other schools (1-5 years, 8-13%)	
Model 3: First major curricular split occurs at lower secondary level: GERMANY			
elementary school (4 years, 100%)	university-preparatory school (6 years, 40%) vocational-technical school (6 years, 29%) general secondary school (5 years, 17%) comprehensive school (5 years, 14%)	university-preparatory school (3 years, 30%) vocational-technical school (2-3 years, 25%) work-school training (2-3 years, 45%)	university (3-7 years, 25%) polytechnics (1-6 years, 25%) professional work-school training (2-7 years, 50%)

Source: Updated from Phelps, et al., *Higher education: An international perspective*, Chapter 2.

<http://nonpartisaneducation.org/Review/Resources/IntlHigherEducation.htm>

* Excludes schools for those with disabilities, behavioral problems, and criminal records. Table includes types of government-funded schools, typical duration in years, and estimated proportions of students in each type. For the U.S., the table excludes homeschooling (over 3%) and sectarian or non-sectarian private schools (~10%).

are wise enough to know that they needn't exert themselves when few consequences are attached to their performance, for them. Those who believe that older students will honestly exert themselves if the test administration room is kept quiet and they are given nothing else to do underestimate the appeal of daydreaming and doodling.¹⁸ Moreover, the more complicated and multi-layered the test items are (i.e., the higher the proportion of constructed-response items, the more elaborate their structure, and the more they require initiative on the student's part), the more likely students will ignore them, if stakes are low.¹⁹

Perhaps being held accountable for their own test performance is what lies behind the fact that when students in high-achieving countries do not score as well as they expect, they say it is because they did not work hard enough.²⁰ Massachusetts students failing MCAS after several tries have said the same.²¹

(3) In addition to tests aligned to multiple pathways, other nations generally offer multiple "targets." A multi-target testing system gives every student, regardless of achievement level or choice of curriculum, a high-stakes test with a challenging but attainable goal. In some systems, tests are set

at differing levels of difficulty related to different certifications (e.g., a “regular” diploma and an “honors” diploma). In other systems, tests cover different subject matter (which may be at a much higher level than what is on Common Core-based tests).²²

Before the arrival of CCSS and Common Core-based tests, several U.S. states had begun the introduction of high school end-of-course examination systems, similar to European systems, with tests in the most basic subject areas required, and a choice among others. These nascent systems may now be eliminated as seemingly duplicative and less needed as all eyes focus on implementing the minimally demanding tests that are federally required.²³ However, as R. James Milgram writes: “The CCMS [Common Core Mathematics Standards] are not for the top 30 percent of high school students, but for the “average” ones.” In addition, because “only 40–45 percent of high school graduates that enter college attend non-selective or community colleges,” the majority of high school graduates that enter college “will be less than minimally ready for a regular four-year college or university.” Milgram further asserts that defining college readiness as mastering weak Algebra II content also “disadvantages students whose school districts do not have high socio-economic status. ...the availability of advanced mathematics courses is strongly related to the socio-economic status (SES) of the school district the student attends. A student attending a high school in the lowest SES quintile has only three/fifths the likelihood of access to calculus when compared with a student in the highest SES quintile; the data are similar for trigonometry and statistics.”²⁴

The single-target problem has two solutions, one passive and one active. The passive solution lets individual students take a minimum-competency test early in their school careers; once they pass it they are allowed to move on.²⁵ If the test is high stakes only for individual students, then no one has an incentive to hold higher-achieving students back, that is, to prevent them from taking accelerated course work afterwards, based

solely on test results.

The original MCAS worked this way. High school students who passed the required exams in grade 10 never took MCAS tests again. They were free in their junior and senior years to take whatever courses they wanted to, for their application to a higher education institution.

The active solution to the single-target problem, and the solution that seems to have worked well, is to offer multiple targets. New York stands out historically as the one state that employed a multiple-target examination system, with a Regents “Competency” exam required for high school graduation with a “regular” diploma, and with a Regents “Honors” exam required for graduation with an “honors” diploma.²⁶

Nothing prevents Massachusetts from adding state assessments in health, the arts, foreign languages, and career and technical education, or in restoring the planned assessment of history in grade 10. *Or in keeping advanced coursework in mathematics and science in its high schools.* But this is unlikely to happen. Once control over core assessment programs moves out of state, state department of education testing offices become dependent on the higher powers that control the ELA and mathematics exams.²⁷ If the controlling entities do not require assessments in other subject areas, other curricular tracks, and more advanced coursework in mathematics and science, little incentive remains for high schools to maintain even existing advanced coursework. Federal requirements take priority not only in law, but also in the media and the court of public opinion.

Milgram foresees little coursework beyond the Common Core—“only the most elementary parts of trigonometry, no pre-calculus content, and no calculus coursework at all. In theory, a local school district could keep these advanced courses, but in practice they most likely will not. Particularly in the lowest SES districts—where financial pressure to cut under-subscribed and unrequired courses is greatest—they will be

axed right off or disappear as soon as their lead teachers retire or leave. All that will be required for graduation is CCMS-based course work. That is all that is needed for entry into credit-bearing (i.e., not remedial or developmental) college math courses at the higher education institutions that participated in applications for Race to the Top grants.”²⁸

(4) The PARCC testing program also differs from our competitors’ in the ways in which local educators are involved. For example, the *abitur*, the exit test for German academic high schools, consists each year of test questions submitted by subject area teachers and university professors. Teachers also take part in scoring the test. Similarly, career-tech assessments involve the direct participation of personnel from businesses, labor unions, and government agencies. Indeed, one of the arguments for adding constructed-response test items (i.e., open-ended essay questions) to high-stakes state tests in the U.S. was the opportunity for further involving teachers in the testing process, in addition to having them serve on test item review committees during test development.²⁹

The constructed-response items on early MCAS tests were graded by Massachusetts educators, helping them connect their instruction with student learning and providing them direct feedback on which instructional strategies worked and which did not. Such activities also helped educators better understand measurement and improve their own classroom test construction.³⁰ But PARCC is moving all test design, analysis, and scoring from Massachusetts to other states in the country.

Pearson, Inc. and Educational Testing Service (ETS) conducted PARCC test item development. Their facilities for this activity are located in Iowa, New Jersey, Minnesota, and Texas, although they also work with itinerant contract employees online.³¹ Pearson analyzes test results; the personnel qualified to do that work are in Iowa and Minnesota. PARCC replaces Massachusetts educators’ scoring of constructed-

response test items with scoring done by itinerant, contingent workers located in Pearson’s regional scoring centers in Arizona, Colorado, Florida, Illinois, Iowa, Michigan, Ohio, Texas, Virginia, and Washington State.³² A college degree in any major is all that is required to be a scorer for 20 hours a week, something College Board and ETS never allowed for scorers of Advanced Placement tests.³³

This massive transfer of control and learning opportunities is justified by a promise of improved comparability of educational outcomes across all state and greater speed in returning timely test results to decision makers. But that comparability already existed by means of regular random sampling of students across all states in tests given by the National Assessment of Educational Progress (NAEP). Moreover, it no longer can exist with PARCC because the number of participant states has declined considerably from the original number and may dwindle more.

(5) The PARCC testing program differs from our competitors’ programs in the purposes it claims for its final tests in grade 11. Although other countries understand the importance of differentiating between retrospective achievement tests—tests aligned to past curricula—and predictive tests—tests aligned to future outcomes—the U.S. seems determined to ignore what other countries do or have learned from experience.³⁴

PARCC proponents have promised that it will both retrospectively measure achievement and prospectively predict college outcomes. But, it cannot do both well.³⁵ High school and college differ substantially, for example, in the populations of students who attend, their goals and choices, and the backgrounds of their teachers. Moreover, some students who prosper in high school struggle in college. Finally, and perhaps most important, post-secondary institutions vary enormously in the U.S. in their academic demands.

PARCC suffers from an identity complex. Our overseas competitors do not expect the same

test to serve two disparate functions, so they use different tests or criteria to separate secondary school exit from entrance to a career or a post-secondary institution.

The role of the early MCAS was clear—a retrospective achievement examination measuring mastery of state standards for the goal of a high school diploma. It left college prediction to other tests designed for that purpose: the old SAT tests and more recent ACT tests before they were aligned to Common Core. In sum, the MCAS program as originally conceived was more similar to our overseas' competitors testing programs than PARCC is. In that sense it was better “benchmarked.”

(6) While a multiple-choice format is still used for most test items in both MCAS and PARCC, with most of the remaining items requiring short written answers or longer essay responses, there are a few “innovations” in PARCC test item formats not used elsewhere. They require students to solve multiple-step problems or use computer-based functions. Inherent flaws beset both innovations, which may explain their rarity outside the United States. Later chapters and Appendix A describe these types in more detail.

(7) Our competitors tend to stick with assessment programs for long periods of time, only occasionally updating them—far different from the U.S.'s peripatetic inclination for constant reform. That consistency offers the advantage of building an historical record of assessment results, with all the useful information that provides. Ironically, by arguing that comparing student results across states is of utmost importance,³⁶ Common Core-based tests rip apart the historical time series of test results accumulated by so many states. In order to enhance a little comparability across geography, comparability across time will be destroyed. What would Massachusetts educators consider most informative—comparing their students' scores to Delaware students' scores, or analyzing trends of Massachusetts student scores over time?³⁷

Not only would dropping MCAS extinguish an informative historical record of test results, it would replace a test whose reliability and validity are long established in order to start from scratch with one whose reliability has yet to be established and whose validity may remain forever unknown.³⁸ Records to back a claim that PARCC's test scores better predict college and career readiness will not be available for years, while all its used test items may never be available for public scrutiny and teachers' information.

3. LINES OF DIFFERENCE: HOW MCAS AND PARCC DIFFER AS TESTING SYSTEMS

How PARCC and pre-Common Core MCAS differ as testing systems can be summarized using Table 1 as a basis for comparison. In this chapter, we discuss general features of the two systems needing elaboration.

PURPOSES/GOALS

As is well known, MCAS was mandated by MERA. Among many of its features, the 1993 Act required and funded DESE to create an assessment system that would be used to hold schools and districts accountable for the effectiveness of their programs by measuring students' knowledge and skills in seven subject areas—English, mathematics, science, history/social science, foreign languages, health, and the arts. PARCC was originally funded by the USED to assess Common Core's standards in English and mathematics. It is now a private entity. Test results will be used for teacher and principal evaluations according to the Memorandum of Agreement (MOA) signed in 2010 by Governor Deval Patrick, Secretary of Education Paul Reville, and Commissioner of Elementary and Secondary Education Mitchell Chester, and implemented by regulations approved by BESE in June 2011 and December 2013 (603 CMR 35.00: M.G.L. c.69, §1B; c.71, §38).

TABLE 1. HOW MCAS AND PARCC DIFFER AS TESTING SYSTEMS

	Pre-Common Core MCAS	PARCC
Purpose	To evaluate school and district performance from grade to grade To determine eligibility for a high school diploma on tests based on grade 10 MCAS standards	To determine if a student in grade 11 is college and career ready or “on track” toward this goal from grade 3 on. To determine if a student in grade 11 is eligible for credit-bearing college freshman courses and can by-pass “developmental” courses based on Common Core standards For use in teacher/principal evaluations*
Standards for	English Language Arts, Mathematics, Science/Technology/Engineering, History and Social Science, Foreign Languages, Health, and the Arts	English Language Arts, Mathematics**
Tests in	ELA, Mathematics, History/Social Science, Science/Technology/Engineering	ELA, Mathematics
Career-technical standards	No	No
Acceleration possible	Yes	Probably No
College freshman placement in reading, math	Determined by higher education teaching faculty	Determined by high school GPA and/or cut score for college readiness in grade 11***
Items released	About half released annually since 2008	Not released
Test reliability	Established	Not Established
Test validity	Established	Not Established
Test results available	Fall following spring tests	Early summer following spring tests
Total testing hours in 2015	See Table 2	See Table 2
How scored for ELA and by whom in past	Machine-scored for multiple-choice items; all writing hand-scored mostly by MA educators in past****	Machine-scored for multiple-choice items; all writing now hand-scored but eventually to be scored by computer program****
Cut score process	See Massachusetts DESE website.***** Based on typical student performance in MA	See PARCC website.***** Based on typical student performance among PARCC members
Governance of system	Totally MA	One vote by state education CEO among member states
Control of test design, administration, analysis, and contractors	MA	PARCC and its member states
Incentives for teachers and students to go beyond minimal competencies or pass/fail	Scholarship for high performance for free merit-based tuition at public colleges for specified number of semesters; grade 12 attendance in MA public high school required for scholarship	None so far

*In a Memorandum of Agreement (MOA), 2010, signed by Governor Deval Patrick, Secretary of Education Paul Reville, and Mitchell Chester, Commissioner of Elementary and Secondary Education.

**Common Core includes general “literacy” standards (reading and writing standards) for science and the social sciences but no specified science or social science content.

***http://www.sudbury.k12.ma.us/index.php?option=com_docman&task=doc_view&gid=396&Itemid=349.%20See%20#5. See Point 5.

**** See http://www.doe.mass.edu/mcas/tech/technical_quality.pdf and <http://www.parcconline.org/news-and-updates/252-did-you-know-who-scores-the-parcc-test>

***** http://www.doe.mass.edu/mcas/tech/technical_quality.pdf

One of MERA's defining features is that it requires high-stakes tests based on grade 10 standards to determine if a student is academically eligible to receive a high school diploma. PARCC does not determine if a student is academically eligible to receive a high school diploma, but the meaning of its grade 11 test is not clear. MCAS is given to students enrolled in vocational-technical schools as well as in other public comprehensive high schools. So is PARCC. But to pass the required grade 10 MCAS tests does not mean that a student is career-ready, only that he or she is eligible for a high school diploma. In contrast, to pass PARCC's college and career readiness tests in grade 11 apparently means that a student is eligible for credit-bearing college freshman coursework, is exempt from a college placement test, and is ready to undertake preparation for an occupation of his/her choosing. But how is all this possible for one test to determine? The phrase "college- and career-ready" links together two goals (readiness for a career and readiness for college) as if they are identical. But, as Chapter 2 suggests, careers and college do not call upon all of the same skills and are measured by different tests in other countries. PARCC purports to capture both goals equally in one test. But how can its tests be valid if the skills needed for the two goals are substantially different?

Moreover, as Table 1 notes, one of the key requirements in the grants USED awarded to PARCC and SBAC is that, in addition to measuring college- and career- readiness, both Common Core-based testing systems have to determine whether or not students are "on track" to achieve this goal. According to a footnote, "on track" to being college- and career-ready means "proficiency" at every grade level. NCLB left states to determine the meaning of proficiency in reading and mathematics. Now it is up to two now-private testing companies. Is this preferable or progress?

Moreover, whereas once it was up to college teaching faculty to determine what they meant by college readiness, it is now by federal regulation

a threshold to be determined by the pass scores on tests given in grade 11 and without any evidence that their cut scores align with faculty understanding of readiness for college coursework or with the cut scores on the placement tests they had been using.

PARCC AND MCAS TEST ITEMS

Pre-Common Core MCAS and PARCC draw upon many of the item types most tests do: multiple-choice questions and short or long essays (or both). Here we focus on the innovative items used by PARCC and MCAS in writing and reading because Common Core-based tests for ELA seem to have received much less public attention than tests for math. However, the problems we find in the new test-item types in PARCC ELA Practice Tests can be also found elsewhere, e.g., in SBAC math Practice Test items, as indicated by Steven Rasmussen in March 2015.³⁹ Ze'ev Wurman, another mathematics expert, further corroborated what we say when he observed on a private listserv that the "large amount of added verbiage in item prompts, particularly in the open response and performance items, makes the test more of an assessment of reading comprehension and of following instructions than of mathematics." These problems "have little to do with computer administration but everything to do with trying to influence classroom instruction to be more *word-based* and less symbol- and procedure-based."⁴⁰

PARCC's Technology-Enhanced Responses (TERs) are item types that make use of drag-and-drop or cut-and-paste functions in many of the reading and writing exercises used on PARCC's computerized versions of the test. TER's presence in PARCC computerized tests is one of the many technology requirements specified in USED's Race to the Top grant application. While seemingly benign, TERs have come to symbolize a broader problem PARCC has faced since computerized test trials began in 2014. Parents, students, and school administrators in the Bay State as well as in other

states have been sharply critical of its online tests.⁴¹ System breakdowns, inadequate teacher training, confusing instructions for accessing the tests, and developmentally inappropriate instructions have all been part of the complaints publicized in the media. An analysis of PARCC's 2014 field trials in the Bay State further substantiates concerns about using PARCC's computerized tests instead of paper-and-pencil tests, especially in the absence of pedagogically useful information on student growth.⁴²

PARCC's Evidence-Based-Selected Response items (ESBRs) are a second new item type. They seem to be an adaptation of what psychometrician James Popham describes as a "multiple, binary choice item" that links *two* multiple choice items.⁴³ The answer selected in Part A is used to frame additional questions in Part B. In PARCC's ESBRs, the answer to Part A sets up a question in Part B that compels students to go back to the question in Part A and into the reading passage (s) used for the question in Part A. Parts A and B are designed to move test-takers back and forth through the text and the answer options. Unless readers make the correct choice in Part A, however, their answer in Part B will not be scored as correct, even if they select the right answer in Part B.

At this point in time, there is apparently no research suggesting that ESBR items are valid measures of reading comprehension. Psychometrician David Frisbie reviewed the research on a variation of this item type (called multiple true/false items or MTFs), conducted on college students, and observed that he could find no research using MTFs for testing younger subjects.⁴⁴ Given the lack of evidence for the value of ESBR test items in measuring reading comprehension in K-12, PARCC needs to explain why it heavily uses this type of test item in a large-scale test with consequences for K-12 teachers and students.

In contrast to PARCC, MCAS makes use of short written responses (ORs), as well as long but open compositions—item types which can

be scored only by trained readers, not computers. PARCC has no item type equivalent to either of these types. The long composition in MCAS requires students to choose the text they wish to write about and is closer to the kind of writing required in college than the structured writing exercises in PARCC. MCAS ORs and multiple choice items also work in tandem to measure reading comprehension and, indirectly, writing effectiveness. PARCC's test designers could have included open response items in their assessment system because there is research evidence for using ORs.⁴⁵ But PARCC chose instead to place a premium on ESBRs, and without an explanation to the public.

ACCESS TO USED TEST ITEMS

MERA mandated release of all used test items annually but, after 2007, DESE released only half the items, stating costs of developing new items and length of testing time as reasons for reducing the number of items released. Complete transparency allowed teachers, administrators, researchers, and parents alike to study how students performed relative to the standards being assessed. Since 2008, their ability to do so has been limited.

So far, PARCC as a private entity will not release its test items to the public and is not compelled to do so. This policy has deleterious consequences for end-users who may want to study the test in order to learn how to prepare students for future test administrations and to understand the basis for the high stakes decisions made on the basis of test results. Today, with both tests, it is a guessing game, much more so with PARCC than MCAS which still releases half of used test items.

RETURN OF TEST RESULTS

MCAS is administered once each year, returning its results in the early fall in order to use the summer months to hand-score long compositions, short-answer questions and "open response" items. In May 2015, PARCC voted to compress its 2015 Performance Based Assessment (PBA) and End of Year (EOY) assessment into a single test to be administered in

the spring of 2016.⁴⁶ As of this writing, PARCC says only that its results will be returned sooner than fall. However, DESE has projected early June as the final date for PARCC's testing period in 2016.⁴⁷ This means that scores returned even just a few weeks later would have little instructional value for teachers whose students will have left for summer vacation and little value for evaluating their teachers (something that usually takes place in the spring). In other words, a vote for PARCC in place of MCAS does not accelerate useful information to the schools.

TESTING TIME

Table 2 indicates testing times for both math and ELA in both tests in 2015. Figure 2 presents a way of visualizing the differences in testing hours through the grades. PARCC is timed, MCAS untimed.

As indicated earlier, PARCC is reducing the number of hours for testing time in 2016. However, schools do not know why PARCC still needs from over eight to over nine hours of testing time per grade. Is it in part because the “modern” item types used by PARCC (described below) provide better instructional information than what can now be obtained through MCAS despite taking up much more test-taking time? Is it in part because PARCC chose to assess three long compositions at each grade? Nowhere has PARCC explained publicly how these new item types and the number of long compositions it chose to assess per grade (as well as the kind of writing it chose to assess) add to our understanding of student growth to justify these many extra and costly hours of testing.⁴⁸ Nor has PARCC indicated where there is evidence for the new test-item types it uses.

These gaps leave basic questions unanswered about what educators will learn about curriculum improvement from PARCC tests. In fact, PARCC testing time raises questions about the breadth of the school curriculum in future years: tests lasting up to or over nine hours each spring leave little to no room for testing more than math and English. MCAS strategically

spaced its science and history/social science tests in different grades to avoid overwhelming elementary classroom teachers and students with too much spring testing. Science and social studies are not tested in every grade, and three other subjects—health, foreign languages, and the arts—are still not tested, even though MERA required testing them. When PARCC becomes the state's new state test for math and ELA, it is not clear that, legally, DESE can discontinue testing science and history/social science and abandon developing tests for the other subjects. But will the USED be calling the shots? What will happen to the breadth of the current school curriculum if only two or three subjects are tested? (DESE is expected to adopt Achieve, Inc.'s Next Generation Science Standards despite the largely negative evaluation these standards received from scientists.)⁴⁹

ESTABLISHMENT OF CUT SCORES

Cut scores on MCAS determine what level of test performance constitutes Advanced, Proficient, Needs Improvement or Failing. A score in the Needs Improvement category for grade 10 is the minimum result a student must attain for each of the subjects tested as part of the state's high school graduation requirements. The methodology for standard setting is detailed in DESE's 2008 publication *Ensuring Technical Quality: Policies and Procedures Guiding the Development of the MCAS Tests*, describing how cut scores have been established for all of the tests in the MCAS battery since 1998.⁵⁰ The last set of cut scores was established for science and technology in 2007. PARCC uses a different process.⁵¹

Different as these standard setting procedures are, what is of utmost importance is *who* will determine what the cut scores are for each PARCC test. By involving personnel from multiple states, PARCC's cut scores are unlikely to reflect Bay State perspectives. For MCAS, cut scores are determined by Massachusetts teachers, parents, and leaders. But this is only one way to look at “who.”

TABLE 2. TIME ALLOTTED FOR ENGLISH LANGUAGE ARTS AND MATHEMATICS TESTS FOR PARCC AND MCAS IN 2015

Grade	Subject	MCAS 2015	PARCC Subject	PARCC 2015 (Unit)
3	ELA Reading Comp	120	ELA PBA	210
			ELA EOY	75
	Math	90	Math PBA	150
			Math EOY	150
TOTAL Grade 3 (minutes)		210		585
TOTAL Grade 3 (hours)		3.50		9.75
4	ELA Reading Comp	120	ELA PBA	225
	ELA Composition	90	ELA EOY	75
	Math	90	Math PBA	150
			Math EOY	150
TOTAL Grade 4 (minutes)		300		600
TOTAL Grade 4 (hours)		5.00		10.00
5	ELA Reading Comp	120	ELA PBA	225
			ELA EOY	75
	Math	90	Math PBA	150
			Math EOY	150
TOTAL Grade 5 (minutes)		210		600
TOTAL Grade 5 (hours)		3.50		10.00
6	ELA Reading Comp	120	ELA PBA	225
			ELA EOY	120
	Math	90	Math PBA	150
			Math EOY	155
TOTAL Grade 6 (minutes)		210		650
TOTAL Grade 6 (hours)		3.50		10.83
7	ELA Reading Comp	120	ELA PBA	225
	ELA Composition	90	ELA EOY	120
	Math	100	Math PBA	150
			Math EOY	155
TOTAL Grade 7 (minutes)		310		650
TOTAL Grade 7 (hours)		5.17		10.83
8	ELA Reading Comp	120	ELA PBA	225
			ELA EOY	120
	Math	100	Math PBA	150
			Math EOY	155
TOTAL Grade 8 (minutes)		220		650
TOTAL Grade 8 (hours)		3.67		10.83
9	ELA Reading Comp	0	ELA PBA	225
	Math	0	ELA EOY	120
			Math PBA	165
			Math EOY	155
TOTAL Grade 9 (minutes)		0		665
TOTAL Grade 9 (hours)		0.00		11.08
10	ELA Reading Comp	135	ELA PBA	225
	ELA Composition	90	ELA EOY	120
	Math (2 sessions)	100	Math PBA	165
			Math EOY	155
TOTAL Grade 10 (minutes)		325		665
TOTAL Grade 10 (hours)		5.42		11.08
11	ELA Reading Comp	0	ELA PBA	225
	Math	0	ELA EOY	120
			Math PBA	165
			Math EOY	165
TOTAL Grade 11 (minutes)		0		675
TOTAL Grade 11 (hours)		0.00		11.25
TOTAL TESTING TIME (hours)		29.75		95.67

Sources:

<http://massparcctrial.org/2014/12/15/2015-parcc-session-times/> 8/20/15
<http://www.doe.mass.edu/mcas/1415schedule.pdf> 8/20/15

Logically, whether or not high school students are college ready is a judgment that should be made chiefly by those who teach at the college level. But PARCC seems to be drawing on a mix of high school teachers, school or college administrators, and college instructors to determine which students are ready for college-level work.⁵² This is almost like asking a group of podiatrists to determine what expertise an anesthesiologist needs in order to treat patients successfully. This is not to discredit high school teachers or school administrators, simply a recognition that their judgments are not based on teaching college students math or English. The problem has been exacerbated by setting grade 11, not 12, for this determination, a grade level where it seems reasonable to involve high school teachers.

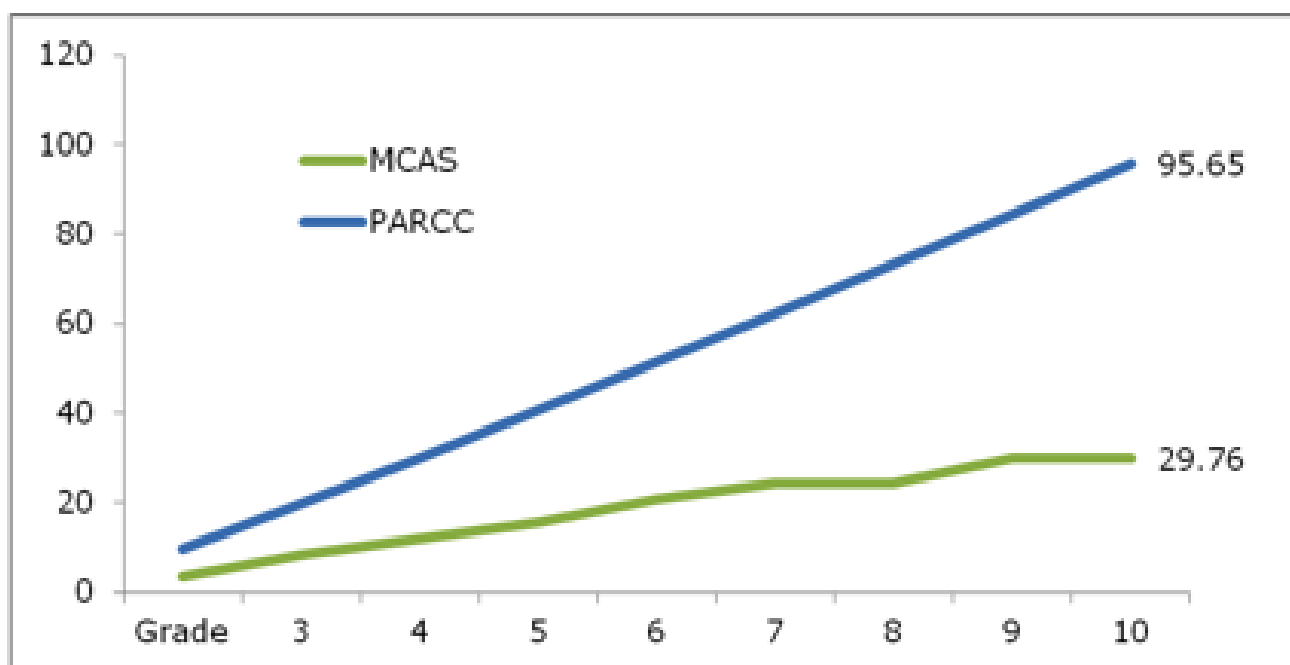
We can also ask who should be judging whether or not a student is career-ready. High school and college teachers or administrators? Shouldn't those judgments be made by business people or instructors in technical institutes? The definition developed by the USED is meaningless—that all students have “the knowledge and skills needed to succeed in college and the workplace.” Who knows what this body of knowledge and these skills are? Business people and college teaching faculty, not K-12 teachers. But they are at best only a small part of a mix for setting cut scores.

To establish performance levels for MCAS in grade 10, the state used chiefly high school teachers. That made sense for students in grade 10. It led to reasonably high standards because the students would attend their high school for another two years, and these teachers would benefit from having more academically competent students. This kind of incentive doesn't exist when high school teachers set performance levels for a “college and career readiness” test.

INCENTIVES FOR HIGH PERFORMANCE

A final difference between PARCC and MCAS is the provision for incentives for high performance. At the inception of MCAS testing BESE approved two merit-based scholarships

FIGURE 2. CUMULATIVE HOURS IN TESTING, MCAS AND PARCC, 2015



for high-performing high school students: the Stanley Z. Koplik Certificate of Mastery and the Abigail and John Adams Scholarship. Among other details, students must demonstrate that they have attended grade 12 in high school, have met the performance requirements specified by law, and intend to use the scholarships within six years to attend a public institution within the state.⁵³

These scholarships address a well-known phenomenon—lack of student effort for low-stakes tests. Student motivation may be greater for the grade 10 MCAS, when, for the first time, test performance has a consequence. At this point, it can mean failure to graduate from high school. But the Koplik and Adams scholarships give students a reason to want to perform well on the MCAS. Multiple scores of Proficient and Advanced will lead to tuition support for any student—rich or poor—to attend college. Common Core-based tests, at this point, offer no incentives for teachers and students to go beyond minimal competencies or pass/fail. Or for students to earn a high school diploma.

4. THE MEANING OF COLLEGE READINESS IN MCAS AND PARCC

Although Common Core promised to make all students college-ready, it has never indicated what exactly students should be able to read once they were declared college-ready. Appendix B in Common Core’s English language arts standards document offers no specifics on what constitutes “college readiness” in reading even though it provides a wide range of exemplars of “quality” and “complexity” at the high school level. Should a grade 12 high school student declared college-ready in grade 11 be able to read the textbooks used in college freshman courses, many of which are written at the college level by college faculty? Or should English teachers in grades 11 and 12 just guess at the reading level and kind of reading curriculum suggested by the passages on the grade 11 Common Core-based practice tests?

Developers of Common Core-based tests themselves offer no clear information on the meaning of college readiness in reading. Making the situation even less clear, the USED and state agencies have sought for about six years to eliminate use of college placement tests by public institutions of higher education. As reported by

Catherine Gewertz in her blog for *Education Week* on April 21, 2015, 600 higher education institutions have agreed to use the cut scores for the grade 11 Smarter Balanced Assessments in place of scores on a college placement test.⁵⁴ However, none is reported as having made an analysis of the grade 11 tests (in advance, under secure conditions) to find out the differences, if any, between the cut scores on their placement tests for entering freshmen and the cut scores for SBAC's tests. In other words, every one of these 600 higher education institutions is apparently allowing grade 11 tests to define readiness for coursework in its own institution when its faculty has neither seen the tests nor been given any evidence that the tests' definition of college readiness (as determined by their cut score and difficulty level of their test items) aligns with their own understanding of readiness for college coursework and with the cut scores on the placement tests they had been using.

Mathematics instructors seem to have been ignored by more than the administrators in their own institutions. Earlier in April, Susan Keene Haberstroh, Chief of Policy and External Affairs at the Delaware Department of Education, issued a press release saying that four major Delaware universities/colleges would accept students' scores on the state's new 11th grade Smarter Balanced Assessments "as an indication of college readiness and in lieu of scores on a separate placement test."⁵⁵ After reading this press release, a co-author of this White Paper contacted Haberstroh for the names of the teaching faculty at these institutions who had examined SBAC's college readiness tests. After two inquiries, the co-author was advised to contact the institutions themselves for the names. Haberstroh gave no indication that teaching faculty had been involved in the decision to allow grade 11 tests to define readiness for credit-bearing coursework in their own institution.

Massachusetts has also ignored most of the teaching faculty in its own post-secondary education institutions on the value of placement tests, even though these tests have long been used

at their recommendation to determine whether students can enroll directly in credit-bearing mathematics, English, and other courses or must enroll in non-credit-bearing "developmental" or remedial courses to prepare them for credit-bearing coursework. But the meaning of the state-determined cut score on the College Board's Accuplacer Computerized Placement Test, the reading test used in the Bay State, is not clear with respect to the reading level of the materials students should be able to read if exempted from a developmental reading course. More is known about the meaning of placement in developmental mathematics courses in the Bay State.

So we look first at how the Massachusetts Board of Higher Education (MBHE) recently reshaped the meaning of college readiness in mathematics. The next section draws chiefly from articles recently published by Professor Richard Bisk and his colleagues at Worcester State University in the *New England Journal of Higher Education* and on his testimony in April 2015 to the MBHE.

THE NEW MEANING OF COLLEGE READINESS IN MATHEMATICS IN MASSACHUSETTS

As we all know, the cost of higher education has dramatically increased in recent decades for many reasons. Incoming college students who require developmental mathematics coursework are one of those reasons. While students themselves pay tuition for these non-credit-bearing courses, the institutions must pay those who teach them without the expectation that tuition will cover their costs. Over the years student fees for college functions have been raised, often sharply, to address the rising costs of higher education in the Bay State, but tuition has risen slowly if it all.

The Common Core standards adopted by the Board of Elementary and Secondary Education in 2010 promised indirectly to solve the problem of rising costs for developmental coursework by claiming they would make all students "college ready" by grade 11. The students' reward, upon enrollment at a public institution of higher education in the state, would be the right to take

their freshman course in mathematics for credit, and without a placement test. They would also be able to transfer credit for their community college coursework to four-year public colleges. In effect, grade 11 tests based on Common Core's standards would define college readiness without any evidence that the tests' definition of college readiness (as determined by their cut score and the difficulty level of their test items) aligned with a college teaching faculty's own understanding of readiness for college coursework and with the cut scores on the placement tests they had been using.

Administrators at public higher education institutions in each state (not their academic senates) were asked to commit their institutions to these conditions in applications for a Race to the Top (RttT) grant in 2010. The MBHE followed up the Bay State's receipt of RttT funds with policy decisions in late 2013 changing how students are placed into their first college-level mathematics class.⁵⁶ These decisions are to be reviewed and finalized in late fall of 2015—at about the time BESE decides on whether to adopt PARCC.

The previous policy for placing incoming students into their first mathematics course was based on a 1998 report from the Mathematics Assessment Task Force. The policy required all incoming students to take the Accuplacer Elementary Algebra exam, which covers topics found in the Algebra I course typically taught in grade 8 or 9. It also established the cut score determining placement in a developmental mathematics course.

In October and December 2013 the MBHE voted to accept recommendations made by a 17-member Task Force on Transforming Developmental Math Education. In contrast to the 1998 task force, only five of the 17 members of the 2013 task force were employed as mathematics faculty.⁵⁷ The new task force recommended that high school graduates with an overall high school Grade Point Average (GPA) of 2.7 or higher be exempt from the

initial placement exam and placed directly into the lowest college-level math course appropriate for their chosen pathway of study. They also recommended that high school graduates with an overall GPA lower than 2.7 but higher than 2.4 who had passed four math courses including math in their senior year would also be exempt from the initial placement exam and placed directly into the college-level math course appropriate for their chosen field of study.

Neither the 2013 task force report nor the MBHE addressed several important questions. Why could students who passed but performed poorly in every high school math class they took be exempt from a placement test so long as their overall high school GPA was 2.7? And why a GPA of 2.7? According to the College Board's 2013 State Profile Report, there does not seem to be a large number of students in the Bay State with a GPA below 2.7, suggesting that it is a low threshold.⁵⁸ Moreover, the MBHE already mandates a minimum GPA of 3.0 for entrance to the state's universities.

Finally, why did the MBHE adopt policies without evidence to support them? None of the studies in the bibliography in the task force report provided evidence for the academic effectiveness of its recommendations for state universities—to the effect that incoming students benefit mathematically from taking math courses that are beyond their mathematical skill level. Indeed, few studies listed in the heavily annotated bibliography were even relevant to state universities. The bibliography was, instead, skewed to community colleges and against the use of placement tests.

It is therefore not clear why community colleges and state universities were subjected to the same policies by the MBHE's vote in 2013, when they have different admission requirements and serve different student populations. Unlike state universities, community colleges are open-enrollment institutions and admit large numbers of unprepared students. According to the task force report itself, 53 percent of incoming

community college students in the Bay State required developmental math education in fall 2010, compared with 23 percent of incoming students at state universities. As Professors Mike Winders and Richard Bisk at Worcester State University asked in their September 2014 article: Why were so few discipline-based experts in the state university system on the 2013 task force?⁵⁹ One wonders who chose its members and why they were chosen.

Placement tests serve a vital purpose at community and state universities or colleges. They do not tell us whether a student will be successful in college math classes, but, rather, if the student has the knowledge base to be successful at a particular level of math. Nor can a grade 11 PARCC test obviate the need for placement tests in the Bay State, when the quality and cut score of the test will remain unexamined by the state's own higher education teaching faculty. A change in the definition of college readiness (by means of changes in the cut scores for credit-bearing courses) has a severe impact on the future of generations of students because it affects the content of post-secondary programs in a vast range of professions and occupations.

In the past, many students enrolling in a two- or four-year public institution of higher education in the Bay State have not taken a mathematics course in their senior year of high school (the 2013 task force report noted but provided no numbers on this phenomenon) because they had not been required to do so, either by local school committees in the Bay State or by the MBHE itself. Instead, many high school students in the Bay State have enrolled in a post-secondary public institution with a gap of one and one-half years between their last mathematics course (in their junior year) and their enrollment in college (and taking the placement test).

Although the MBHE is now requiring all college freshmen in public institutions to have taken four years of math in high school, beginning with the freshman class of 2016, the new MBHE policy does not require that the fourth

year of high school math be more advanced than (or as advanced as) the previous math course. Thus, regardless of the BESE vote on PARCC, full approval and implementation of the MBHE's 2013 decisions at its fall meeting in 2015 will likely result, as Professors Bisk and Winders suggest, in pressure to lower standards in entry-level mathematics courses in order to avoid an increase in the number of students failing their first college-level math course.⁶⁰ The MBHE's decisions are unlikely to result in more mathematically able students in our public colleges.

THE TEACHING AND LEARNING GAP THROUGH THE GRADES

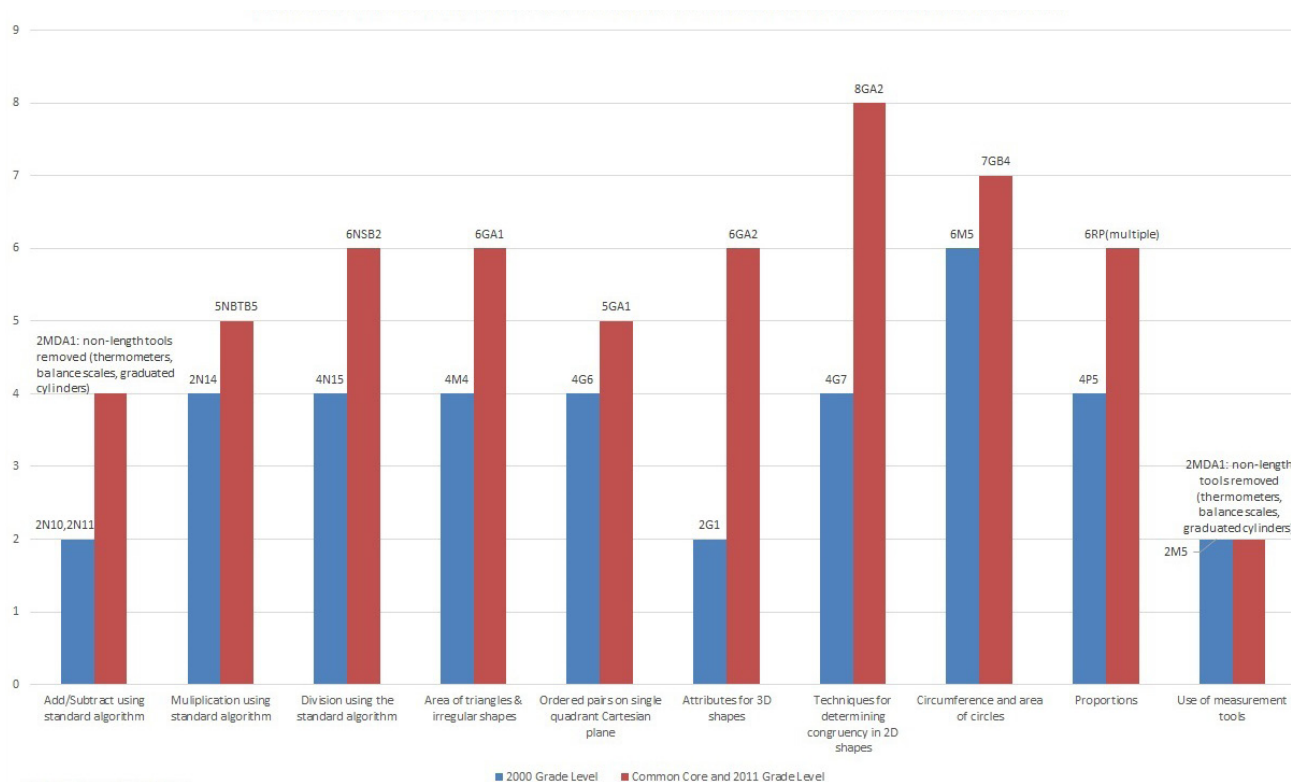
The final academic meaning of college readiness in mathematics in the Bay State depends first on curbing course title inflation in high school, a national phenomenon in the past three decades that has been well documented.⁶¹ It also depends on whether use of PARCC tests in the state could stimulate development of a mathematics curriculum that is more rigorous than the mathematics curriculum stimulated by the original MCAS tests and the standards on which they were originally based. It is unlikely that a more rigorous curriculum will emerge. The blue bars in Table 3B show that the specified topics listed under each set of bars have been taught almost consistently at earlier grades under the 2000 Massachusetts mathematics standards than they now are under Common Core and the 2011 Massachusetts mathematics standards. These topics will be assessed on Common Core-based tests at the grade level of the Common Core standard.

As Tables 3A and 3B show, the gaps between when these topics were taught and then assessed by MCAS and when these topics are taught and then assessed by PARCC (and SBAC) are often quite large, confirming Professor R. James Milgram's judgment that by high school most American students will be several grades behind their international peers in mathematics.⁶²

TABLE 3A. GRADE LEVEL WHEN EACH MATHEMATICS TOPIC IS FIRST TAUGHT IN THE 2000 MASSACHUSETTS MATHEMATICS CURRICULUM FRAMEWORK AND IN COMMON CORE AND THE 2011 MASSACHUSETTS MATHEMATICS CURRICULUM FRAMEWORK

Topic	MASS 2000	Common Core
Use of measurement tools	2	2
Add/subtract using standard algorithm	2	4
Attributes for 3D shapes	2	6
Multiplication using standard algorithm	4	5
Ordered pairs on single quadrant Cartesian plane	4	5
Division using the standard algorithm	4	6
Area of triangles and irregular shapes	4	6
Proportions	4	6
Techniques for determining congruency in 2D shapes	4	8
Circumference and area of circles	6	7

TABLE 3B. WHEN MATHEMATICS TOPICS ARE FIRST TAUGHT IN THE 2000 MASSACHUSETTS MATHEMATICS CURRICULUM AND IN COMMON CORE AND THE 2011 MASSACHUSETTS MATHEMATICS CURRICULUM FRAMEWORK



Source: V. Mollo, Software Engineer

“...It is important to remember that the hierarchical nature of mathematics implies that incomplete courses in lower grades will always haunt students and their teachers in higher grades. ...defining college readiness as mastering weak Algebra II content disadvantages students whose school districts do not have high socio-economic status.... The CCMS effectively end at a weak version of Algebra II, defined as sufficient to make students “college and career ready.” Unless our high schools provide the coursework they need, mathematically capable students will no longer be able to prepare for careers in science, technology, engineering, and mathematics (STEM).”

WHY THE OVERALL LEVEL OF READING DIFFICULTY MATTERS IN HIGH SCHOOL

The basic problem is that most American high school graduates today cannot read college-level textbooks. We know that the average American high school student today is not ready for college-level reading from two independent sources: (1) Renaissance Learning’s latest report (2015) on the average reading level of what students in grades 9-12 choose or are assigned to read,⁶³ and (2) the average reading level of the books that colleges assign incoming freshmen to read (the 2013-2014 “Beach Book” report).⁶⁴ From two sources that are independent of each other, we can infer that average American high school students read at about the grade 6 to 7 level. Some high school students can read high school-level material, of course, while many others are still reading at an elementary school level even though they are in high school.⁶⁵

Based on the information available, it seems that our colleges are not demanding a college-level reading experience for incoming freshmen. Nor are they sending a signal to the nation’s high schools that high school-level reading is needed for college readiness. Indeed, they seem to be suggesting that a middle school reading level is satisfactory, even though most college textbooks require mature reading skills. However, our colleges can’t easily develop college-level reading skills if most students admitted to a post-secondary institution in this country have difficulty reading even high school-level

textbooks. No wonder community colleges spend a lot of money on developmental coursework for new freshmen.

WHAT COLLEGE READINESS SEEMS TO MEAN IN GRADE 10 MCAS ELA IN 2000 AND 2001

To understand why MCAS ELA tests may have helped to strengthen student reading and writing skills, especially through grade 10, we look first at the reading selections on the grade 10 MCAS ELA tests in two of the years when all test items were released annually. We then look at the types of questions that elicited student writing (at all grade levels) and how they were scored. These types of questions are important because of the models they provided to the state’s teachers, especially the Open-Response (OR) questions. MCAS used a variety of question types for math as well as ELA tests, not just multiple-choice options.

1. Reading Selections on MCAS ELA Tests in Grade 10 in 2000 and 2001

Grade 10, 2000

Excerpt from *The Perfect Storm* by Sebastian Junger

Poem by Robert Frost, “Acquainted with the Night”

Excerpt from *The Changing Year* by Rachel Carson

Essay or short memoir by Jesus Colon, “Kipling and I”

African myth: “The Three Calabashes” and Greek myth: “Pandora”

Short Story, “Early Autumn,” by Langston Hughes

Grade 10, 2001

“Lego” by David Owen, article in the January 14, 1991 *New Yorker* magazine.

Poem by Mitsuye Yamada, “A Bedtime Story”

Sonnet 116 by William Shakespeare

Excerpt from *The Grapes of Wrath* by John Steinbeck

Essay by Loren Eiseley, “The Angry Winter”

Excerpt from *The Autobiography of an Ex-Colored Man* by James Weldon Johnson

2. *Types of questions on MCAS tests and how they were scored*⁶⁶

Multiple-choice questions are included on all MCAS tests except the ELA Composition and require students to select the correct answer from a list of four options. Responses to multiple-choice questions are machine scored.

Short-answer questions are included only on Mathematics tests and require students to generate a brief response, usually a numerical solution or a brief statement. Responses to short-answer questions are scored on a scale of 0-1 points by one scorer at grades 3-8 and by two independent scorers at grade 10.

Short-response questions are included only on the grade 3 ELA test and require students to generate a brief response to a reading comprehension question. Responses to short-response questions are scored on a scale of 0-2 points by one scorer.

Open-response questions are included on all MCAS tests except the ELA Composition and require 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. Students can respond correctly using a variety of strategies and approaches.

Responses to open-response questions are scored using a scoring guide and anchor papers (student work), for each question. The scoring guides indicate what knowledge and skills students must demonstrate. Open-response questions are scored on a scale of 0-4 points, with the exception of grade 3 Mathematics, which is scored on a scale of 0-2 points.

Answers to open-response questions are not scored for spelling, punctuation, or grammar. Responses are scored by one scorer at grades 3-8. Grade 10 ELA and Mathematics tests and high school Science and Technology/Engineering tests are scored by two independent scorers.

Writing prompts are included only on ELA Composition tests and require students to respond by creating a written composition. Student compositions are scored independently by two scorers for topic development, based on a six point scale, with students receiving from 2 to 12 points (the sum of scores from each of the two scorers), and for standard English conventions, based on a four-point scale, with students receiving from 2 to 8 points (the sum of

the scores from each of the two scorers).

Student compositions that do not address the prompt are deemed non-scorable (NS), earning them 0 out of 20 possible points.

3. *MCAS Writing Prompts and Questions for OR in 2000 and 2001*

In all cases, students were given two test sessions for generating an authentic essay, the first for brain-storming, outlining, and note-taking, and the second for drafting a scorable essay. This two-part process was established to honor the concept of a writing process within testing conditions.

Grade 10, 2000

Composition Writing Prompt: Often in works of literature, there are characters—other than the main character—whose presence in the work is essential. From a work of literature you have read in or out of school, select a character, other than the main character, who plays a key role. In a well-developed composition, identify the character and explain why this character is important.

Question for an OR to an excerpt by Sebastian Junger: “Explain how the quotation by Herman Melville is appropriate for this excerpt. Use specific evidence from the text to support your explanation.”

Question for an OR to an excerpt by Rachel Carson: “The author uses both literary and scientific language in this excerpt. Choose one example of literary language and one example of scientific language and explain how each contributes to the development of the excerpt.”

Question for an OR to the essay by Jesus Colon: “Explain the author’s attitude throughout this essay toward the poem “If—.” Use specific evidence from the essay to support your answer.”

Question for an OR to the African and Greek myth: “Explain the similarities and differences between “The Three Calabashes” and “Pandora.” Use specific evidence from both myths to support your answer.”

Grade 10, 2001

Composition Writing Prompt: “A frequent theme in literature is the conflict between the individual and society. From a work of literature you have read in or out of school, select a character who struggles with society. In a well-developed composition, identify the character and

explain why this character’s conflict with society is important.”

Question for an OR to the article “Lego” from *The New Yorker* magazine: “Mark Twain said, “Make your vocation your vacation.” Explain how this quotation relates to this article. Use specific evidence from the article to support your answer.”

Question for an OR to Sonnet 116 by Shakespeare: ““Sonnet 116” does not have a title linked to the text; rather its title distinguishes it from Shakespeare’s other sonnets. What title would you give to “Sonnet 116”? Provide evidence from the poem to support your answer.”

Question for an OR to an excerpt from *The Grapes of Wrath*: “In the last paragraph, the author writes, “Then she knew, and her control came back, and her hand dropped.” Based on the description of Ma Joad in this excerpt, explain what she knew and how that influenced her actions. Use specific information from the entire excerpt to support your answer.”

Question for an OR to an essay by Loren Eiseley: “Explain the significance of the statement in lines 82 and 83, “It was he who was civilized now,” as it applies to both the man and the dog. Use specific evidence from the essay to support your answer.

WHAT COLLEGE READINESS MAY MEAN IN GRADES 10 AND 11 ON PARCC ELA PRACTICE TESTS

1. Reading Selections on Grades 10 and 11 PARCC Practice Tests in 2015

We can get some sense of what college readiness means by examining the reading selections and writing prompts on both the computer-based and pencil/paper versions of PARCC practice tests for the PBA and EOY for grades 10 and 11 in 2015.

Grade 10, PBA

Excerpt from short story “Red Cranes,” by Jacey Choy, 2008

Excerpt from short story “The Firefly Hunt,” by Jun’ichiro Tanizaki, 1956

Excerpt from U.S. Supreme Court majority decision, written by Justice Abe Fortas, in *Tinker v. Des Moines Independent Community School District* No. 21

Excerpt from U.S. Supreme Court dissenting decision, written by Justice Hugo Black, in *Tinker v. Des Moines Independent Community School District* No. 21

Short audio clip (or transcript) of radio interview with Law Professor Catherine Ross on the impact of the decision

Excerpt from *Three Men on the Bummel*, by Jerome K. Jerome, in public domain

Grade 10, EOY

Excerpt from *The Red Badge of Courage* by Stephen Crane, in public domain

Excerpt from *Woman on the Other Shore*, by Mitsuo Kakuta, 2004

Excerpt from the short story, “A White Heron,” by Sarah Orne Jewett, in public domain

Excerpt from the speech, “The Sinews of Peace,” by Winston Churchill

Excerpt from the article, “Plastic: A Toxic Love Story,” by Susan Freinkel, 2011

Grade 11, PBA

Excerpt from *Quicksand*, by Nella Larsen, 1928

Excerpt from *Autobiography of an Ex-Colored Man*, by James Weldon Johnson, in public domain

Declaration of Independence by Thomas Jefferson, 1776

Excerpt from “Speech to the Second Virginia Convention,” by Patrick Henry, 1776

Video with music, or transcript, by the Kettering Foundation on “From Subjects to Citizens”

Excerpt from the short story, “The Overcoat,” by Nicolai Gogol, 1842

Grade 11, EOY

Excerpt from *Cranford*, by Elizabeth Gaskell, 1853

Excerpt from *Heart of Darkness*, by Joseph Conrad, 1899

Excerpt from *Frankenstein*, by Mary Shelley, 1818

Excerpt from the speech, “The Solitude of Self,” by Elizabeth Cady Stanton, in public domain

Blog post on antibiotic resistance by Beth Skwarecki, 2012

2. Writing Prompts on Grades 10 and 11 PARCC Practice Tests in 2015

Grade 10, in PBA but none in EOY

Essay 1 (literary analysis): “Though Mie and Sachiko, the main characters in the passages, have certain similarities, the authors develop their characters in very different ways. Write an essay in which you analyze the different approaches the authors take to develop these characters. In your essay, be sure to discuss how each author makes use of such elements as *the main characters’ interactions with other characters, *the presentation of the main characters’ thoughts, and *the strong feelings each character experiences at the end of each passage. Use specific evidence from both passages to support your analysis.”

Essay 2 (argument/informative/explanatory): “Consider the points made by each source about the issues surrounding the *Tinker v. Des Moines* case. Write an essay analyzing the argument of those who believe that certain kinds of speech should be prohibited within an educational setting and those who believe the opposite. Base the analysis on the specifics of the *Tinker v. Des Moines* case and the arguments and principles put forth in the three sources. The essay should consider at least two of the sources presented.”

Essay 3 (narrative): “After discovering that his wife has gone missing from the bicycle they were sharing, Mr. Harris returns “to where the road broke into four” and seems unable to remember where he has come from. Using what you know about Mr. Harris, write a narrative that describes how he chooses what road to take and the experiences he has on his return journey. Be sure to use details from the passage in developing your narrative.”

Grade 11, in PBA but none in EOY

Essay 1 (literary analysis): “...Write an essay in which you identify a theme that is similar in both passages and analyze how each author uses the characters, events, and settings in the passages to develop the themes.”

Essay 2 (argument/informative/explanatory): “An important idea presented in the three sources involves the colonists’ notion of the purpose of government. Write an essay in which you explore the perceptions of the government’s purpose presented in the sources. In writing your essay, consider how the authors of the two written documents describe the ideal relationship

between a government and its people and how they describe the actual relationship between Great Britain and the colonists. Consider also the perspective presented in the video. Remember to use evidence from all three sources to support your ideas.”

Essay 3 (narrative): “Near the middle of paragraph 1, the author describes a “young man, a newcomer” who shows sympathy for Akakiy. Write an imagined journal entry from the young man’s point of view as he reflects back on the situation later in life and the effects it has had on his life. Use what you have read in the passage to provide specific details relevant to the young man and Akakiy.

THE MEANING OF COLLEGE-READINESS IN MCAS AND PARCC

The MCAS selections reveal the strong hand of well-read high school English teachers. They also suggest the kind of curriculum that may have been in place in English classes across the state—recognized literary and non-fiction writers, black and white. The selections in grade 8 MCAS tests (not shown in this chapter) also reveal the kind of reading preparation that students need for the selections in grade 10. The questions for the ORs in grade 10 support the readings and give evidence of some coherence in the literature and reading curriculum (e.g., they call attention to authors who have long been in the high school curriculum, such as Melville and Twain).

The cumulative value of the ORs may not be readily discernable. The four questions eliciting ORs at every grade level can easily serve as models to reading or English teachers at all grade levels: (1) They compel the student to return to the text for the information needed for a response. In contrast to the open-ended composition required at only three grade levels (4, 7, and 10), they regularly demand analytical reading and content-oriented writing. (2) No more than one or two paragraphs of writing are required, lessening the anxiety many students feel when asked to write. (3) The questions never ask for personalized responses. And (4) the quality of the texts sets a standard for grade 10 reading for all students. Almost all of the texts are by

well-recognized and long-recognized authors, whether or not they are complete literary texts. Students will have two more years of high school to complete, but the overall level of authors and/or texts in these grade 10 selections indicate that high school level reading is required in high school. Students who perform at a Proficient level or above on grade 10 tests like these are college ready, as indicated in the longitudinal study prepared by the MBHE and DESE in 2008.⁶⁷

In contrast, while the overall difficulty level of the passages on the PARCC practice tests seems to be as strong as those in the grade 10 MCAS tests, and many of the selections are superb, they do not suggest the influence of high school English teachers—indeed, of high school teachers at all. The passages, mainly excerpts, sprawl across centuries and cultures, and together comprise an incoherent group of readings that straddle the history and the English class. While most of the authors should be known to all American high school students, and most passages require high school level reading, whether the actual test items in 2015 are at that level cannot be publicly verified. What is troubling is the low reading level of the novels excerpted for the PARCC grade 8 practice tests where we could find a level using ATOS for Books as the readability formula.

We found that *Confetti Girl* by Diana Lopez, © 2009, has a reading level of 4.1; *Tortilla Sun* by Jennifer Cervantes, © 2010, has a reading level of 4.0, and *Seven Keys of Balabad* by Paul Haven, © 2009, has a reading level of 5.9. The reading levels of the informational selections are all above grade 4, to judge impressionistically, skewing the ease with which students will be able to read both types of selections. But many passages cannot serve as appropriate staging for students who should be preparing for high school level reading.

So, who will be better prepared for college-level reading and writing? Those who take a grade 10 MCAS test or a grade 10 or 11 PARCC test? To better address this question, the next chapter analyzes how vocabulary knowledge is assessed

by both sets of tests through the grades before students are judged ready or not for college and career.

5. CHOICE OF VOCABULARY AND FORMAT IN PRE-COMMON CORE MCAS TESTS AND COMMON CORE-BASED TESTS

We know from a hundred years of research that knowledge of word meanings is the key component of reading comprehension.⁶⁸ So it stands to reason that a reading test should assess this component of reading instruction directly, in addition to assessing word knowledge indirectly in test items that purport to assess comprehension of selected passages. This chapter looks at the choice of words and the format for direct assessment of vocabulary knowledge in (1) pre-Common Core MCAS tests from 1998 to 2004 in grades 3, 4, 8, and 10,⁶⁹ (2) Common Core-based online practice tests in grades 3, 4, 8, and 10 in PARCC's 2015 PBA and EOY assessments (both computer- and paper-based),⁷⁰ and (3) Common Core-based online practice tests in grades 3, 4, 8, and 11 provided by SBAC in 2015.⁷¹

CHOICE OF VOCABULARY AND FORMAT FOR ASSESSMENT IN PRE-COMMON CORE MCAS TESTS

The general standard that vocabulary test items addressed in pre-Common Core Massachusetts was: "Students will acquire and use correctly an advanced reading vocabulary of English words, identifying meanings through an understanding of word relationships." No particular pedagogy was implied by this statement. As a result, the format for vocabulary test items almost consistently asked what a specific word or phrase meant in the context of its use in a reading passage—i.e., direct assessment of its meaning. Typically, the word or phrase was repeated in the question as it was used in the reading selection, or the test-taker was referred to the paragraph in which it was used. Typically, the test-taker could choose from among four options. (The format

for assessing vocabulary in the Common Core-oriented MCAS tests for grade 10 ELA in 2012 was still direct assessment.)

MCAS test questions rarely asked students to locate words, phrases, or details in the reading passage that were related to a word's meaning (i.e., as support or evidence for their answer). However, for six years (1998-2003), students were asked, in more than one item on a test, to demonstrate their understanding of the various parts of a dictionary entry for a word, the meaning of selected affixes or roots (in the elementary grades), and the distinction among the kinds of references available for determining word or phrase meanings (in the upper grades).

CHOICE OF VOCABULARY AND FORMAT FOR ASSESSMENT IN 2015 PARCC PRACTICE TESTS

PARCC claims it assesses “words that matter most in the texts, which include words essential to understanding a particular text and academic vocabulary that can be found throughout complex texts.”⁷²

PARCC explains that “Assessment design will focus on student use of context to determine word and phrase meanings.”⁷³ As a result, it uses one particular format for its vocabulary items fairly consistently. Part A of a two-part multiple-choice answer format typically asks directly for the meaning of a word or phrase, as on MCAS tests. PARCC then requires students in Part B to locate “evidence” or “support” in the text (guided by the content of the four optional answers—or Evidence-Based Selected Responses, or EBSR) for their choice of answer to Part A. PARCC consistently uses this two-part multiple-choice answer format for assessing both vocabulary and reading comprehension.

It should be pointed out that Part B is less critical than Part A for the test score. If the test-taker gets Part B correct, he must also get Part A correct to get full credit. If the test-taker gets only Part B correct, he gets no credit for Part A or B. How many children know that when they take the test is unknown.

The impetus for this design feature is most likely Common Core's vocabulary standards. Its general standard is: “Determine or clarify the meaning of unknown and multiple-meaning words and phrases based on [grade-level] reading and content, choosing flexibly from a range of strategies.” Despite the word “flexibly,” the first strategy at every grade level is always “use context as a clue to the meaning of a word or phrase,” regardless of the source of the unknown word or genre in which it is used. As PARCC indicates, “Tier III vocabulary—also referred to as domain-specific vocabulary—may also be assessed, when the meaning of the word(s) can be determined through the context of the informational text.”⁷⁴

In examining its vocabulary practice test items, we encountered a number of issues in addition to PARCC's consistent use of a misguided and misleading format for assessing vocabulary knowledge: developmentally inappropriate test directions; a puzzling choice of words for vocabulary assessment; and a mismatch between answer options and a dictionary meaning of the word. The following examples illustrate these issues.

First question set in grade 3

(For a story about animals by Thornton Burgess)

Part A. “What does *cross* mean....?” The answer to Part A has to be a synonym for *cross* in the context of this story. “Upset” is the only one that could make sense as the other options are “excited,” “lost,” and “scared.” However, that is not what *cross* means in a dictionary. Google gives us: “marked by bad temper, grumpy” or “annoyed, angry.” Moreover, a grade 3 student might have picked up some understanding of the meaning of the word from hearing about a *cross* grandmother or a *cross* look on someone's face (or from hearing Burgess stories as a pre-schooler when read to from their picture book editions). Since the three wrong choices are much farther away in meaning than is “upset,” “upset” might well be chosen as the correct answer by a process of elimination even if not quite right in the reader's own experience with the word.

The most serious issues concern the wording and meaning of the question in Part B. The question in Part B is: “Which statement best supports the answer to Part A? The correct answer is: “... hadn’t found The Best Thing in the World.” But two of the four choices, including the correct answer, are phrases, not “statements” Moreover, it is not clear what the question itself means. What can it mean to a third grader to have a question about a “statement” that “best supports the answer to Part A” (keeping in mind that the correct answer in Part A may not seem to be the correct answer to some children)? This is not child-friendly language.

A grade 3 teacher would have asked orally something like: “Why were all the animals unhappy or angry at the end of the story? Agreed, it doesn’t force the reader to go back to the story to find specific words that some test-item writer thinks “supports” the answer. But a grade 3 teacher would have been unlikely to use “supports.” If a teacher had written the test question, she might have worded the Part B question as “What phrase (or words) in the story best explains why the animals were unhappy at the end of the story?” In this case, the child doesn’t have to look back at the answer to Part A. The question is about comprehension of the story, not the question and answer in Part A. And the point is still made that the answer to the Part B question is in the text.

First question set in grade 4

(For a 2012 story about children in an elementary classroom by Mathangi Subramanian)

Part A asks for the meaning of *drift*. The correct answer is “wander.” The other choices are “hover,” “consider,” and “change.” Google gives us: “to move slowly, esp. as a result of outside forces, with no control over direction.” As in: “He stopped rowing and let the boat drift.”

Part B asks “Which detail from the story helps the reader understand the meaning of *drift*?” Alas, none of the answers is correct. The

intended correct answer is: “Lily, Jasper, and Enrique make comments about the drawings as the students come close enough to see them.” But only Lily and Jasper make comments in the story. Enrique asks a question. A careful reader would be very bothered by a poorly-worded question and no fully correct answer. These details (not just one detail, as the question implies) do not help any reader to understand the meaning of *drift*.

First question set in grade 8

(For a 2009 novel about a Hispanic American teenager by Diana Lopez)

Part A asks for the meaning of *sarcasm* as used in the Lopez novel. The correct answer is “a remark indicating mockery and annoyance.” However, Google defines the word as “the use of irony to mock or convey contempt. Synonyms: derision, mockery, ridicule, scorn, sneering, scoffing.” The word “annoyance” is not there. It is not clear why the test-writer didn’t use “contempt” instead of “annoyance”? The right answer would then have accurately pointed to the young girl’s lack of respect in speaking to her father, an attitude that helps to explain why she uses the book her father gave her as a coaster for a glass of soda. Although the right answer for another question points to resentment rather than contempt as the motivation for her behavior, her behavior may be better understood as contempt. The questions thus frame a somewhat inaccurate interpretation, confusing children who have been taught that sarcasm to parents or other elders is a sign of disrespect.

Question set for an informational article on elephants in grade 8

Confusion may also result from the answer to the Part A question about the meaning of *anecdotal observations* in an article on elephants. The only possible right answer is “a report that is somewhat unreliable because it is based on a personal account.”

The test developers had to have known they were skating on the edge of a precipice.

Personal accounts of the texts students read or of the events in their neighborhood have been pedagogically elicited for decades in efforts to “engage” students through their own daily life while, at the same time, relieving them of learning how to make supported interpretations of texts, events, people, and/or movements in place of groundless or simply emotionally-driven personal opinions—something Common Core promised to remedy.

Part B asks for the “best evidence” for this meaning in the article. But the correct answers for both Part A and Part B are misleading and scientifically wrong. A report based on an anecdotal observation is unreliable not because it is based on a personal account (which is characteristic of observation-based field reports in many disciplines) but because it has too few subjects (maybe only one, idiosyncratically chosen) and does not include a large enough random sample to serve as the basis for a defensible generalization. Thus, the only sentence that seems to make sense as the right answer for Part B (“But it’s one thing to witness something that looks like consolation, and another to prove that this is what elephants are doing.”) is highly misleading. Unreliability does not necessarily result from someone witnessing as opposed to proving something. In this article it refers to claiming that some observed animal behavior demonstrates consolation, implying that animal behavior is motivated in the same way that human behavior is. Students are misled by Part A to think that the unreliability of an anecdotal observation is a result of its being a “personal account,” not a result of a false assumption. They are then led to choose a largely wrong answer because the author of the article didn’t indicate correctly why an anecdotal observation of elephant behavior is scientifically unreliable.

First question set in grade 10

(For an excerpt from a story by an American writer about Japanese children and cranes.)

Part A asks for the meaning of *resonant*. The choices are “intense,” “familiar,” “distant,” and

“annoying.” Google offers this definition:

“(of sound) deep, clear, and continuing to sound or ring as in “a full-throated and resonant guffaw”

synonyms: deep, low, sonorous, full, full-bodied, vibrant, rich, clear, ringing; loud, booming, thunderous such as in “a resonant voice”

“(of a place) filled or resounding with (a sound) as in “alpine valleys resonant with the sound of church bells”

synonyms: reverberating, reverberant, resounding, echoing, filled as in “valleys resonant with the sound of church bells”

By a process of elimination, the word least likely to be wrong in the four choices is “intense,” even though *resonant* usually refers to the continuing nature of a sound. “Intense” is not the right answer to a student who knows from experience or reading that a resonant sound is one that continues long after the action that caused the sound.

Thus, we find no answer in the choices in Part B, which asks: What quotation from Paragraph 3 helps clarify the meaning of *resonant*? None does. The intended right answer: “they’re so loud...” doesn’t clarify the meaning of *resonant*. A more relevant choice is: “I wasn’t sure where their calls were coming from,” although it does not so much clarify the meaning of *resonant* as reflect it. (In other words, the cranes’ calls last so long that it is difficult to figure out where they actually are as they fly around.) This sentence also precedes *resonant* in the text so that if the reader doesn’t already know the meaning of *resonant* as continuing sound, then the child’s comment in the story makes little sense to the reader. In a reading lesson, the word would be one of the vocabulary items that are pre-taught before students read the selection (which is more suitable for middle than high school students).

CHOICE OF VOCABULARY AND FORMAT FOR ASSESSMENT IN 2015 SBAC PRACTICE TESTS

We explored the practice tests provided by the other consortium developing tests of Common

Core's standards to find out its format for vocabulary assessment. SBAC does not seem to assess as many vocabulary words as PARCC does. When it does assess the meaning of a word or phrase, it may ask directly for the meaning or, in a format reversal, for a word in the selection that means what the question itself provides as its meaning (see the format for stacked in grade 3 and for *whispered* in grade 4). SBAC does not use for vocabulary test items the two-part multiple-choice answer format used by PARCC for all vocabulary test items and for all other multiple-choice test items. SBAC uses the two-part format occasionally but only for other types of items.

Interestingly, SBAC does expect an advanced vocabulary to be used by teachers and test-item writers even in the primary grades and even when below-grade-level reading passages are used, and it provides a long list of such words described as “construct relevant vocabulary.”⁷⁵ This list “refers to any English language arts term that students should know because it is essential to the construct of English language arts. As such, these terms should be part of instruction.” For example, the list for grade 3 includes: *source(s)*, *specific word choice*, *spelling errors*, *stanza*, *supporting details*, *trustworthy source*, and *verb tense*. SBAC notes that these words will not be explained on the test. It expects teachers to “embed” them in instruction.

These grade-level lists for teachers and test-item writers help to explain the absence of child-friendly language in both SBAC and PARCC test items. They also send the strong message that elementary teachers are going to have to learn precisely what these terms mean and use them regularly as part of daily instructional talk. It is not at all clear where that learning is to take place. Our elementary teaching force does not normally take the kind of linguistics coursework that helps them internalize the exact meanings of many of these terms (e.g., *verb tense* and *tense shifts*).

However, while SBAC is very clear about the language that teachers and test-item writers can

use in test items and instruction, it is not clear about its criteria for choosing words to assess for student knowledge of their meanings. It is possible that the difficulty of the words chosen for assessment or for an answer option was determined by the reading level of the selections in which they were found. In addition to a readability formula, SBAC seems to be using a set of subjective variables for determining “text complexity” (e.g., knowledge demands, language features, and text structure).⁷⁶ No clear statements can be found on criteria for determining grade-level difficulty of literary or informational reading passages or for words and phrases to be assessed.

FINAL OBSERVATIONS

Format for Assessing Vocabulary

We do not know why PARCC consistently focuses on student use of context to determine word and phrase meanings.⁷⁷ Such a focus assumes context can be relied on to determine word and phrase meanings, most if not all of the time. Indeed, the assumptions seem to be that the acquisition of most reading vocabulary depends on use of context and that context is there for most reading vocabulary, in literary as well as informational text.⁷⁸ These two huge and different assumptions raise unanswered but answerable questions. First, is it the case that we learn the meaning of most new words by using context? (The general consensus is that we learn the meaning of most new words in context, a very different statement.) Second, is an informational text apt to provide a context, never mind enough context, for determining the meaning of new, domain-based words? And third, is the use of context a sound strategy to promote pedagogically for determining the meaning of unknown words in any text, regardless of genre, discipline, domain, or research evidence?

Unfortunately, no body of research shows that we learn the meaning of most new words by using context for that purpose. (We may learn them in context, but that is not the same thing as using context to learn them.) Or that most new and

difficult words students encounter in their literary reading have sufficient context (never mind any relevant context) to enable students to determine the meaning of these words from this context. Indeed, it is more likely that some understanding of the meaning of a new word helps students to understand the context. How they learned its meaning we may never find out. This seems to be the theory underlying NAEP's assessment of vocabulary knowledge (a new feature since 2009). "NAEP assesses vocabulary in a way that aims to capture students' ability to use their understanding or sense of words to acquire meaning from the passages they read....Students are asked to demonstrate their understanding of words by recognizing what meaning the word contributes to the passage in which it appears."

In sum, there is *no* research showing that sufficient context exists in literary or informational texts to justify an assessment format implying that students of any age can determine the meaning of a hard word by using its context. Worse yet, the almost exclusive use of this format may encourage teachers to teach students not to use a dictionary or other references for determining the meaning of an unknown word or phrase but to use its context instead, as if there would always be useful and sufficient context available for that purpose. It should be noted that, occasionally, PARCC's Part B question asks students to indicate a word in the options given that is opposite in meaning to the synonym answer in Part A—a small step away from the use of context.

According to one well-known reading researcher, "there are four components of an effective vocabulary program: (1) wide or extensive independent reading to expand word knowledge, (2) instruction in specific words to enhance comprehension of texts containing those words, (3) instruction in independent word-learning strategies, and (4) word consciousness and word-play activities to motivate and enhance learning."⁷⁹ What are those "independent word-learning strategies" teachers should teach? The National Reading Panel's 2000 report recommended four

"Word-Learning Strategies": 1) dictionary use, 2) morphemic analysis, 3) cognate awareness for ELL students, and 4) contextual analysis.⁸⁰ Yet, PARCC stressed only one pedagogical model, with nothing in its documents indicating that the best (though not the most efficient) way to acquire new vocabulary is through wide reading, followed by advice to teachers on ways to stimulate leisure reading. Nor did PARCC (or SBAC) in their practice tests assess dictionary skills, morphemic analysis, or cognate awareness. Assessments that influence classroom teachers to ignore the critical importance of broad independent reading, the need for specific vocabulary instruction, and the information provided in a dictionary do an enormous disservice to those children who most need to expand their vocabulary.

The pedagogy that the vocabulary standards promote, as well as the need for students to locate "evidence" in the text to show that they have determined the meaning of an unknown word from what is in the text (or could determine it if need be), more often than not seem to have led to poorly constructed test items and incorrect information on word meanings. What this pedagogical and assessment model is leading teachers to do in their classrooms as part of test preparation is unknown.

Words/Phrases Selected for Assessment

Although PARCC claims it assesses "words that matter most in the texts, which include words essential to understanding a particular text and academic vocabulary that can be found throughout complex texts," it is not apparent why many of the chosen words were selected. For example, *cross* in the grade 3 Thornton Burgess story and *drift* in the grade 4 selection about children in an elementary classroom are not important to the meaning of these reading selections for children, nor are they apt to be considered part of an academic vocabulary. The plot in these selections helps young readers to understand these words if they don't already know their meanings by the age of 8 or 9.

TABLE 4. WORDS ASSESSED ACROSS TESTS: GRADES 4, 8, AND 10/11

MCAS 1998-2002	PARCC samples 2015	SBAC samples 2015
<p>MCAS Grade 4 snapshots of our ancient past culprits, uneasy abroad massacre massive squall, Screee!/ Scree! millimeter, hydrometer microscope, wield, unfriendly, generated whoosh, enable, great vaults, anguish, bitter cold, holed up/denned up, ceased, stumpy saddle, dreadful bumbershoot unsuitable for speed clam tide, pop</p>	<p>PARCC Grade 4 drift adapted get a glimpse of channel dedicated torrential downpour disputing</p>	<p>SBAC Grade 4 flat as a pancake whispered</p>
<p>MCAS Grade 8 spoil once was quite enough vulnerable, arbiter demeanor, prematurely Charon's bark foul, chill sea dog, aloof, image elusive, urchin career, ebony disdainful, accommodate quest, wan, belt, collide forked, descry serene, detect, stout</p>	<p>PARCC Grade 8 permissive anecdotal observations cognition/cognitive strategy restrained discord niche sport established dexterous</p>	<p>SBAC Grade 8 caretaker bystanders</p>
<p>MCAS Grade 10 clamor sinuous grandiose epigrams ascent nabobs, edict, enumeration, jocund, swell, inundate, luminary, ominous, maxims, hovered, supplication, impediments, augment, simultaneous, tenaciously, trace, oblivion, seamless, repression, mercurial</p>	<p>PARCC Grade 10 permissive arrogate enunciate entranced bewilderment ostracizing totalitarian arresting litter classics uniformity crimson blotches on the pages of the past</p>	<p><i>No SBAC Practice Test in 10</i></p> <p>SBAC Grade 11 touts lethargic problem stick encore reprise mass-produced</p>

Why SBAC seems to have decided to assess few word meanings directly is not known (recall that we were looking only at its practice tests). Maybe SBAC does not view vocabulary teaching worthy of highlighting by assessments despite a major conclusion reached many years ago by researchers on how to develop students' reading and writing vocabularies: while no one method is superior to other methods, some attention to vocabulary is better than no attention.

In addition, the words selected for the high school grades in SBAC do not seem to be

advanced high school English vocabulary. Table 4 shows the words assessed on MCAS tests in its early years (when the state's English teachers clearly influenced the format and content of the test items) and in the sample test items provided by PARCC and SBAC in 2015. The grade 10 MCAS words have more of a literary flavor, which is understandable because of English teachers' stress on literary selections.⁸¹ SBAC's words for grades 8 and 11 are relatively less difficult as vocabulary items (and because SBAC does not stress word assessment, there are fewer

words chosen for assessment). PARCC's words for grades 8 and 10 are more difficult than SBAC's, but have little literary flavor, suggesting the type of selections it may stress at the high school level (we do not know what test items are on PARCC's actual tests, either). We do know that MCAS grade 10 selections in 1998-2002 were chosen mainly by high school English teachers; we do not know exactly who chose the selections or the vocabulary for PARCC's and SBAC's sample tests for high school, and the public can never see all the actual test items at the high school level.

Regardless of who chooses the reading passages, there should be a match between what is in Google's definition (which incorporates what is in major dictionaries) and the correct answer to a test question on the meaning of a word. If a passage depends on an unusual meaning for a word, the word should not be used in a test question.

Language of Assessment

It is not clear why the questions in Part B in PARCC were worded as they were. In the early grades, they do not reflect how a teacher talks. Nor were they always precise. The lists of "construct relevant vocabulary" on the SBAC website explain the presence of difficult or cumbersome terminology in PARCC and SBAC practice test items, while the use of the Part A/Part B format in both tests suggests joint planning. But we may never have any definitive body of observational research showing whether teachers embed this terminology (correctly) in their daily instruction and whether students understand them.

Nor is it at all clear why vocabulary test items so often in PARCC come at the beginning of a test or set of questions about a selection. It would seem more reasonable for questions about the theme or purpose of a passage to lead off the test questions.

6. HOW WRITING IS ASSESSED ACROSS TESTING SYSTEMS

PARCC and MCAS differ fundamentally on how best to measure writing skill and what kinds of writing should be tested. In MCAS students are tested for a "long" composition in grades 4, 7, and 10. It is administered over two untimed sessions, and the scores are derived from judges specially trained to evaluate the compositions holistically. MCAS's tests for grades 3-8 and 10 also call for students to write four short pieces of writing annually as "Open Responses" (ORs) that answer specific questions about the content or meaning of text passages they have just read. These pieces of writing (one to two paragraphs long) are also scored holistically by trained readers. These long and short pieces of writing (3 long and 28 short pieces over 8 test years) help ensure that writing skills are continuously evaluated each year.

PARCC asks students to write three compositions at every grade from 3-11 in three timed sessions for its Performance Based Assessments. PARCC claims to measure students' ability to: (1) analyze works of literature; (2) organize and evaluate given information; and (3) write "narratives." PARCC's measures of writing skills are obtained only through the PBAs, not through its End of Year (EOY) tests. In 2016, both tests will be combined and shortened. Unless there are changes, students moving through the PARCC system will produce 27 pieces of writing, or three each year for the 9 grades tested.

WRITING PROMPTS IN THE TWO TESTING SYSTEMS

Figure 3 below provides a side-by-side comparison of writing prompts (i.e., directions to elicit writing) in grades 3, 4, 8, and 10 in both testing systems. The examples for PARCC appear on the Practice Tests made available in 2015 to the public; those shown for MCAS are used items released to the public since 1998.

FIGURE 3. EXAMPLES OF MCAS AND PARCC WRITING PROMPTS IN GRADES 3, 4, 8, AND 10

MCAS	PARCC
<p>Grade 3: 2010 Open Response</p> <p>Reading: "George Washington Carver: The Peanut Scientist"</p> <p>Writing Prompt Based on the article, explain why George Washington Carver is famous. Support your answer with important information from the article.</p> <p>Grade 4: 2004 Long Composition</p> <p>Writing Prompt Think about a friend who has been an important part of your life. How did you become friends with this person? Think about when you met, what you did, and how your friendship grew. Write a story about this friendship. Give enough details to tell the reader about this friendship.</p> <p>Open Response Question Reading: "America's Best Girl"</p> <p>Explain what made Trudy's swim across the English Channel so dangerous. Use important and specific information from the article to support your answer.</p>	<p>Grade 3: 2015 Research Task</p> <p>Readings: "A Howling Success" and "The Missing Lynx."</p> <p>Writing Prompt Write an essay comparing and contrasting the key details presented in the two articles about how endangered animals can be helped. Use specific details and examples from both articles to support your ideas.</p> <p>Grade 4: 2015 PBA Narrative Task</p> <p>Writing Prompt In "Those Wacky Shoes," a girl has to outsmart a pair of shoes. Think about the details the author uses to create the characters, settings, and events. Imagine that you, like the girl in the story, find a pair of wacky shoes that won't come off. Write a story about how you find the pair of wacky shoes and what happens to you when you are wearing them. Use what you have learned about the wacky shoes when writing your story.</p>
<p>Grade 8: 2000 Long Composition</p> <p>Writing Prompt Write a persuasive essay to the school committee describing one change that will improve your school. Give at least two reasons to show how your suggestion will improve your school.</p> <p>Remember, you must argue in a convincing manner so that the school committee will understand and agree with your position.</p> <p>Open Response Question Reading: "Fire Ants on the March" The fire ant's name (<i>Solenopsis invicta</i>) means invincible or undefeatable. Why is invincible a good term to describe fire ants? Explain your answer using information from the article.</p>	<p>Grade 8: 2015 PBA Research Task</p> <p>Writing Prompt You have read three passages about studies involving the behavior of elephants: "Elephants Can Lend a Helping Trunk" "Elephants Know When They Need a Helping Trunk in a Cooperative Task" "Elephants Console Each Other"</p> <p>Write an essay analyzing each author's purpose in describing the studies of elephant behavior, and compare the information about the behavior of elephants each author presents in the passages. Remember to use evidence from all three passages to support your responses.</p>
<p>Grade 10: 2010 Long Composition</p> <p>Writing Prompt Often in works of literature, a character's life is affected by a single act or mistake. From a work of literature you have read, in or out of school, select a character whose life is affected by a single act or mistake. In a well-developed composition, identify the character, describe how he or she is affected by a single act or mistake, and explain how the character's experience relates to the work as a whole.</p> <p>Open Response Question Reading: <i>Heart of Darkness</i> (excerpt) Based on the excerpt, explain how the narrator is affected by the jungle environment. Support your answer with relevant and specific information from the excerpt.</p>	<p>Grade 10: 2015 PBA Literary Analysis Task</p> <p>Writing Prompt You have read two passages, one from Jacey Choy's "Red Cranes" and one from Jun'ichiro Tanizaki's "The Firefly Hunt." Though Mie and Sachiko, the main characters in the passages, have certain similarities, the authors develop their characters in very different ways.</p> <p>Write an essay in which you analyze the different approaches the authors take to develop these characters. In your essay, be sure to discuss how each author makes use of such elements as: the main characters' interactions with other characters; the presentation of each character's thoughts; and the strong feelings each character experiences at the end of each passage. Use specific details from both passages to support your analysis.</p>

Figure 3 shows how PARCC and MCAS follow different pathways. In PARCC, the prompts focus on what PARCC calls essay writing as early as grade 3. Each prompt establishes a formal relationship between the writer and the imagined reader, with instructions clearly signaling what must be done in order to meet the reader's expectations. In no instance is the writer encouraged to go outside the prescribed text(s). Each prompt calls upon the writer to adopt an objective stance to his/her subject matter,

and speak to an unknown audience, and understand terms typically associated with the study of English (e.g., compare and contrast, character development, point of view, narrator, author's purpose etc.). One might ask, however, "Are these the hallmarks of academic writing, or are they contrivances made to look like students are engaging in academic writing?"

In MCAS's grade 4 prompt, a very different reader-writer relationship is established. The

imagined reader is interested in what the writer personally has to say, and instructions are aimed at eliciting personal information in order to complete the composition. In MCAS prompts for long compositions in grades 8 and 10, there is an implied recognition that as students move from concrete thinking to logical reasoning, the assignment must change to reflect changes in a child's thinking, both as readers and writers. In grade 8, we see an argument-based assignment with purpose and audience given; in grade 10, we see a formal reading-based assignment. MCAS's approach, it can be said, is developmental, PARCC's prescriptive.

STRUCTURE

MCAS long compositions deliberately make no reference to specific texts (suggesting there is no one curriculum across school districts), and provide students considerable freedom in choice of text when writing. In contrast, as shown in Figure 3, PARCC prompts are highly structured and part of a larger assessment format designed to link reading and writing. The format is consistently presented in four parts: Test takers must: (1) read instructions and information about what they will write about; (2) read two or more specific texts of varying length; (3) answer multiple-choice or two-part questions on the readings; and (4) write a "text based response" (PARCC terminology), based on the texts they have just completed. All of the parts fit together as a "unit."

However, PARCC prompts, when separated from the other parts of the unit test, only seem to be "rigorous." In the grade 3 prompt in Figure 3, for example, writers are asked to complete five distinct tasks: (1) identifying the key details in two texts; (2) organizing these details into two categories in order to compare and contrast the writers' use of them; (3) determining how these uses are different and alike; (4) deciding which group of details and examples to focus on first; and then (5) writing a typical "comparison/contrast" essay that centers on how the two texts differ, all in a timed situation. For developing

writers, as cognitive psychologists might observe, the information processing load of the task surpasses what can realistically be understood, much less executed in writing.⁸²

The same can be said of a grade 5 prompt in another PARCC Practice Test:

You have read three articles about penguin rescue efforts after an oil spill.

* From "The Amazing Penguin Rescue" by Lauren Tarshis

* "The Amazing Penguin Rescue" by Dyan DeNapoli

* "Update on Penguin Efforts from an Oil Spill in South Atlantic"

Write an essay explaining the similarities and differences in each article's point of view about penguin rescue efforts after an oil spill. Support your essay with information from all **three** sources.

Think about what the reader-writer must do to successfully respond to this prompt: First, he or she has to understand what the assignment is calling for. This, in and of itself, is no mean feat, given the fact that "point of view" is a literary term typically used to discuss a narrator's role in fiction but utilized here for an author's stance in an informational article. The objectives of the assignment are: (1) to determine the "*article's*" point of view in three separate texts after being told throughout the unit that there are three different points of view (one imaginative, the other two first-person eyewitness); (2) to identify relevant details from each text to show and explain what this point of view is; and then (3) to explain how each is similar to and different from the others. The degree of planning and re-reading needed to compose an essay like this under timed conditions is altogether staggering, even when, in the computerized version, the writer is given assistance in planning by being given textual details that have to be organized by a drag-and-drop function.⁸³

PARCC prompts fail to meet what Gertrude Conlon suggests are basic criteria for effective essay questions. Effective prompts, she

suggests, should be: (1) “clear” enough that a test taker “should not have to puzzle over the instructions”; (2) as “brief as clarity allows”; (3) balanced enough so that “average students should be able to write average answers to the question and...bright students... able to show their brightness;” (4) focused by an “organizing principle” to point writers toward the features of the essay that evaluators will expect to see; and (5) simply written, using “vocabulary and concepts that are not too difficult for the ordinary student to understand immediately.”⁸⁴ To judge with these criteria, PARCC’s format and elaborate directions for writing usually violate basic expectations for an effective prompt. Requiring students to perform multiple tasks as a precondition for writing is not rigor at all, but rather a way to confound, unintentionally, what is being measured in the first place. Understanding directions, not writing, becomes the objective.

Finally, if we consider how MCAS and PARCC address these differences—in grade 3 and elsewhere in Figure 3—it can easily be argued that the 28 ORs used throughout MCAS are as powerful, if not better placed developmentally, as PARCC tools for measuring both reading comprehension and writing. MCAS ORs are often linked to sophisticated reading passages drawn from the best works of English, American, and World literature, and they are not dependent on elaborate formats like those used for PARCC to integrate reading and writing activities. MCAS ORs seem to provide a more efficient way to gauge and stimulate intellectual development than PARCC’s long, complicated units and prompts. PARCC prompts in grade 3 are not much different from those in grade 10, suggesting little understanding of developmental differences over time. The creators of PARCC writing prompts seem more focused on making the prompts complicated than on creating challenging but suitable assignments.

ACCESSIBILITY

Smith and Swain argue that writing tests are best served by prompts that are both

“accessible” and grounded in “authentic” forms of communication.⁸⁵ Accessible and authentic prompts (1) motivate and interest students; (2) suggest or imply a real or imagined audience that cares about what is said; (3) allow all students to draw upon personal knowledge that is uniquely theirs; and (4) are sufficiently open-ended to allow writers to choose *what* to write and in what *form*. Choice is especially important because it makes it possible for students to write about what they know as individuals and to create responses that appeal to their personal likes and dislikes and their acquired knowledge. Accessible prompts engage students’ curiosity and creativity in formulating an appropriate, authentic response to the topic or question posed.

The MCAS prompts in Figure 3 are highly accessible, allowing students to draw upon their personal knowledge and prior experience. Accessibility here does not mean writing that is self-absorbed or centered entirely on the writer, as is often the case with expressive discourse or free-writing, but rather writing that taps into what a writer thinks as an individual. The analytic essays required in Grade 10, for example, consistently encourage students to write about something they have read in or outside school. The same cannot be said of PARCC’s grade 10 prompt in Figure 3 or others in the Practice Tests. All of PARCC’s prompts require the use of a “formal register” deliberately designed to mirror the “objectivity” of academic language used in college. This single difference demonstrates how PARCC and MCAS pursue markedly different paths to achieve their goals. PARCC prompts are so tightly linked to a narrow range of texts in grade 11 that their lack of flexibility may leave students unmotivated to write much if at all. Voice, authenticity, personal experience, even style are not priorities in PARCC.⁸⁶ The writing is strictly utilitarian.

TYPES OF WRITING

Both PARCC and MCAS have designed their prompts to measure writing taught in schools—stories, expository essays, literary analysis, etc.—

and used in nearly all large-scale assessments. These types of writing mix the terminology of classical rhetoric (i.e., the “modes of discourse:” *narration, description, exposition, and argument*) with such terms as *persuasion*, frequently ascribed to Aristotle or, more recently, to discourse theorists such as James Kinneavy⁸⁷ or James Britton.⁸⁸ Research on written communication shows no clear lines between different modes of discourse or “domains,” be they “transactional or poetic,” “expressive,” “persuasive” or “referential” or simply “fiction” or “non-fiction.” The terms used to classify types of writing are still ill-defined despite the best efforts of scholars seeking to untangle what James Moffett described as the “universe of discourse.”⁸⁹ Few high school English teachers today may even recognize the works of these scholars.⁹⁰

Of special interest, therefore, is how PARCC test designers have attempted to integrate past and modern rhetorical terms. PARCC divides writing performance into three broad domains—literary analysis, research “simulation,” and narrative writing—which serve as the basis for the three prompts in each grade from 3-11. From a design perspective this framework maps onto the areas of writing specified in Common Core. But in creating this framework, PARCC test designers had to eliminate other areas that may be as important as, if not more important than, the three they have chosen, such as persuasive writing. But are the domains chosen essential categories from which to measure the skills 11th graders need or have before entering college? Tom Newkirk, an English professor, certainly does not think so, and, in fact, he outlines how, in very fundamental ways, Common Core and PARCC have failed to get their taxonomy right and thereby distort what we should be teaching in our schools today.⁹¹ We comment on its taxonomy in what follows.

LITERARY ANALYSIS

Both MCAS and PARCC agree on the value of literary analysis or analytical essays as an important area of writing to measure. Writing

analytically about literary and non-literary texts is an integral part of every school curriculum and a significant part of what 11th graders will need to be able to do well if they are to be successful in college. Analytical writing is unquestionably an essential domain to test. But PARCC and MCAS differ on when and how the skills entailed by literary analysis are tested. PARCC begins literary analysis in grade 3; MCAS does not. MCAS looks only to see how well students can perform on an “open response” item. Why? To a large degree the answer brings us back to the differences between a developmental and prescriptive approach, discussed above, and to whether the analysis of literary texts might best begin in secondary schools when students are well beyond the first stages of reading.

MCAS is closely tied to the importance of reading in the early grades and the emphasis that the state’s previous curriculum framework placed on developing strong readers before they embark upon the formal analysis of what they read. The state’s pre-Common Core curriculum framework for ELA presented carefully sequenced standards that expected children to learn to read fluently by grade 4, and then to use this foundational skill to read to learn the rich subject matter to be studied in the years ahead. These distinctions—between learning to read and reading to learn—were first described by education researcher Jeanne Chall in 1979.⁹²

Today, they are widely accepted descriptions of the early stages children go through before attaining full competency as readers by the time they enter college. The state’s 2001/2004 ELA curriculum framework incorporated Chall’s observations into a fully elaborated scope and sequence of preK-12 skills outlining what children must know and be able to do as readers and writers over the course of twelve years of public school. Based on the Common Core, PARCC seems to have pushed past the evidence and is, instead, presenting complex tasks that have little to do with children’s developing abilities as readers or writers.

RESEARCH “SIMULATION”

MCAS makes no pretense of measuring research skills even though standards for the research process are quite visible in the state’s 2001/2004 ELA curriculum framework. In fact, research projects were explicitly left to the local level for assessment. But what is meant by PARCC’s research “simulation” writing tasks?⁹³ The term warrants skepticism when no open-ended controlling idea is ever asked for. It is also misleading to call its syntheses of given information a simulation of a research process. Synthesis might well be considered the least important aspect of the process. Do these tasks capture the essential skills students must learn if they are to do the type of research required in college or, for that matter, in a business organization? The solitary, painstaking work of developing an open-ended research question (and revising it regularly), identifying relevant sources, synthesizing studies, organizing one’s findings, and then writing a coherent report are not what is being measured by PARCC, even though so-called research simulations constitute the longest and largest section of PARCC tests in each grade.

In fact, it is doubtful that even the term “essay” as used at all grade levels and for all types of writing by PARCC is an appropriate word to use for the kind of referential, information-based writing required in college. No prompt can validly or practically elicit what it takes to write a research paper or report (which is why the standard on writing a research paper in the pre-2011 Massachusetts ELA standards was left for assessment at the local level). Research simulation (or argument) is an inappropriate designation of the competencies needed to write research papers when the writing expected for research is largely, if not entirely, general exposition.⁹⁴ The only tangible difference between the prompts used for literary analysis and research simulation in PARCC seems to be whether the subject matter is fictional or factual.

If PARCC cannot validly measure research skills through the prompts it has developed, are all

of the essay questions simply a waste of time? Perhaps how students write about historical and scientific texts is useful information to gather, but where is the evidence that the ability to write expository prose in these areas (not the writer’s content knowledge) is what is needed for success in a college history course or an introductory course in, say, ecology? We have found none.

It is also striking that PARCC’s writing tests pay so little attention to persuasive writing when, according to its own literature, Common Core’s standards “put particular emphasis on students’ ability to write a sound argument on substantive topics, as this ability is critical to college and career readiness.” Nearly all of PARCC’s argumentative writing assignments are essays designed to convey information based on limited evidence or carefully selected examples rather than a logical argument to persuade an audience to adopt a particular point of view. Authentic argumentative writing, from what we observe in its Practice Tests, is largely unelicited despite its importance as an indicator of college readiness.

NARRATION

Narrative writing is typically taught in grades K-8, because students enjoy writing stories. This is undoubtedly one of the reasons that PARCC has selected story-writing for its grade 4 prompt as well as for eight other PBA prompts. PARCC’s decision to allocate this much space and time to administer and score nine narratives poses a question similar to the one raised about simulated research prompts. If measuring college preparedness by the 11th grade is, in fact, the central goal of PARCC, why is narration singled out as a domain of discourse?

This is not hair-splitting. Newkirk rightly observes that the underlying structure of all written communication is narration. So why classify narration as a “type” when so many other critical choices could have been made? The issue is how to prioritize important domains of writing to the developmental paths children normally follow as readers over time. PARCC’s classification scheme grows directly out of

Common Core's theory of text complexity outlined in its Appendix A. But as Newkirk also rightly observes, the map is a creation of a top-down and backwards design that starts with what 11th graders are supposed to know and then moves down, grade by grade, until we get to grade 3 where students are called upon to compose a compare-and-contrast essay. This method results in an artificial progression from 3-11 that strives to accelerate writers to only one important destination—college readiness. There are many important milestones along the K-12 continuum. In the end, PARCC has created a large number of prompts that fail to capture essential writing skills from K-12.

FINAL OBSERVATIONS

So, will MCAS or PARCC better prepare high school students for college reading and writing? Based on the foregoing analysis, it would appear that pre-2011 MCAS is more clearly focused on age-appropriate tasks aligned to what can be taught in K-12 curricula. PARCC addresses the standards it is based on, and its writing assignments highlight Common Core's priorities and very real shortcomings. Neither testing system captures the full measure of skills and abilities needed to write well in college. MCAS looks at writing as it develops up until 10th grade but cannot delineate the essential components of what it means to be a strong writer in college. PARCC attempts to do so by testing the so-called learning progressions outlined in Common Core. But its effort is a failure for different reasons.

PARCC has created an extensive system to measure writing ability but its tests are cumbersome, artificially academic, often inaccessible, and based on writing types that are of limited value for developing and assessing the skills 11th graders must have for college coursework. In addition, its writing prompts are mismatched to the developmental capacities of young writers. It is doubtful if PARCC, as currently designed, can yield little more than a superficial understanding of the writing skills

K-12 students need because these skills are so greatly influenced by age, gender, the rhetorical demands of the writing prompt, and especially the writer's reading level.⁹⁵

Over the years, many scholars have attempted to study how children's reading and writing abilities develop and pattern over time, revealing different milestones and stages along the way. For example, Marie Clay documented the patterns of growth of emergent readers;⁹⁶ Jeanne Chall the stages of growth in reading;⁹⁷ Walter Loban the stages of growth in oral language;⁹⁸ and Sandra Stotsky the gradual acquisition of vocabulary needed for writing.⁹⁹ Donald Graves opened a new window on how young children learn to write.¹⁰⁰ Yet another study showed how writing tasks given to 4th and 6th graders produced surprisingly mixed and different scores for boys and girls in both grades, and how expressive, referential and persuasive tasks led to differing scores based on the apparent difficulty of the three modes.¹⁰¹ Mark McQuillan obtained similar results in his study of writers in grades 7-9 and included measures of students' reading levels to determine their impact on students' writing scores.¹⁰²

What this short review of the research literature suggests is that neither PARCC nor MCAS measures the broad set of writing skills that underlie growth over time. MCAS comes closer, but what we can learn about writers from MCAS stops at 10th grade. PARCC ends with an assessment of 11th grade skills, but the fatal flaw with PARCC ELA exams is its dependence on Common Core, where writing is portrayed as a unitary phenomenon, if not in theory, then in practice. Writing abilities are far more complex and nuanced than Common Core acknowledges or PARCC assessments allow. We are left with a rigid set of writing assignments organized in such a questionable way that they cannot tell us whether 11th graders can write well enough for a high school diploma or college.

7. HOW TO PREPARE BAY STATE STUDENTS FOR AN ACADEMICALLY MEANINGFUL HIGH SCHOOL DIPLOMA OR COLLEGE: CONCLUSIONS AND RECOMMENDATIONS

We can sum up many of the differences between PARCC and MCAS as testing systems by highlighting the mindset seemingly behind each one. It is clear that in MERA the legislature intended a discipline-based array of tests, one for each of the major subject areas in the school curriculum.¹⁰³ Only four sets of standards ever got tested (English language arts, mathematics, science and technology/engineering, and history/social science), even though standards were developed for all seven subject areas mandated by MERA (no tests were developed for health, foreign languages, or the performing arts). But a wealth of used test items are available for researchers to scrutinize in order to determine what the test items themselves may have contributed to improved classroom instruction and “the Massachusetts education miracle.”¹⁰⁴ The mindset guiding PARCC as a testing system may be described as skills-oriented, in contrast to the discipline-oriented focus of MCAS.

Clearly, the gains in academic achievement by Bay State students since the mid-2000s may reflect the changes in teacher/administrator licensing regulations and teacher licensing tests implemented after 2000.¹⁰⁵ But few would deny that the focus as well as the quality of the MCAS tests students took since 1998 contributed in no small part to the increases in achievement. The big question is how they might have done so and whether PARCC and the standards it is based on are capable of moving Bay State students farther along the academic highway and at a faster pace than MCAS did or whether they may halt or retard the growth MCAS stimulated in all students and accelerated in low-income students.

The Board of Elementary and Secondary Education is to decide officially in late fall of

2015 whether to replace an effective state-owned testing system created and vetted by Massachusetts educators with a problematic privately-owned testing system created and vetted by unknown others outside of Massachusetts. Puzzlingly, local superintendents have been led to believe that the transition to PARCC is already a *fait accompli*¹⁰⁶ despite the absence of evidence that students who took Common Core-based PARCC tests in 2015 made greater gains than those who should have been able to take MCAS tests in 2015 based on the standards and goals for which it was designed. If DESE in 2010 had allowed a number of randomly chosen representative schools to continue with pre-Common Core standards and MCAS tests in mathematics and ELA and an equal number of randomly chosen representative schools to move directly to Common Core standards and PARCC tests, with five years of implementation time to prepare their teachers and students, we might have had results in 2015 that could answer the big question: Which testing system, based on the standards and goals for which it was designed, produces better academic results, judging by TIMSS, the one test that is independent of the USED and reflects the school curriculum? But DESE, BESE, and then Secretary of Education Paul Reville did not have the foresight in 2010 to set up the kind of comparison needed by 2015.

They did, however, sign a Memorandum of Understanding in 2010 that committed members of PARCC to, among other things, the following obligations in PARCC’s application for USED funds:

- To provide assessments and results that: (1) are comparable across states at the student level,
- (2) meet internationally rigorous benchmarks,
- (3) allow valid measures of student longitudinal growth, and (4) serve as a signal for good instructional practices.

For this White Paper, we have examined used MCAS test items in English Language Arts from 1998 on, practice test items posted online by PARCC in 2015, as well as many documents, research studies, reports, and other

material that could inform our judgment about MCAS and PARCC as testing systems in ELA and mathematics. As the evidence now suggests, PARCC has not met even the four basic obligations enumerated above. First, the number of states now left in PARCC means the Bay State can compare itself to only eight states, assuming that kind of information is useful to Bay State teachers. Second, as shown in Chapter 2, Common Core-based tests do not meet internationally rigorous benchmarks for high school, by definition. Third, PARCC writing prompts do not allow for valid measures of student growth, and fourth, as Chapter 5 demonstrates convincingly, PARCC tests do not model sound practices for teaching vocabulary knowledge or developing young writers. PARCC tests actually model ineffective pedagogical practices that have already been studied and found wanting.

We describe PARCC's notable flaws in more detail below:

1. Most PARCC writing prompts do not assess the kind of writing done in college or the real world of work. The “narratives” (most of which is creative writing) are curriculum-relevant chiefly in the early grades and are not desirable in college or the world of work, while the “simulated research” tasks do not address the most important skills needed for a research project and college writing—finding a researchable topic and relevant sources.

2. PARCC uses a format for assessing word or phrase knowledge that seriously misleads the state's teachers. There is little research to support the use of context to determine the meaning of a word, according to the National Reading Panel report in 2000. NAEP itself asks students “to demonstrate their understanding of words by recognizing what meaning the word contributes to the passage in which it appears.” It does not ask students to figure out what the passage contributes to the word's meaning. An assessment format for vocabulary knowledge needs to accord with 100 years of research evidence.

3. PARCC's computerized tests have not shown more effectiveness than paper-and-pencil-tests or a return of useful information to the teachers of the students who took the tests. Why must the state's assessment system be computerized when a paper-and pencil-test may be more effective and cheaper than a computerized system? Moreover, there is no indication that a computerized system can return useful information to the teachers of the students who took the test.

4. PARCC uses “innovative” item-types for which no evidence exists to support claims that they tap deeper thinking and reasoning as part of understanding a text. PARCC has presented no independent evidence that its “innovative” test items known as ESBRS and TERs can promote deeper thinking and reasoning. They also waste instructional and testing time.

5. PARCC tests require too many instructional hours to administer and prepare for. They also do not give enough information back to teachers or schools to justify the extra hours or costs. MCAS appears to accomplish as much as PARCC in many fewer hours of preparation and test time and at a much lower cost to the school districts.

6. PARCC test-items do not use student-friendly language and its ELA reading selections do not look as if they were selected by secondary English teachers.

WHAT MUST BE DONE?

PARCC's inadequacies as a test can be traced to the design features specified by USED, as well as to the dozens of questionable choices made by PARCC's assessment development team and governing board. At the root of PARCC's weaknesses are the Common Core standards. In addition to being inadequately supported by research, they require states to pursue spurious, ill-defined goals. As a first step to repairing the damage, BESE must do three things: (1) phase out use of Common Core standards and discontinue use of PARCC by 2018, (2) develop a MCAS 2.0 beginning with the state's pre-Common Core curriculum frameworks,

updated by any pertinent new research, and (3) build a coalition with representatives from the legislature, higher education, and the governor to implement the changes needed to restore the excellence of MCAS.

Within this framework, MCAS 2.0 would incorporate the following policy changes:

1. Implement MCAS testing only in grades 3, 4, 8, and 10.
2. Continue with grade 10 MCAS tests for the high school diploma, but require relevant academic teaching faculty (in mathematics, science, literature, and composition) in state higher education institutions to review and publicly report on before they are given.
3. Require a mix of long compositions and questions for open response (ORs) at every grade level tested in ELA and short-answer and open-responses for tests in history, science, and mathematics.
4. Use Massachusetts high school English teachers to choose and review reading passages and questions and to design and evaluate student writing; these functions should not be jobbed out to non-teachers, nor should essay tests be consigned to computerized evaluations.
5. Require that cut scores for all performance levels on all state tests be set by Massachusetts teachers.
6. Require DESE to discourage teachers' reliance on the use of context to figure out the meaning of new words and to encourage teachers to instruct students in the use of glossaries for the precise meaning of technical terminology in mathematics and science textbooks.
7. Establish a junior/senior-year interdisciplinary research paper requirement as part of the state's graduation requirements—to be assessed at the local level following state guidelines—to prepare all students for authentic college writing.
8. Eliminate online testing; computerized testing has not demonstrably improved test quality or led to a more expeditious return of useful data to educators. Handwriting skills need to be taught and stressed because they are related to correct spelling, beginning reading, and other skills needed for writing.
9. Require all scored test items to be released every year to serve the diagnostic purposes required by MERA.
10. Provide realistic expectations for touch typing and instructional uses of technology in K-12, as MCAS 2.0 incorporates affordable and desirable technical tools for assessment.

In addition to these specific recommendations to the state's secretary of education and board of elementary and secondary education, we recommend that the Board of Higher Education disallow students from taking credit-bearing college freshman math or science coursework without an advanced math course in grade 12 or a college placement test.

LACK OF PUBLIC CONFIDENCE IN THE DEPARTMENT OF ELEMENTARY AND SECONDARY EDUCATION

What drives academic achievement and a meaningful high school diploma? There are no independent analyses by undergraduate teaching faculty in mathematics, science, and English to indicate that Common Core-based tests in grade 10 or 11 indicate readiness for authentic college coursework. Even if there were, a state testing system needs to do more than drive continuous improvement in academic gains among all students, especially low-achieving students, as MCAS has done for over a decade. The public needs to have confidence that the testing system not only uses evidence-based types of test items for assessing reading, writing, and mathematics, but is also in the hands of an agency responsive to the concerns of the teachers and parents of the children in its public schools. That confidence in DESE was rarely heard in the testimony given by

parents and teachers at the five public hearings in 2015 (see Appendix B). That confidence clearly deserves to have been shattered by poorly constructed MCAS tests in grade 10 in recent years and misleading results.

WHY LACK OF PUBLIC CONFIDENCE IS DESERVED

As the MBAE report noted, the most recent MCAS tests in grade 10 used items that addressed standards far below a grade 10 level. Independent corroboration can be seen in the percentage of students in the Bay State scoring

Proficient and above on the 2013 NAEP Pilot Test in mathematics (34%) and in ELA (43%) in grade 12, and the much, much higher percentage of students in the Bay State scoring Proficient and above on the 2013 MCAS test in mathematics (80%) and in ELA (91%) in grade 10.¹⁰⁷ There has been little discrepancy between comparable percentages in grades 4 and 8 and, in fact, Massachusetts has been commended in the past for having performance level percentages on state tests that correspond to its performance level percentages on NAEP—the yardstick for all states since 2002. But to go from 80% Proficient

TABLE 5. MCAS PERFORMANCE PERCENTAGES FROM 2005-2012 IN MATHEMATICS IN GRADES 6, 7, 8, AND 10

GRADE 10

ACHIEVEMENT LEVEL	2005	2006	2007	2008	2009	2010	2011	2012
ADVANCED	35	40	42	43	47	50	48	50
PROFICIENT	27	27	27	29	28	25	29	28
NEEDS IMPROVEMENT	24	21	22	19	18	17	16	15
FAILING	13	12	9	9	8	7	7	7

GRADE 8

ACHIEVEMENT LEVEL	2005	2006	2007	2008	2009	2010	2011	2012
ADVANCED	13	12	17	19	20	22	23	22
PROFICIENT	26	28	28	30	28	29	29	30
NEEDS IMPROVEMENT	30	31	30	27	28	28	27	28
WARNING	30	29	25	24	23	21	21	19

GRADE 7

ACHIEVEMENT LEVEL	2005	2006	2007	2008	2009	2010	2011	2012
ADVANCED		12	15	15	16	14	19	20
PROFICIENT		28	31	32	33	39	32	31
NEEDS IMPROVEMENT		33	30	29	30	27	27	30
WARNING		28	24	24	21	19	22	18

GRADE 6

ACHIEVEMENT LEVEL	2005	2006	2007	2008	2009	2010	2011	2012
ADVANCED	17	17	20	23	24	27	26	27
PROFICIENT	29	29	32	33	33	32	32	33
NEEDS IMPROVEMENT	30	29	28	26	27	25	25	24
WARNING	23	25	20	18	16	16	16	16

and above in grade 10 MCAS math in 2013 (and 91% Proficient and above in grade 10 MCAS ELA in 2013) down to 34% Proficient and above in grade 12 NAEP math in 2013 (and 43% Proficient and above in grade 12 NAEP ELA in 2013) requires an explanation from DESE or BESE or the secretary of education, and none has been forthcoming.

Those responsible for the undemanding quality of the grade 10 MCAS tests in recent years were able to evade accountability in part because full

public scrutiny of all test items was not possible. But it seems accountability was evaded mainly because, as the 2015 MBAE report noted, the “Proficient bar on the MCAS high school tests is set very low compared to all other indicators of students’ college- and career-readiness.” A low bar for Proficient means that the number of points needed for Proficient was set (originally or at a later time) below what “proficient” should mean in grade 10 math and ELA. A low bar can also reflect test items that are too easy for grade

TABLE 6. MCAS PERFORMANCE PERCENTAGES FROM 2005-2012 IN ENGLISH LANGUAGE ARTS IN GRADES 6, 7, 8, AND 10

GRADE 10

ACHIEVEMENT LEVEL	2005	2006	2007	2008	2009	2010	2011	2012
ADVANCED	23	16	22	23	29	26	33	37
PROFICIENT	43	53	49	51	52	52	51	51
NEEDS IMPROVEMENT	26	24	24	21	15	18	13	9
FAILING	9	7	6	4	4	4	3	3

GRADE 8

ACHIEVEMENT LEVEL	2005	2006	2007	2008	2009	2010	2011	2012
ADVANCED		12	12	12	15	17	20	18
PROFICIENT		62	63	63	63	61	59	63
NEEDS IMPROVEMENT		19	18	18	15	16	15	14
WARNING		7	6	7	6	7	6	6

GRADE 7

ACHIEVEMENT LEVEL	2005	2006	2007	2008	2009	2010	2011	2012
ADVANCED	10	10	9	12	14	11	14	15
PROFICIENT	57	55	60	57	56	61	59	56
NEEDS IMPROVEMENT	27	26	23	23	23	21	21	21
WARNING	7	9	8	8	7	7	6	7

GRADE 6

ACHIEVEMENT LEVEL	2005	2006	2007	2008	2009	2010	2011	2012
ADVANCED		10	9	15	16	15	17	18
PROFICIENT		54	58	52	50	54	51	48
NEEDS IMPROVEMENT		28	25	24	24	21	23	22
WARNING		8	7	8	9	9	9	11

10, as the MBAE report indicated. Why the bar is set so low the MBAE did not explore or explain. Perhaps the Fordham et al report will. Appendix C shows a randomly chosen test item from the 2014 grade 10 math MCAS so that readers can see what a test item far below grade 10 level looks like.

The use of below-grade 10 test items was not a one-test or one-year phenomenon. As Tables 5 and 6 show, the remarkable rise in percentages in Proficient and Advanced in grade 10, in math especially, began in 2008, so that by 2013 the percentage of students judged to be Proficient and above was much, much higher on grade 10 MCAS than on NAEP's grade 12 tests. While Bay State high school students did very well compared with their peers in other states in mathematics and ELA, they did not do nearly as well as their MCAS scores suggested.

If BESE simply chooses to replace MCAS with PARCC, it will not be addressing the growing lack of parent and teacher confidence in Common Core's standards and the tests based on them. At present, PARCC cannot ensure the integrity of a test for readiness for college or career, or for a high school diploma, for that matter. PARCC does not plan to release all test items used for college readiness (and other grades) and because it is now a private entity, it cannot be required to provide documentation of its decision-making process for test-item use or the cut scores used for each performance level, but especially at the high school level. In contrast, DESE released all used test items from 1998-2007 and about half after that date. Massachusetts high school teachers set the performance levels for all the grade 10 MCAS tests and because so many of them taught grade 11 or 12, they have had every incentive to want a high standard for passing.

RECOMMENDATIONS

Based on all that we have examined, coupled with our concerns over what will be lost if MCAS is abandoned, our final remarks boil down to two central recommendations:

- (1) that Massachusetts use a testing system for K-12 that is much less costly, more rigorous academically, and much more informative about individual student performance, and with much less instructional time spent on test preparation and administration, than the current PARCC tests. Both the PARCC and current MCAS tests are weak, albeit for different reasons, and cannot indicate eligibility for a high school diploma, college readiness, or career readiness.
- (2) that BESE reject the PARCC assessment system and vote for continuation of the MCAS system on the condition that the responsibility for developing and administering K-12 standards and tests be assigned to an organization in Massachusetts independent of DESE and the state's education schools. This organization must focus squarely on providing the best possible content standards from disciplinary experts in the arts and sciences and engineering throughout the state and be capable of providing oversight of high school standards and tests.

If carried out, these recommendations will ensure the legacy and future promise of MERA.

About the Authors

Mark McQuillan is former dean of the School of Education at Southern New Hampshire University. He served as commissioner of education in Connecticut (2007-2011), deputy commissioner of education in Massachusetts (2002-2004), and superintendent of schools for ten years in Massachusetts. He began his career as an English teacher in the Newton Public Schools.

Richard P. Phelps is editor or author of four books: *Correcting Fallacies about Educational and Psychological Testing* (APA, 2008/2009); *Standardized Testing Primer* (Peter Lang, 2007); *Defending Standardized Testing* (Psychology Press, 2005); and *Kill the Messenger* (Transaction, 2003, 2005), and founder of the Nonpartisan Education Review (<http://nonpartisaneducation.org>).

Sandra Stotsky is professor *emerita* at the University of Arkansas where she held the 21st Century Chair in Teacher Quality. She was senior associate commissioner in the Massachusetts Department of Elementary and Secondary Education (1999-2003) and a member of the Common Core Validation Committee (2009-2010).

About Pioneer

Pioneer Institute is an independent, non-partisan, privately funded research organization that seeks to change the intellectual climate in the Commonwealth by supporting scholarship that challenges the “conventional wisdom” on Massachusetts public policy issues.

Recent Publications

Modeling Urban Scholarship Vouchers in Massachusetts, White Paper, July 2015

Federal Overreach and Common Core, White Paper, July 2015

The Pacheco Law Has Cost the MBTA More than \$450 Million, White Paper, July 2015

Mass Hospitals Weak on Price Transparency, Policy Brief, June 2015

MBTA Reform – The Case of Full, Final and Binding Interest Arbitration, June 2015

Support and Defend: The K-12 Education of Military-Connected Children, White Paper, June 2015

Analyzing the Convention Center Authority's Inflated Claims, Policy Brief, June 2015

Why Massachusetts Should Abandon the PARCC tests and the 2011 Coleman et al English Language Arts Standards on which the MCAS Tests are Based, Public Testimony, June 2015



Endnotes

1. It is not clear if post-2011 MCAS tests have also assessed the standards BESE voted to add in the fall of 2010. At the time, BESE voted to add a few standards in both subjects as part of the extra 15 percent allowable by the Common Core project. Thus, the state's 2011 Mathematics and English Language Arts Curriculum Frameworks contain all of Common Core's standards and the (few) mathematics and ELA standards voted for addition by BESE at the commissioner's recommendation. It was understood from the start that PARCC would not assess standards that states added after adopting Common Core's.
2. <http://www.mbae.org/wp-content/uploads/2015/02/MBAE-MCAS-PARCC-Report-Web.pdf>
3. Nichols-Barrer, I., Place, K., Dillon, E., & Gill, B. (2015, October 5). *Final Report: Predictive Validity of MCAS and PARCC: Comparing 10th Grade MCAS Tests to PARCC Integrated Math II, Algebra II, and 10th Grade English Language Arts Tests*. Cambridge, MA: Mathematica Policy Research. Shortcomings:

First, the report attempts to calculate only general predictive validity. The type of predictive validity that matters is “incremental predictive validity”—the amount of predictive power left over when other predictive factors are controlled. If a readiness test is highly correlated with high school grades or class rank, it provides the college admission counselor no additional information. It adds no value. The real value of the SAT or ACT is in the information it provides admission counselors above and beyond what they already know from other measures available to them.

Second, the study administered grade 10 MCAS and PARCC tests to college students at the end of their freshmen years in college, and compared those scores to their first-year grades in college. Thus, the study measures what students learned in one year of college and in their last two years of high school more than it measures what they knew as of grade 10. The study does not actually compute predictive validity; it computes “concurrent” validity.

Third, student test-takers were not representative of Massachusetts tenth graders. All were volunteers; and we do not know how they learned about the study or why they chose to participate. Students not going to college, not going to college in Massachusetts, or not going to these colleges in Massachusetts could not have participated. The top colleges—where the SAT would have been most predictive—were not included in the study (e.g., U. Mass-Amherst, any private college, or elite colleges outside the state). Students not going to college, or attending occupational-certificate training programs or apprenticeships—for whom one would suspect the MCAS would be most predictive—were not included in the study. <http://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>
4. See <http://blog.socrato.com/partial-release-of-mcas-test-items-is-a-step-backwards/>. The 2015 MBAE report also notes: “Releasing all items on all tests each year for the first nine years of the program, and continuing to release half of items in subsequent years” (p. 4).
5. <http://edexcellence.net/publications/stars-by-which-to-navigate.html>;
<http://www.brookings.edu/research/papers/2013/01/09-timss-pisa-loveless>
6. National Governors Association Center for Best Practices, and Council of Chief State School Officers. (June 10, 2012), *Reaching Higher: The Common Core State Standards Validation Committee*. Washington, DC: NGA and CCSSO, p.3. <http://nonpartisaneducation.org/Review/Reviews/CommonCoreValidationCommitteeReport.pdf>
7. See Special Section: The Common Core Dissenters. (2013). *Nonpartisan Education Review*, <http://nonpartisaneducation.org/CommonCoreDissenters.htm>
8. See Sandra Stotsky. (2014). Letter to sponsors, *Nonpartisan Education Review*, <http://nonpartisaneducation.org/Review/Reviews/v9n4.doc>
9. R.P. Phelps. (Fall 2006). Characteristics of an effective student testing system, *Educational Horizons*, 85(1), 19–29. <http://nonpartisaneducation.org/Foundation/EffectiveTestingSystem.pdf>
10. Similar abhorrence is projected toward ability grouping or curricular tracking within schools and retention in grade.
11. U.S. Department of Education. (April 9, 2010). Overview Information: Race to the Top Fund Assessment Program; Notice Inviting Applications for New Awards for Fiscal Year (FY) 2010, *Federal Register Notices*, 75(68), p.18172.

12. Clifford Adelman. (1999). *Answers in the Toolbox*. Washington, DC: U.S. Department of Education, <http://www2.ed.gov/pubs/Toolbox/toolbox.html>; Clifford Adelman. (2006). *The Toolbox Revisited: Paths to Degree Completion from High School Through College*. Washington, DC: U.S. Department of Education, <http://www2.ed.gov/rschstat/research/pubs/toolboxrevisit/toolbox.pdf>. A NAGB panel studied this issue and concluded that more research was needed on the issue. See <http://www.nagb.org/content/nagb/assets/documents/publications/PreparednessFinalReport.pdf>. Also see the panel chair's comment here: <http://collegepuzzle.stanford.edu/?p=466/>
13. As currently proposed, those who achieve a PARCC test score that statistically predicts a 75 percent probability of achieving at least a "C" in beginning college algebra or introductory statistics at a two-year college will be designated "college ready." See U.S. Department of Education. (April 9, 2010). Overview Information: Race to the Top Fund Assessment Program; Notice Inviting Applications for New Awards for Fiscal Year (FY) 2010, *Federal Register Notices*, 75(68), p. 18172.
14. Supporters of the one-size-fits-all U.S. system often label European and East Asian education systems as "elitist" and our system as a more "democratic," "second chance" system. That contrast may have been valid 60 years ago but is no longer. It is now easier to enter upper-academic levels in current European systems, and most countries now offer bridge programs for, say, a dissatisfied vocational-track graduate to enter a university or an advanced technical program. Typically, bridge programs are free of charge. For example, Germany offers the *abendgymnasium* and France the DAEU. (See "DAEU: le diplôme de la seconde chance," *Le Monde de l'Education*, 1996, 241, 81–84.)
15. Our public education system is neither less elitist nor more conducive to "second chances." In typical European or East Asian systems, multiple programs and tracks offer multiple opportunities for students to attain high achievement in something. A German student who enters a vocational-technical program at the lower-secondary level and finishes by passing the industry-guild certification examination as a machinist enters an elite of the world's most skilled and highly remunerated craftsmen.
16. See, for example, Zeng, Kangmin. (2001). *Dragon Gate: Competitive Examinations and Their Consequences*. London: Cassell; Britton, E. D., Hawkins, S., & Gandal, M. (1996). Comparing examinations systems. In E.D. Britton & S.A. Raizen (Eds.), *Examining the examinations* (pp. 201–218). Boston, MA: Kluwer Academic; Eckstein, M. A., & Noah, H. J. (1993). *Secondary school examinations: International perspectives on policies and practice*. New Haven, CT: Yale University Press; Stevenson, Harold W. and Shin-ying Lee, et al. 1997. International comparisons of entrance and exit examinations: Japan, United Kingdom, France, and Germany. U.S. Department of Education, Office of Educational Research and Improvement; Phelps, R. P. (2001, August). Benchmarking to the world's best in Mathematics. *Evaluation Review*, 25(4), 391–439; and Mullis, Ina V.S. 1997a. "Benchmarking toward World-class Standards: Some characteristics of the highest performing school systems in the TIMSS" Paper presented at panel "International Benchmarking: New Findings," at the Annual Meeting of the American Educational Research Association, Chicago, IL.
17. Richard P. Phelps and Mónica Silva. (2014). The effect of testing on student achievement, 1910–2010: Meta-regression analysis. Paper presented at the 9th Conference of the International Test Commission, San Sebastian, Spain, July 5, 2014.
18. See, for example, Howard Wainer. (2011). Chapter 9: Assessing Teachers from Student Scores: On the Practicality of Value-Added Models, in *Uneducated Guesses: Using Evidence to Uncover Misguided Education Policies*. Princeton, NJ: Princeton University Press, p. 120+.
19. See, for example, Eklof, H. (2007). Test-taking motivation and mathematics performance in TIMSS 2003. *International Journal of Testing*, 7(3): 311–326; ... (2010). Skill and will: Test-taking motivation and assessment quality. *Assessment in education: Principles, Policy & Practice*, 17(4), 345–356; Phelps, R. P. & Moskowitz, J. (1995). *Report on the U.S. pilot administration of the OECD's Cross Curricular Competency (CCC) Test*, for the Organisation for Economic Co-operation and Development and the National Center for Education Statistics, Office of Educational Research and Improvement, U.S. Department of Education, Washington, DC: Pelavin Associates; Wise, S.L. & DeMars, C.E. (2003, June). Examinee motivation in low-stakes assessment: Problems and potential solutions. Paper presented at the annual meeting of the American Association of Higher Education Assessment Conference, Seattle http://www.jmu.edu/assessment/wm_library/Examinee_Motivation.pdf

20. Harold W. Stevenson, Shin-Ying Lee, and James W. Stigler (1986). "Mathematics Achievement of Chinese, Japanese, and American Children." *Science* 231 (4739): 693–699; Harold W. Stevenson and James W. Stigler (1992). *The Learning Gap: Why Our Schools are Failing and What We Can Learn From Japanese and Chinese Education*. New York: Summit Books, pp. 13–27 in particular.
21. Mass Insight Education, 2002. *Taking Charge: Urban high school students speak out about MCAS, academics and extra-help programs*. Mass Insight.
22. For example, the French baccalauréat—the secondary level exit exam for their academic track—is subdivided, first by *séries* (literature, economics and social science, science), then by *spécialités* (options) and then by *travaux personnel encadrés* (personal project work). A French student might take just one from among dozens of possible exit exams and then a completely different higher education entrance exam from among a variety of such exams.
23. See, for example, Catherine Gewertz. (April 27, 2015). Florida's Biggest District Cuts Nearly All End-of-Course Exams. *Education Week*. http://blogs.edweek.org/edweek/curriculum/2015/04/miami_dade_cut_nearly_all_end_of_course_exams.html. (2015, July 15). MCPS eliminates middle school final exams. Montgomery County, Maryland: Author. <http://www.mymcmedia.org/mcps-eliminates-middle-school-final-exams/>; Drellinger, D. (2013, July 2). Louisiana high school end-of-course test results rise slightly. *New Orleans Times-Picayune*. http://www.nola.com/education/index.ssf/2013/07/louisiana_end-of-course_test_r.html.
24. R. P. Phelps and R.J. Milgram. (2014). *The Revenge of K-12: How Common Core and the New SAT Lower College Standards in the U.S.* Pioneer Institute White Paper 122, pp. 24-25. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2519889
25. A minimum competency test is a high-stakes test that requires performance at or above a single threshold test score before a certain attainment (e.g. a high school diploma) will be recognized.
26. Economist John Bishop found that New York State high school graduates achieved superior academic and labor market outcomes than graduates from other states, other factors held equal. See Bishop, J.H. (Summer 2000). Diplomas for learning: Not for seat time. *Economics of Education Review*, 19, 333–349.
27. See, for example, Cross, C.T. & Roza, M. (2007, March 14). *How the federal government shapes and distorts the financing of K-12 schools*. School Finance Redesign Project, Center on Reinventing Public Education, University of Washington, Daniel J. Evans School of Public Affairs. <http://digitalarchives.wa.gov/wa.media/do/676303ce60013b67ca781ff4049161d9.pdf>; Vinovskis, M. (2009). *From A Nation at Risk to No Child Left Behind: National Education Goals and the Creation of Federal Education Policy*. New York: Teachers College Press.
28. Phelps and Milgram, 2014, p. 24.
29. See, for example, Eleanor Chelimsky, *Student Achievement Standards and Testing*, T-PEMD-93-1. Washington, DC: US GAO.
30. See, for example, Ferrara, S., Willhoft, J., Seburn, C., Slaughter, F., & Stevenson, J. (1991). Local assessments designed to parallel statewide minimum competency tests: Benefits and drawbacks. In R.G. O'Sullivan & R.E. Stake (Eds.), *Advances in program evaluation: Effects of mandated assessment on teaching* (pp. 41-74). Greenwich, CT: JAI Press; Goldberg, Gail and B.S. Roswell 1999/2000. "From perception to practice: the impact of teachers' scoring experience on performance-based instruction and classroom assessment" *Educational Assessment*, 6(4), 257-290.
31. A Pearson subsidiary, Pearson Evaluation Systems, is located in Hadley, Massachusetts, but its sole product is teacher licensure testing, and not K-12 student testing.
32. Pearson Regional Scoring. <https://sites.google.com/a/pearson.com/regional-scoring/project-dates>
33. See, for example, Valerie Strauss. (2015, January 16). "Pearson criticized for finding test essay scorers on Craigslist," *Washington Post*. <http://www.washingtonpost.com/blogs/answer-sheet/wp/2013/01/16/pearson-criticized-for-finding-test-essay-scorers-on-craigslist/>; Claudio Sanchez. (2015, July 8). How Standardized Tests Are Scored (Hint: Humans Are Involved) [audio], National Public Radio. <http://www.npr.org/sections/ed/2015/07/08/420559529/unveiling-how->

- [standardized-tests-are-scored](#); Motoko Rich. (2015, June 22). "Grading the Common Core: No teaching experience required," *New York Times*. http://mobile.nytimes.com/2015/06/23/us/grading-the-common-core-no-teaching-experience-required.html?referrer=&_r=1
34. Richard P. Phelps and R. James Milgram. (2014). *The Revenge of K-12: How Common Core and the new SAT lower college standards in the U.S.* Boston, MA: Pioneer Institute White Paper No. 122, pp. 15-18.
 35. See, for example, Howard Wainer. (2011). Chapter 2: On Substituting Achievement Tests for Aptitude Tests in College Admissions, in *Uneducated Guesses: Using Evidence to Uncover Misguided Education Policies*. Princeton, NJ: Princeton University Press, p. 20.
 36. Despite the fact that the National Assessment of Educational Progress already compares educational achievement across states. And, despite the fact that negotiations to keep states from dropping out of the assessment consortia have now produced differences in testing procedures across states that threaten comparability. See, for example, Christine Amario. (2015, August 29). As Common Core results trickle in, initial goals unfulfilled. *Associated Press & Yahoo! News*. <http://news.yahoo.com/common-core-results-trickle-initial-goals-unfulfilled-132810793.html>
 37. See John Fensterwald (August 26, 2015). "State removes 15 years of test results before releasing new scores." *EdSource*. <http://edsources.org/2015/state-removes-15-years-of-test-results-before-releasing-new-smarter-balanced-scores-aligned-to-common-core/85561>.
 38. Whereas reliability can be established fairly quickly after a few test administrations, establishing validity depends on a test's claimed purpose. PARCC officials claim it will be predictive of college and career success. That claim of predictive validity will require a large number of data records that connect individual student PARCC scores with the same students' actual college and career outcome data, such as, for college, grade-point average and enrollment, persistence, and completion indicators, and, for career, salary, employment rate, and occupation type indicators.
 39. <http://www.mathedconsulting.com>
 40. Comments used with permission.
 41. See, for example, <http://www.patriotledger.com/article/20150716/NEWS/150717598>
 42. <http://www.doe.mass.edu/parcc/MCASChoicePres.pptx>; http://renniecenter.org/research/PARCC_TestingTest.pdf
 43. J. W. Popham. (2004). *Classroom assessment: What teachers need to know*. Chicago, Ill.: Pearson
 44. D, Frisbie. (1992). The multiple true-false test item format: A status review. *Educational Measurement Issues and Practices*. 11. Winter 21-26.
 45. For example, see S. Gultekin and N.C. Mimirtash. (2012). Comparing the test information obtained through multiple-choice, open-ended, and mixed item tests based on item response theory. *Elementary Education Online*: 11(1) 251.
 46. In May 2015, the PARCC Governing Board voted, among other things, to reduce testing time and compact PARCC's two assessments (PBA and EOY) into one. See "PARCC Testing Time Gets Trimmed." <http://www.edweek.org/ew/articles/2015/05/21/parcc-shortens-its-common-core-test.html>
 47. Commissioner's email to district superintendents in July 2015 re: 20015-2016 Statewide Testing Schedule and Administrative Deadlines. (<http://www.doe.mass.edu/news/news.aspx?id=15687>)
 48. The 2015 MBAE report (<http://www.mbae.org/wp-content/uploads/2015/02/MBAE-MCAS-PARCC-Report-Web.pdf>) suggests that PARCC's added testing time for writing is a highly positive new direction for Massachusetts, but it does not explain what new information it provides.

49. <http://edexcellence.net/next-generation-science-and-college-readiness>; http://edex.s3-us-west-2.amazonaws.com/publication/pdfs/20130612-NGSS-Final-Review_7.pdf; see comment by a mathematician at <http://www.edexcellencemedia.net/publications/2013/20130820-NGSS-Appendix-Review-and-State-Comparisons/20130820-Commentary-on-Appendix-L-FINAL.pdf>; see comments by non-scientists: <http://edexcellence.net/commentary/education-gadfly-weekly/2013/august-22/next-generation-science-standards-revisited.html>
50. http://www.doe.mass.edu/mcas/tech/technical_quality.doc
51. See C. Riley. (2013). PARCC: An update and overview of standard setting research at http://highereducationcolorado.gov/academics/Groups/RemedialReview/Meetings/Handouts/oct25_PARCCPP.pdf
52. http://www.nj.com/education/2015/08/parcc_scores_setting_the_standard.html
53. http://www.uxbridgeschools.com/pdf_info/uhs_guid_stanleykoplik_diff_between_koplik_abigail.pdf
54. <http://www.edweek.org/ew/articles/2015/04/22/more-universities-take-up-sbac-college-ready-cutoffs.html>
55. <https://exceptionaldelaware.wordpress.com/2015/04/14/breaking-news-delaware-colleges-universities-to-use-smarter-balanced-scores-for-acceptance-credentials/>
56. <http://www.mass.edu/bhe/lib/documents/AAC/AAC14-12DevelopmentalMathEducationTaskForceRecommendations-supersededbyOct22ndAACmeetingedit.pdf>
57. <https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbXniaXNrbWF0aHxneDoyOGRiNmQzMjViMGE3Y2Zk>. The full report is on pp. 2-32.
58. http://media.collegeboard.com/digitalServices/pdf/research/2013/MA_13_03_03_01.pdf
59. Mike Winders and Richard Bisk. September 30, 2014. "Math Task Force's Bad Calculation." *New England Journal of Higher Education*. See the response in NEJHE in October 2014 by Carlos Santiago, the chairman of the MBHE, and the comment on his response by Professor Bisk. <http://www.nebhe.org/newslink/remedying-nebhes-journal-piece-on-developmental-math-a-letter-to-the-editor/>
60. <http://commonwealthmagazine.org/education/lowering-math-standards-not-the-answer/>
61. http://www.nytimes.com/2011/04/26/education/26inflate.html?_r=0
62. R.P.Phelps and R.J. Milgram. (September 2014). *The Revenge of K-12: How Common Core and the New SAT Lower College Standards in the U.S.* Pioneer Institute White Paper #122. <http://pioneerinstitute.org/featured/common-core-math-will-reduce-enrollment-in-high-level-high-school-courses/>
63. <https://www.google.com/#q=2015+renaissance+learning+report+reading>
64. http://www.nas.org/images/documents/Beach_Books_10-21-14.pdf
65. According to *Beach Books: 2013-2014*, the top seven books assigned as summer reading by 341 colleges are as follows (together with a reading level [RL], if available, based on Renaissance Learning's readability formula) <http://www.arbookfind.com/UserType.aspx>
The Immortal Life of Henrietta Lacks by Rebecca Skloot (RL: 8.1)
This I Believe by Jay Allison and Dan Gediman
The Other Wes Moore by Wes Moore (RL: 7.1)
Wine to Water by Doc Hendley

Little Princes: One Man's Promise to Bring Home the Lost Children of Nepal by Conor Grennan (RL: 6.1)

Full Body Burden: Growing Up in the Nuclear Shadow of Rocky Flats by Kristen Iversen (RL: 7.0)

Half the Sky: Turning Oppression into Opportunity for Women Worldwide by Nicholas Kristof and Sheryl WuDunn (RL: 9.5)

The average RL for the five of the top seven books whose RL is available is 7.56 (meaning grade 7, sixth month). When we go deeper into the reading list, the reading level seems to get lower. Of the 53 most frequently mentioned titles listed in Beach Books: 2013-2014, the reading levels of 23 were available, with an average level of 6.8.

As for Renaissance Learning's own reports, its 2014 report showed that the average reading level (using its own readability formula—ATOS for Books) was 6.7 for the 25 most frequently read works of fiction by grade 12 students. This number was higher than the average reading level for the top 25 informational texts read by grade 12 students. The average reading levels at other high school grades were lower for both the top 25 works of fiction and informational texts, calculated separately. Its 2015 report shows a reading level of less than grade 6 for grade 11 students, judging by the reading level of the books they report reading. This is hardly the reading level needed for college textbooks and other readings assigned in college.

66. <http://www.doe.mass.edu/mcas/overview.html?faq=2>.
67. Massachusetts Board of Higher Education and Massachusetts Department of Education. (February 2008). *Massachusetts School-to-College Report: High School Class of 2005*. See p. 7.
68. See, for example, <http://eric.ed.gov/?id=ED492062>; <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4031673/>; and <http://www2.ed.gov/programs/readingfirst/support/rmcfinal1.pdf>
69. 1998: <https://archive.org/details/massachusettscompr00mass>;
1999: <https://archive.org/details/masscomprehensiv00mass>;
<https://www.brocktonpublicschools.com/uploaded/TeachingLearning/MathResourcesK-8/MCAs-Questions/MCAS-2000.pdf>;
https://www.brocktonpublicschools.com/uploaded/TeachingLearning/MathResourcesK-8/MCAs-Questions/MCAS_2001.pdf;
<https://www.brocktonpublicschools.com/uploaded/TeachingLearning/MathResourcesK-8/MCAs-Questions/MCAS-2002.pdf>;
<https://www.brocktonpublicschools.com/uploaded/TeachingLearning/MathResourcesK-8/MCAs-Questions/MCAS-2003.pdf>;
<https://www.brocktonpublicschools.com/uploaded/TeachingLearning/MathResourcesK-8/MCAs-Questions/MCAS-2004.pdf>;
70. <http://parcc.pearson.com/practice-tests/english/>
71. https://login4.cloud1.tds.airast.org/student/V111/Pages/LoginShell.aspx?a=Student&c=SBAC_PT&logout=true
72. Louisiana Guide to the PARCC Assessments for Grades 6-8: English Language Arts/Literacy, posted February 10, 2015, p. 2.
73. Item Guidelines for ELA/Literacy PARCC Summative Assessment, 8/9/2014, pp. 3-4.
http://www.jcboe.org/boe2015/images/PARCCCCSS/Updated_Formatted_Item_Guidelines_.pdf
74. Item Guidelines for ELA/Literacy PARCC Summative Assessment. 8/9/2014, p. 19.
http://www.jcboe.org/boe2015/images/PARCCCCSS/Updated_Formatted_Item_Guidelines_.pdf
75. http://www.smarterbalanced.org/wordpress/wp-content/uploads/2011/12/ELA_Construct_Relevant_Vocabulary.pdf

76. <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/EnglishLanguageArtsLiteracy/ELAStimulusSpecifications.pdf>
77. Item Guidelines for ELA/Literacy PARCC Summative Assessment, 8/9/2014, pp. 3-4. http://www.jcboe.org/boe2015/images/PARCCCSS/Updated_Formatted_Item_Guidelines_.pdf
78. “Definitions will be provided for words that are important to understanding the text but do not have sufficient context.” Louisiana Guide to the PARCC Assessments for Grades 6-8: English Language Arts/ Literacy, posted February 10, 2015, p. 9.
79. <http://www.readingrockets.org/article/teaching-vocabulary>
80. <http://www.readingrockets.org/article/teaching-vocabulary>
81. 2007 MCAS Technical Report. Under Content-Related Test Specifications, we find: 2.1.4.1 English Language Arts Test Development...The grade 3 test included 50 percent literary and 50 percent informational passages. The grades 4–8 and 10 tests and the Retests included 60 percent literary and 40 percent informational passages. Approximately 50 percent of the authors of test passages are listed in the Framework’s Appendices A and B. p. 9. <http://iservices.measuredprogress.org/documents/MA/Technical%20Report/2007/04-09-08%202007%20MCAS%20Tech%20Rpt%20Final%20PDF.pdf>
82. M. Scardamalia, C. Bereiter, and R. Steinbach. (1984). Teachability of reflective processes in written communication. *Cognitive Science*. 8(2), 173-190. <http://csjarchive.cogsci.rpi.edu/1984v08/i02/p0173p0190/MAIN.PDF>; See also A. Becker. (2006). A review in writing model research based on cognitive processes in Revision: *History, theory, and practice*. A. Horning and A. Becker (Eds.). Parlor Press: West Lafayette, Indiana. Ch.3. http://wac.colostate.edu/books/horning_revision/chapter3.pdf;
83. Nowhere in this paper do we discuss the issue of touch-typing as a genuine problem for developing writers, nor do we examine what different skills—handwriting, spelling, punctuation, recopying first drafts etc.—are tapped when writing with pen and paper. These are substantial issues beyond the scope of what can be discussed here. Some writing teachers like Ann Trubeck look upon touch-typing as an essential 21st century skill that is not taught in schools, largely based on the conventional wisdom that today’s digital natives (i.e., kids now in school) already know how to use a keyboard, which is actually not true. The lack of touch-typing skill, she argues, significantly impedes “cognitive automaticity” an essential ability needed to compose with a computer. See A. Trubek’s August 15, 2011 post “Out of Touch with Typing” in *MIT Technology Review*. The consensus among researchers is that touch-typing should be taught sometime between 3rd and 5th grade (see A. Bullock’s review of the literature at http://chiron.valdosta.edu/are/litreviews/vol4no2/AllenBullock_LitRev.pdf). If this is true, PARCC’s decision for an online test may pose genuine problems for elementary test-takers if (a) they have not been taught to touch-type and can only “keyboard” or (b) they do not have access to a computer to complete a written exam and are therefore forced to lose time and focus by completing tasks with pen-and-paper. Both scenarios place writers at a disadvantage. PARCC has no formal policy on touch-typing.
84. See G. Conlon in Appendix A in Gray et al. (1982). 211-213). For a partial review of earlier research on writing prompts see Godshalk, F.I., Swineford, F. & Coffman, W.E. (1966) The measurement of writing ability. New York: CEEB; Gray, J., Ruth, L., Murphy, S., Keech, C., Carroll, K., Kinzer, C., Leu, Robyns, A., & Warantz, E. (1982) *Properties of writing tasks: A study of alternative procedures for holistic writing assessment*. Berkeley, CA: Bay Area Writing Project, University of California, Graduate School of Education. ERIC (No. ED 230 576) <http://www.eric.ed.gov/PDFS/ED230576.pdf>; and C. Keech. (1982). “Practices in defining writing test prompts: Analysis and recommendations.” In Gray et al. (1982). 132-198.
85. M. Smith and S. Swain. (2011). *Wise eyes: Prompting for meaningful student writing*. Berkeley, CA: National Writing Project. Fall. 2-3. http://www.nwp.org/cs/public/download/nwp_file/15440/Wise_Eyes.pdf?x-r=pcfile_d
86. These priorities are routinely emphasized in much of the literature on teaching writing. See, for example, C. R.

- Cooper and L. Odell. (1977). *Evaluating writing*. Urbana, Ill: NCTE. See also P. B. Diederich, (1974). *Measuring growth in writing*. Urbana, Ill: NCTE.
87. J. L. Kinneavy. (1971). *A theory of discourse*. Englewood Cliffs: Prentice Hall.
 88. J. Britton, N. Martin, T. Burgess, A. McLeod, and R. Rosen. (1975). *The development of writing abilities 11-18*. London: Macmillan Education Ltd.
 89. James Moffett. (1968). *Teaching the universe of discourse*. Boston: Houghton Mifflin.
 90. If high school teachers teach writing directly at all, their practice centers mainly on teaching students how to compose, revise, and edit expressive or narrative writing, with an occasional exploration of literary criticism and interpretation. This is not much different from what children learn in upper elementary and middle school. For a review of literature on writing instruction, see G. A. Troia and S. Graham. (2003). Effective writing instruction across the grades: What every consultant should know. *Journal of Educational and Psychological Consultation* 14 (1) 75-89.
 91. T. Newkirk. (2013). Speaking back to the Common Core.
https://www.heinemann.com/shared/onlineresources/e02123/newkirk_speaking_back_to_the_common_core.pdf
 92. J. S. Chall. (1979). The great debate: Ten years later with modest proposal for reading stages. In L. Resnick and P. Weaver (Eds.) *Theory and practice in early reading*. Vol. 1. Hillsdale, NJ: Erlbaum Associates.
 93. As used by PARCC, research “simulations” are supported by no evidence that these test items validly capture what researchers do, either in the humanities or in the sciences. Educational Testing Service, Pearson’s co-contractor, cites no studies on research simulation in the research studies cited on its website.
 94. For an extensive discussion of how exposition and argument differ, see W. Brandt. (1970). *The rhetoric of argumentation*. New York: Bobbs-Merrill.
 95. C. R. Cooper, L. Odell, and K. Coats (1978). Discourse theory: Implications for research in composing. In C. R. Cooper and L. Odell (Eds.). *Research on composing: Points of departure*. Urbana. NCTE. 1-12.
 96. For a review of Clay’s work and subsequent research, see D. Johnson. (1999). *Critical issues: Addressing the literacy needs of emergent and early readers*. North Central Regional Education Lab: Oak Rock, Ill. ERIC ID. ED480227.
 97. J. S. Chall. (1979). The great debate: Ten years later with modest proposal for reading stages. In L. Resnick and P. Weaver (Eds.) *Theory and practice in early reading*. Vol. 1. Hillsdale, N.J.: Erlbaum Associates.
 98. W. Loban. (1976) *Language development: Kindergarten through grade 12*. Urbana, Ill. NCTE.
 99. S. Stotsky, (1982). A model of written language development for teachers. *Fforum* (A Newsletter of the English Composition Board). University of Michigan: V.4. 89-96.
 100. D. H. Graves. (1975). An examination of the writing processes of seven year-old children. *Research in the Teaching of English*. 9. 227-242
 101. D. Prater and W. Padia. (1983). Effects of modes of discourse in grades 4 and 6. *Research in the Teaching of English*, 17 (2).
 102. M.K. McQuillan. (1984). *The measurement of writing abilities*. Unpublished dissertation: Harvard Graduate School of Education. Cambridge, MA.
 103. Massachusetts Department of Education. (September 28, 1999). Massachusetts School and District Accountability System, Malden, MA. <http://www.doe.mass.edu/apa/accountability/cyclei/sdas.pdf>. “At the core of the academic changes brought about by Education Reform are the Curriculum Frameworks for the core subjects specified in

the Act (English Language Arts, Mathematics, Science and Technology, History and Social Science, Foreign Languages, Health, and the Arts). The frameworks establish the standards for what students should know and be able to do at particular stages of their education. The Massachusetts Comprehensive Assessment System (MCAS), a criterion referenced test aligned with those standards, is designed to test whether students have learned the content and skills set out in the frameworks” (p. 1).

104. <http://pioneerinstitute.org/education/how-massachusetts-promoted-achievement-before-common-core-parcc/>
105. See a review of the 2015 book by Sandra Stotsky detailing these changes at http://www.goacta.org/the_forum/from_the_bookshelf_an_empty_curriculum_the_need_to_reform_teacher_licensing
106. See Superintendent William Lupini’s letter to the Brookline School Committee in June 2014 in Appendix B of <http://pioneerinstitute.org/education/why-massachusetts-should-abandon-the-parcc-tests-common-core/>
107. <http://nces.ed.gov/nationsreportcard/subject/publications/stt2013/pdf/2014082MA12.pdf>; see also <http://www.doe.mass.edu/mcas/2013/results/summary.pdf>

APPENDIX A. CRITIQUE OF CRITERIA FOR EVALUATING COMMON CORE-ALIGNED ASSESSMENTS

According to the non-disclosure agreement that the Massachusetts Department of Elementary and Secondary Education (DESE) signed with the Thomas B. Fordham Institute in early 2015, the forthcoming Fordham et al. report will use the criteria conveniently issued by the Council of Chief State School Officers (CCSSO) in March 2014 to evaluate test content and design in MCAS and PARCC in grades 5 and 8, while its partner will examine high school test content and design.¹ The project is funded by the High-Quality Assessment Project, described by DESE as a coalition of national foundations (among them the Bill and Melinda Gates Foundation) and just recently founded, it seems.² So, where did these criteria come from, what are they, and how useful can they be?

According to the CCSSO, its *Criteria for Procuring and Evaluating High-Quality Assessments* built on a commitment the states made to high-quality assessments aligned to college and career readiness and it will simply “assist states in operationalizing their commitment ...” However, this commitment was actually made by CCSSO on behalf of the states, not by the states themselves. In an October 2013 letter from Chris Minnich, Executive Director of the CCSSO, titled *States’ Commitment to High-Quality Assessments Aligned to College and Career Readiness*, he asserts that:

“CCSSO, on behalf of the states, hereby commits to further states’ proactive leadership in promoting college and career readiness for all students by establishing or adopting high quality systems of assessments, including summative, interim, and classroom assessments, based on college- and career-ready (CCR) standards...” and that these assessment systems will: “assess higher-order cognitive skills; assess critical abilities with high-fidelity; be based on CCR standards that are internationally benchmarked; be instructionally sensitive and educationally valuable; and be valid, reliable, and fair.”³

Moreover, a footnote in this five-page letter indicates that these criteria are taken from a June 2013 report titled *Criteria for High-Quality Assessment* written by Linda Darling-Hammond and many others at the Stanford Center for Opportunity Policy in Education.⁴ In other words, CCSSO is urging states to evaluate their Common Core-based assessments with criteria written by the founder of one of the testing companies funded by the USED to develop Common Core-based tests (the Smarter-Balanced Assessment Consortium or SBAC).

But CCSSO’s criteria are more problematic than that. In 2014 CCSSO claimed that its *Criteria for Procuring and Evaluating High-Quality Assessments* were based chiefly on “*Standards for Educational and Psychological Testing* (AERA, APA, and NCME, 1999).”⁵ This volume contains the standards or criteria that professionals use to evaluate new tests. The problem is that Common Core-based assessments violate these professional standards.

For example, the first assessment standard in *Standards*—Standard 1.0—is the standard many professionals consider the testing field’s “prime directive.” It reads as follows:

“Clear articulation of each intended test score interpretation for a specified use should be set forth, and appropriate validity evidence in support of each intended interpretation should be provided.”⁶

In short, a test should be validated for each purpose for which it is used. But PARCC cannot be validated for its purpose of predicting college and career readiness until data are collected in years to come on the college and career outcomes of PARCC test-takers in 2015. And it is possible it may never be validated. So, PARCC, in effect, has violated Standard 1.0.

There is an even deeper problem in using *Criteria for Procuring and Evaluating High-Quality Assessments*. As CCSSO makes very clear, its criteria apply specifically (and only) to assessments of “college and career” standards. Clearly, Massachusetts faces a dilemma. The

criteria designed to evaluate PARCC can't be used to evaluate MCAS. Yet, that is what DESE has agreed to, without a public hearing on the matter.

MCAS was conceptualized in the law as an assessment to determine eligibility for a high school diploma, and the standards on which it was based before 2011 were conceptualized to designate the “knowledge and skills” students needed for a high school diploma, not a freshman year at Bunker Hill Community College, Salem State University, Northeastern University, or MIT. On the other hand, DESE wants to be sure that PARCC both reflects Common Core's standards and is of high enough quality so that Bay State students accepted at these colleges don't need remedial or developmental coursework—an impossible twin goal. (Since the use of Common Core's standards may not lead to a high-enough score on a college placement test that results in an exemption from remedial coursework, it is understandable why there is now pressure on post-secondary institutions to accept a determination of college readiness by a Common Core-based readiness test in grade 11 and forgo use of a placement test in college.)

The bias in CCSSO's criteria probably accounts for the particular adjectives and adverbs used over and over in the 17 page-document: “high-quality” 24 times; “higher” 18 times; “higher-level” four times; “deep”, “deeply”, or “deeper” 14 times; “critical” or “critically” 17 times; and “valuable” nine times. These words have been used for years by the Fordham Institute and other advocates of Common Core's standards to describe them.

These oft-repeated words also suggest that much of the thinking behind these criteria can be traced to education psychologist Benjamin Bloom's 1956 *Taxonomy of Educational Objectives*.⁷ In an attempt to impose more definition on the learning process, Bloom and his colleagues identified six different thought processes involved in learning: 1) knowing; 2) comprehending; 3) applying; 4) analyzing; 5) synthesizing; and 6) evaluating. Interestingly, Bloom and colleagues listed

knowledge first and evaluation last.⁸ They argued that their taxonomy represented a hierarchy, but not because any of the thought processes were superior to any of the others. Rather, the order of the taxonomy represented the natural flow of reasoning: one must know and recall information before one can understand it; one must understand information before one can apply it; and one must apply information before one can analyze, synthesize, or evaluate it (although no empirical research has ever produced evidence for this particular order). Indeed, one could argue that knowledge and its recall could be considered the most important thought process for, without it to start with, none of the other thought processes is possible.

The CCSSO document does not list Bloom's Taxonomy directly in its bibliography but, rather, Darling-Hammond et al's 2013 document, another one issued in 2013 by the Stanford Center for Opportunity Policy in Education, as well as a document issued in 2013 by the Center for Research on Education Standards and Student Testing (CRESST), written by a co-author of one of the Stanford Center monographs.⁹ To understand the origins of “deeper learning,” a phrase commonly used to describe the effects of Common Core's standards and tests by its advocates, we must look at this document's hierarchy of “depth of knowledge” (DOK), summarized on p. 5 as follows:

- DOK1: Recall of a fact, term, concept, or procedure; basic comprehension.
- DOK2: Application of concepts and/or procedures involving some mental processing.
- DOK3: Applications requiring abstract thinking, reasoning, and/or more complex inferences.
- DOK4: Extended analysis or investigation that requires synthesis and analysis across multiple contexts and non-routine applications.

The language of DOK4 begins to have a familiar ring. But no empirical evidence is provided for this hierarchy.

Is there evidence that PARCC’s new test-item types tap “deeper” learning?

Among its dozens of criteria, the CCSSO document suggests that evaluators find out if there are “rationales for the use of specific item types,” such as “selected-response, two-part evidence-based selected-response, short and extended constructed-response, technology-enhanced, and performance tasks.” So, does PARCC use item-types that lead to deeper learning? Its tests include many two-part evidence-based selected-response test items and multi-step problems.

TEST ITEMS IN TWO PARTS OR WITH MULTI-STEP PROBLEMS

As one kind of selected-response item—the general category of test items of which multiple-choice is the most familiar member—multi-step problems are problems inside one test item rather than spread across a series of test items. No evidence has been cited by PARCC that such test items lead to deeper learning. Nevertheless, CCSSO’s criteria suggest that these test-item types are desirable on tests. We ask: even if the ultimate effect may be to degrade item fairness? How so?

When multi-step problems are presented in separate items, students are afforded the maximum opportunity to display their knowledge. Say, a process runs from step A through steps B and C to step D. If the process is presented in a test as four separate test items, students can obtain credit for knowing correct answers to later steps even if they are wrong about step A. When a multi-step problem is inside a single test item, credit is all-or-nothing. Students get the test item right only if they get all the steps right. When two multiple-choice items are paired in a two-part Evidence-Based Selected Response (EBSR), students must get Part A right (as well as Part B) in order to get credit for the item. Not only is this unfair, but educators also gain less useful information about a student’s strengths and weaknesses. Nevertheless, CCSSO’s criteria suggest that the test items are desirable.

DO WE NEED DIFFERENT TEST ITEM FORMATS TO ASSESS “HIGHER-ORDER” THINKING?

Intriguingly, the bulk of Bloom’s Taxonomy of Educational Objectives includes many examples of multiple-choice test items designed to measure all six of the Taxonomy’s listed thought processes. Thus, the foundational work on the subject starkly contradicts the pervasive argument that different kinds of assessment formats are needed to assess “higher-order” thinking.

However, multiple-choice items have always been capable of measuring intellectually ambitious expectations (as exemplified in the MCAS ELA grade 10 questions on theme, developed by the state’s English teachers). It’s just that they did it with multiple items, which is a fairer and better way to do it, instead of within single, often convoluted multi-step test items.¹⁰ Many accusations leveled at multiple-choice items have little substance—to the effect that multiple-choice items demand only factual recall and “lower-order” thinking, while “performance-based” tasks do neither. It is the structure of the question that determines the character of the cognitive processing necessary to reach a correct answer.¹¹ There is no necessary correlation between the difficulty of a problem and its response format. Even integrative tasks that may require fifty minutes to classify, assemble, organize, calculate, and analyze can, in the end, present the test-taker with a multiple-choice response format. Just because the answer to the question is among those provided, it is not necessarily easy or obvious how to get from the question to the right answer.¹²

Straightforward and familiar testing formats are fairest, for they are most likely to measure a student’s mastery of the subject matter. The more complexity one adds to test and test item formats, the more likely the test will measure not mastery of subject matter, but knowledge, skills, and familiarity related to the formats instead (or, in psychometric-speak, they will produce high levels of “construct irrelevant variance”).¹³

WHAT ABOUT TECHNOLOGY-ENHANCED TEST ITEMS?

PARCC's chief "innovation"—the incorporation of multiple steps within individual test items—makes them difficult to navigate online. A co-author of this White Paper, who works in quantitative statistical research and sometimes employs "equation editors" when writing up his study results for scholarly journals, attempted to take the PARCC online practice test in grade 3 mathematics. He couldn't get the PARCC equation editor to work because the text boxes were too small. (Nor were there any instructions directing one to enter responses inside the tiny little box instead of anywhere else in the much larger blank space.) He did his best with work-arounds. But, even though the several test items that require use of equation editor must be hand-scored, he received no feedback as to whether his work-arounds worked or not. He received feedback only for his responses to the more traditional multiple-choice items.

He did encounter a much wider variety of test item formats in the PARCC practice tests for the higher grades. Indeed, a bewildering variety of test item formats. Building all these "innovative" products may help explain the enthusiasm of many test developers for PARCC and SBAC. But, what do unfamiliar formats measure? Knowledge of the subject matter, or an ability to decode unfamiliar symbols and structures? PARCC's wide variety of "innovative" (and untested) formats may inject "test prep" with steroids.¹⁴

PARCC tests are not quite as revolutionary in their delivery technology, though, as their proponents have claimed. In many venues in industry, government, and education, electronic tablets now deliver tests and surveys. Tablets share most of the better features of paper-and-pencil or desktop-computer delivery, while avoiding many of their drawbacks. We cannot make the point any better than did the bloggers at *CCSSI Mathematics*.¹⁵

"...none of the SBAC or PARCC extended tasks as of yet take advantage of technology's capabilities in such a way to justify the transition to computer-based assessments. ...Rather than demonstrating more authentic and complex tasks, they present convoluted scenarios and even more convoluted input methods. Rather than present multimedia in a way that is authentic to the tasks, we see heavy language describing how to input what amounts to multiple choice or fill-in the blank answers. ...it is hardly a 'next generation' set of items that will allow us to attain more accurate measures of achievement.

"Computer-based mouse and keyboard-entry assessments face other obstacles. Shifting from handwritten answers to typed answers to facilitate computer scoring isn't a sufficient justification, especially when primary school students now have to struggle to learn and physically handle what was at one time a skill taught in junior high school: touch typing on a keyboard that was designed 140 years ago.¹⁶

"Computer-based assessments must have seemed cutting-edge to the old fogies that drafted Common Core, but to many youngsters growing up nowadays with smart phones and tablets, computers are relics of their parents' era.... Assessments, when they become tech-based worthy, should be neither device-dependent nor exclusory."

FINAL OBSERVATIONS

With nine very costly hours of testing time per grade level required,¹⁷ and no research evidence publicly cited by PARCC anywhere to support the extensive use of the new types of test items in Common Core-based assessments of millions of children, one may wonder why assessments were constructed to feature them and why CCSSO promotes them in its criteria. The money that school districts and states are raising to support PARCC and other Common Core-based tests might have been better used to raise teachers' salaries or re-construct teacher and administrator training programs. The answers to these puzzling questions can be found in Section A of the application for a USED grant for test development.

Applicants were told what criteria would be used to assess the quality of their proposal. Proposals

were to contain assessment designs that were “innovative, feasible, and consistent with” the applicant’s “theory of action” They were also urged to design a system that would measure standards “traditionally difficult” to quantify and make use of “types of items (e.g., performance tasks, selected responses, brief or extended constructed responses)” sufficiently “varied” and able to elicit “complex student demonstrations” or applications of knowledge, with “descriptions to include concrete examples of each item type proposed, the rationale for using these items types, and their distributions” (pp. 31-32).

Keep in mind that the USED was in charge not only of dispensing stimulus money for test development but also of the Technical Review of the progress its grantees were making in developing Common Core-based tests. Interestingly, no mathematicians or literary scholars were selected by the USED to be on the technical review teams even though Common Core-based tests address only mathematics and English language arts. Yet, the psychometric, non-content experts on the technical team evaluating PARCC’s progress in 2013 felt confident enough to applaud “PARCC’s use of authentic texts and thoughtful combinations of texts,” to highlight “that the consortium has effectively selected texts worth reading,” and to find “the constructed response items (assessment prompts requiring students to write essays) to be strong” (p. 2).

We then find the USED criteria for evaluating proposals to develop Common Core-based tests elaborated in a document produced by a copyright owner of Common Core’s standards (CCSSO). The document is to be used by, among others, state departments of education to evaluate the virtues of their Common Core-based assessments. As a result, we end up with a neat, circular process that is too obvious, if not unprofessional and unethical. For a range of independent data-driven comments on MCAS or PARCC as testing systems, see the studies in the endnotes.¹⁸

Appendix A Endnotes

1. <http://www.ccsso.org/Documents/2014/CCSSO%20Criteria%20for%20High%20Quality%20Assessments%2003242014.pdf>
2. Massachusetts Department of Elementary and Secondary Education. (2015). "Exhibit A: Description of Study: Evaluating the Quality of Common Core Assessments."
3. <http://www.ccsso.org/Documents/2013/CCSSO%20Assessment%20Quality%20Principles%2010-1-13%20FINAL.pdf>
4. Linda Darling-Hammond, Joan Herman, James Pellegrino, Jamal Abedi, J. Lawrence Aber, Eva Baker, Randy Bennett, Edmund Gordon, Edward Haertel, Kenji Hakuta, Andrew Ho, Robert Lee Linn, P. David Pearson, James Popham, Lauren Resnick, Alan H. Schoenfeld, Richard Shavelson, Lorrie A. Shepard, Lee Shulman, and Claude M. Steele. (2013). *Criteria for high-quality assessment*. Stanford, CA: Stanford Center for Opportunity Policy in Education; Center for Research on Student Standards and Testing, University of California at Los Angeles; and Learning Sciences Research Institute, University of Illinois at Chicago (<https://edpolicy.stanford.edu/publications/pubs/847>).
5. American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME). (2014). *Standards for Educational and Psychological Testing*, Washington, DC: AERA. CCSSO cited the 1999 version of the *Standards*; *Standards* was revised and re-issued in 2014.
6. AERA, APA, and NCME, *Standards for Educational and Psychological Testing*, p. 23.
7. Benjamin S. Bloom, Ed. (1956). *Taxonomy of educational objectives: Handbook 1: Cognitive domain*. New York: David McKay Company.
8. Bloom, p. 18.
9. L. Darling-Hammond and F. Adamson. (2013). **Developing assessments of deeper learning: The costs and benefits of using tests that help students learn**. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education (<https://edpolicy.stanford.edu/publications/pubs/745>); J. L. Herman and R.L.Linn. (2013). *On the road to assessing deeper learning: The status of Smarter Balanced and PARCC assessment consortia*. (CRESST Report 823). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
10. For the general superiority of the "fill-in-the-bubble" multiple-choice test item format, see, for example, Roediger, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 31, 1155–1159; Roediger, H. L. & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1, 181–210; Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249–255; Ross D. Hudson and David F. Treagust. (2013). Which form of assessment provides the best information about student performance in chemistry examinations? *Research in Science & Technological Education*, 31:1, 49–65, <http://eric.ed.gov/?id=EJ1001413>; Rudman, H. C. (Fall 1992). Testing for learning (book review). *Educational Measurement: Issues and Practice*, 31–32; Terry, P.W. (January 1934). How Students Study for Three Types of Objective Tests. *The Journal of Educational Research*, 27(5), pp. 333–343. <http://www.jstor.org/stable/27525804>; Traub, R. E. (1993). On the equivalence of the traits assessed by multiple choice and constructed-response tests. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement*. Hillsdale, NJ: Lawrence Erlbaum;
11. D. E. Powers and J. C. Kaufman. (2002). *Do standardized multiple-choice tests penalize deep-thinking or creative students?* (Research Report RR-02-15). Princeton, NJ: ETS.
12. Bridgeman, B. (1991). Essays and multiple-choice tests as predictors of college freshman GPA. *Research in Higher Education*, 32(2), 319–332; Croft, M. and Guffy, G. (2015). *Reviewing your options: The case for using multiple-choice test items*. Issue Brief, ACT Research & Policy. <http://act.org/research/policymakers/pdf/2015-Issue-Brief->

[Reviewing-Options.pdf](#); Feinberg, L. (Fall 1990). Multiple-choice and its critics: Are the ‘alternatives’ any better? *The College Board Review*, 157, 13–17, 30–31.

13. AERA, APA, NCME. *Standards for Educational and Psychological Testing*, pp. 54–57.
14. Richard P. Phelps. (Fall 2011). Teach to the test? *Wilson Quarterly* Washington, DC: Woodrow Wilson International Center for Scholars. <http://wilsonquarterly.com/stories/teach-to-the-test/>
15. *CCSSI Mathematics*. (November 7, 2013). “Computer-based assessments.” <http://ccsimath.blogspot.com/2013/11/computer-based-assessments.html>
16. Lyndsey Layton. (October 13, 2013). Elementary students learn keyboard typing ahead of new Common Core tests, *Washington Post*
17. PARCC has proposed cutting 90 minutes, presumably at each tested grade level. If implemented, a cut of that magnitude would represent about a 14% time reduction. PARCC’s current time accumulation of 95.65 hours would decrease to around 82 hours, still almost three times more than MCAS’ 29.76 hours. However, PARCC needed all of those 95.65 hours in order to sufficiently cover the content domain with its convoluted and time-consuming test item types. PARCC officials have yet to explain how shortening test administration times will degrade the technical quality of the PARCC test.
See: <http://hechingerreport.org/some-common-core-tests-are-getting-shorter-what-are-they-losing/> ;
<http://www.edweek.org/ew/articles/2015/05/21/parcc-shortens-its-common-core-test.html> ;
<http://www.parcconline.org/assessments/test-design/design-changes>
18. *Achieve, Inc. (October 2001). *Measuring Up: A Report on Education Standards and Assessments for Massachusetts*.
*Driscoll, Linda, et al. (2003). *Education Reform: Ten Years after the Massachusetts Education Reform Act of 1993, Education Connections*, Amherst, MA: University of Massachusetts School of Education.
<http://nonpartisaneducation.org/Review/Resources/EdConnection.htm>
*Garcia, Joseph and Robert Rothman. (2001). *Three paths, one destination: Standards-based reform in Maryland, Massachusetts, and Texas*. Achieve, Inc.
*Hambleton, R.K. (2006). Update on MCAS: Is it Working? Is it Fair? *Nonpartisan Education Review / Essays*, 2(7).
<http://nonpartisaneducation.org/Review/Essays/v2n7.htm>
*Jirka, Stephen, & Hambleton, Ronald K. (July 4, 2004). *Comparison of trends in NAEP, Massachusetts-NAEP, and MCAS Results*. Center for Educational Assessment MCAS Validity Report No. 6. (CEA-537). Amherst, MA: University of Massachusetts, Center for Educational Assessment.
*Mass Insight Education. (2002). *Taking Charge: Urban high school students speak out about MCAS, academics and extra-help programs*.
*Papay, J.P., Murnane, R.J., and Willett, J.B. (2010). The consequences of high school exit examinations for low-performing urban students: Evidence from Massachusetts. *Educational Evaluation and Policy Analysis*, 32(1), 5–23.

APPENDIX B. LINKS TO PUBLIC HEARINGS AND OTHER SOURCES OF PUBLIC COMMENT ON MCAS OR PARCC

1. Public hearings conducted in 2015 by the Massachusetts Secretary of Education and Board of Elementary and Secondary Education. The first hour or so usually consists of testimony invited by the Board of Elementary and Secondary Education. This testimony is followed by members of the general public (parents, teachers, others) who signed up to give testimony.

Beacon Hill

<https://www.youtube.com/watch?v=omiC6zDJvKE>

<https://www.youtube.com/watch?v=og74iEZoYdc>

<https://www.youtube.com/watch?v=UXARpQ-M1VY>

Bridgewater

<https://www.youtube.com/watch?v=CDLJZlFLYs>

https://www.youtube.com/watch?v=C18VqLdsG_8

<https://www.youtube.com/watch?v=1IRpyYkbBV4>

Bunker Hill

<https://www.youtube.com/watch?v=hgGqGYMpiXk>

<https://www.youtube.com/watch?v=cogTA8Jow1Q>

<https://www.youtube.com/watch?v=0MmGWjimtCk>

Fitchburg

<https://www.youtube.com/watch?v=dUgbp8jlK6I>

Lynn

<https://www.youtube.com/watch?v=HEaWSdmjZGs>

<https://www.youtube.com/watch?v=czX4Wee11OE>

<https://www.youtube.com/watch?v=HtAB2EsevZk>

Springfield

<https://www.youtube.com/watch?v=7aSgMr5teVY>

<https://www.youtube.com/watch?v=afwPHCgfk04>

<https://www.youtube.com/watch?v=jQA19FMPVMQ>

2. Communities that have passed a non-binding resolution/petition/ballot question in Town Meeting or a local election, on Common Core and/or Common Core-based tests: **Abington, Brookfield, Halifax, Hampden, Hansen, Holland, Lakeville, Norfolk, Orange, Tewksbury, Uxbridge, Whitman, Wilbraham, Worcester**

3. Other sources of public comment on MCAS or PARRC

[*http://www.baystateparent.com/baystateparent.com/commoncorema/](http://www.baystateparent.com/baystateparent.com/commoncorema/). See hyperlink to spreadsheet on Gates Foundation Education-Related Giving in Massachusetts, including:

Teach Plus, Boston—representatives advocated for the PARCC test at Board of Education forums

\$7,500,000.00 For general operating support

*<http://danverspublicschools.org/wp-content/uploads/2015/08/MASS-PARCC-Position-paper.pdf>

*<http://www.patriotledger.com/article/20150716/NEWS/150717598>

*<http://superintendentlps.blogspot.com/>

*<https://www.bostonglobe.com/metro/regionals/south/2015/05/28/two-views-whether-state-should-replace-mcas-with-parcc-test/0F3FPIbodjqH6icnNtZE1K/story.html>

*<http://www.doe.mass.edu/parcc/CommTool/WhyPARCC.pdf>

APPENDIX C. A RANDOMLY-CHOSEN TEST ITEM USED ON THE 2014 GRADE 10 MCAS MATH TEST

ID: 303264 C Common

A farmer harvested a total of 364 pumpkins. The pumpkins had an average weight of 10.9 pounds. Which of the following is closest to the total weight, in pounds, of the pumpkins the farmer harvested?

- A. 3,000 B. 3,300 C. 4,000 D. 4,400

This test item was coded to Standard 10. N.4 in the state's 2000 Mathematics Curriculum Framework. That standard read as follows: "Use estimation to judge the reasonableness of results of computations and of solutions to problems involving real numbers." The test item could be construed as meeting that standard but not at a grade 10 level. The test item itself requires knowledge only of the decimal system, multiplication, and estimation—all taught in the elementary school. Yet, as a "common" item (used in all forms of the grade 10 math test), it had gone through layers of review for two years.

As explained in DESE technical monographs, the process for test-item approval on MCAS is roughly as follows: First, each item that appears on a grade-level test may originate with the testing company. It is brought to the appropriate assessment development committee, which comes to consensus on recommendations for the wording of each item, the coding to a standard (or standards) for the item, its difficulty level, and type of skill (procedural, conceptual, or problem-solving). This committee and all other assessment development committees are chosen by DESE. According to DESE, the Bay State math teachers on the committee reviewing the "pumpkin" item (among others) for a recommendation on grade-level appropriateness were: Sharon DeCicco, Patricia Tranter, Denise Sessler, Ann-Marie Belanger, Alison Kellie, Kimberly Donovan, Michelle Bussiere, Paula Sweeney, Patricia Izzi, Clare Brady, and Deatrice Johnson.

The members of this DESE-chosen committee, however, were not the only ones that looked at the "pumpkin" item. A bias committee, also selected by DESE, examined the item for bias. Two paid mathematics professors from out-of-state, also selected by DESE, examined the item for mathematical accuracy.

DESE always does its own review. And it may reject an item if it chooses to. But it also gives final approval to what the test developer assembles as the set of items to appear on the next year's grade-level tests. This final set of items is based on statistics from tryouts and may be based further recommendations from DESE and the relevant assessment development committee.

In the final analysis, the commissioner of education and his staff at DESE are responsible for the "pumpkin" item and all the other test items on the grade 10 tests, whether or not they are on-grade-level, above-grade-level, or below-grade-level.



Copyright © 2015 Pioneer Institute for Public Policy Research. All rights reserved.

PIONEER INSTITUTE
PUBLIC POLICY RESEARCH

185 Devonshire Street, Suite 1101 | Boston, MA 02110 | 617.723.2277
www.pioneerinstitute.org | [Facebook.com/PioneerInstitute](https://www.facebook.com/PioneerInstitute) | [Twitter.com/PioneerBoston](https://twitter.com/PioneerBoston)