MATHEMATICA
Policy Research

REPORT

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

October 5, 2015

Ira Nichols-Barrer

Kate Place

Erin Dillon

Brian Gill

Submitted to:

Massachusetts Executive Office of Education

One Ashburton Place, Room 1403

Boston, MA 02108

Project Officers: Jill Norton and Tom Moreau

Contract Number: PO-15-1026-DOE02-DOE01-00000004280:1

Submitted by:

Mathematica Policy Research

955 Massachusetts Ave

Suite 801

Cambridge, MA 02139

Telephone: (617) 491-7900

Facsimile: (617) 491-8044

Project Director: Ira Nichols-Barrer

Reference Number: 50141

This page has been left blank for double-sided copying.

ACKNOWLEDGEMENTS

This page has been left blank for double-sided copying.

CONTENTS

This page has been left blank for double-sided copying.

TABLES

FIGURES

This page has been left blank for double-sided copying.

EXECUTIVE SUMMARY

In fall 2015, the state of Massachusetts will decide whether to continue using the Massachusetts Comprehensive Assessment System (MCAS) or adopt the new Partnership for Assessment of Readiness for College and Careers (PARCC) exam for testing the achievement of students in the state's public schools. Unlike the MCAS, which was designed to measure students' proficiency relative to statewide curriculum standards, the stated goal of the PARCC is to measure whether students are on track to succeed in college. Whether the PARCC examination succeeds in measuring college preparedness better than the MCAS has been an open question: no prior research has analyzed the extent to which PARCC test scores predict students' outcomes in college.

The Massachusetts Executive Office of Education commissioned this study to provide timely, rigorous evidence on the extent to which MCAS and PARCC test scores can accurately assess whether students will succeed in college (recognizing that this was not the original aim of MCAS). To examine this question, at the end of the 2014–2015 academic year state education agencies coordinated the administration of 10th-grade MCAS and corresponding PARCC assessments for a sample of first-year college students at 11 public higher-education campuses in Massachusetts. (PARCC's math assessments are designated for courses rather than grades, so the state selected the PARCC end-of-course assessments that should best align with the content assessed by the 10th-grade MCAS test.) Ideally, a study of predictive validity would be longitudinal, tracking the outcomes of students over at least three years from the point when they complete each exam to the end of their first year in college. But the state cannot wait that long to choose its assessments. By testing first-year students who are already in college, this study can immediately provide evidence regarding the college outcomes of students relative to their performance on the MCAS or PARCC exams.

For each test, we examine whether high-scoring students perform better in college than low-scoring students; if this is true, we would conclude that the scores have validity in predicting college outcomes. We also examine whether students who meet designated standards on the tests ("proficient" in the terms of MCAS and "college-ready" in terms of PARCC) are likely to be ready for college as indicated by their need for remedial coursework and by their ability to earn "B" grades in college.

The study's key findings are as follows:

- **Both the MCAS and the PARCC predict college readiness.** Scores on the assessments explain about 5 to 18 percent of the variation in first-year college grades, depending on the subject. MCAS and PARCC scores are comparable to SAT scores in predicting college outcomes. Similarly, MCAS and PARCC test scores both provide statistically significant predictions regarding which students will need remedial coursework in college.

- **The validity of scores on PARCC assessments in predicting college grades is similar to the validity of scores on the MCAS.** In English language arts, the PARCC end-of-year and performance-based assessment scores have a combined correlation with college grades (0.23) that is virtually identical to the corresponding correlation between MCAS English language arts test scores and college grades (0.23). For mathematics, the correlation with

college grades for scores on the two PARCC integrated math components (0.43) is also statistically indistinguishable from the association for MCAS math test scores (0.36).

- **Scores on both the MCAS and PARCC provide similarly strong predictions about which students need remedial coursework in college.** We find no consistent pattern suggesting that the PARCC test components outperform the MCAS in predicting students' scores on the Accuplacer exam (an assessment that determines assignment to remedial coursework at many campuses). Overall, both exams also have an equally predictive relationship with enrollment in remedial courses during the first year of college.

- **In math, meeting the PARCC standard for college readiness predicts a higher level of college performance than meeting the MCAS standard for proficiency, while in English language arts the two standards are similar.** In English language arts, students meeting the college-ready standard on the PARCC exam earn a 2.76 grade point average (GPA) in first-year college English courses, whereas students meeting the proficient standard on the MCAS earn a 2.66 GPA, and the difference is not statistically significant. In math, the difference is larger and statistically significant: students meeting PARCC's college-ready standard earn a 2.81 GPA in first-year college math courses and students at the proficient standard on the MCAS earn only a 2.39 GPA. In other words, the PARCC math standard is more closely aligned with achieving B-level college grades.

- **In math, students who achieve the college-ready standard on PARCC are also less likely to need remediation than students who achieve the proficient standard on MCAS, while in English language arts the two standards are not statistically distinguishable**. In English language arts, students meeting PARCC's college-ready standard were 8 percentage points less likely to have been assigned to remedial classes than students meeting the MCAS proficient standard, but the difference was not statistically significant. In math, students meeting PARCC's college-ready standard were 11 percentage points less likely to have been assigned to remedial classes than students meeting the MCAS proficient standard, and the difference was statistically significant.

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

It should be noted that PARCC offers assessments in other high-school courses that could not be included in this study; some of these assessments are typically given to 11th-grade students who would be one year closer to college entry than the students who now are required to take MCAS. We were able to examine only one of the PARCC assessments that would typically be given to 11th-grade students: the Algebra II end-of-year assessment. Scores on that assessment were not correlated with first-year college outcomes at rates any higher than the correlations found for the 10th-grade assessments.

Because the underlying scores on the MCAS and PARCC assessments are equally predictive of college outcomes, Massachusetts policymakers have more than one option to align high-school mathematics test standards with college readiness: either adopt the PARCC exam, or continue using MCAS while simply setting a higher score threshold for college readiness. Either

of these options would achieve the goal of ensuring that the state's high-school assessments provide better information about college readiness to students, parents, educators, and policymakers.

Predictive validity is only one consideration that is relevant to the state's selection of an assessment system. Even though PARCC and MCAS are similar in terms of their predictive validity relative to college grades and college remediation outcomes, they might differ in other ways. For example, the content knowledge, writing, and problem-solving skills measured across the tests' various components are not identical, meaning that the exams could differ in their validity with respect to other factors beyond college performance. Differences in the content and structure of the assessments could also promote different kinds of instructional practices as schools seek to prepare students for the assessments. A full comparison of all of the possible differences between the two examinations falls outside of the scope of this study, which was designed for the specific purpose of comparing the effectiveness of the two exams in identifying which students succeed in college.

This is the first study to examine the predictive validity of the PARCC examination compared with the state assessment it would replace. By examining rigorous evidence about the validity of these two exams, Massachusetts provides a model for other states weighing difficult choices about whether to keep or reform current statewide educational assessments.

I. BACKGROUND AND PURPOSE OF THE STUDY

Massachusetts faces an important policy choice regarding its educational assessment system. The Commonwealth is a member of the multi-state Partnership for Assessment of Readiness for College and Careers (PARCC), which has been engaged in developing next-generation student assessments in conjunction with the adoption of the Common Core State Standards. Like many other states, Massachusetts is considering whether to replace its current assessment system with PARCC assessments. In fall 2015, the Massachusetts Board of Elementary and Secondary Education will decide whether the state should continue to use the Massachusetts Comprehensive Assessment System (MCAS) or adopt the new PARCC exams.

Massachusetts is a national leader in establishing high quality learning standards for its students (Carmichael et al. 2010). In 1993, Massachusetts implemented some of the most rigorous state standards in the nation. Five years later, it began administering the MCAS to determine whether students met those standards. The MCAS, with its tight alignment to state standards and variety of question types, was widely considered a state-of-the-art assessment when it was introduced (Massachusetts Business Alliance for Education et al. 2015). Since the MCAS was implemented, Massachusetts' students have outperformed the rest of the country on the National Assessment of Educational Progress, and Massachusetts now ranks among the top-performing education systems in the world (National Center for Education Statistics 2013; Chang 2013).

Although the MCAS sets a high bar for assessing students' performance against state standards, it was not explicitly designed to assess whether students are also college and career ready. Despite the fact that 90 percent of students obtain a "proficient" or "advanced" score on the 10th grade MCAS exam, 35 percent of Massachusetts high school graduates require remediation when they enter college (Massachusetts Department of Elementary and Secondary Education 2015). In contrast, the stated goal of the PARCC is to measure whether students are on track to succeed in college and their careers. To pursue this objective, the 2015 version of the PARCC exam includes performance-based assessments (PBAs) in mathematics and English language arts (ELA) that use questions with open-ended responses, in addition to more conventional end of year (EOY) components that rely on multiple-choice questions. The MCAS exam also uses a combination of multiple-choice questions and short answer questions. Whether the PARCC examination succeeds in measuring college preparedness better than the MCAS is an empirical question; answering it requires a rigorous, independent empirical analysis examining which test better predicts college outcomes.

This study provides timely evidence to help reveal whether the MCAS or PARCC better predicts students' success in college, and to help the state of Massachusetts decide whether to replace the MCAS with the PARCC assessment. The study's primary outcome of interest is the strength of associations between first-year college grades and MCAS or PARCC test scores. In addition, the study examines whether each test predicts scores on the Accuplacer exam (a measure of students' academic ability when they first enter college) and whether students are assigned to remedial coursework in their first year of college. We also conduct exploratory and descriptive analyses to examine the distribution of scores on these tests and associations with students' prior academic performance before college as measured by SAT scores (an additional measure of college preparedness).

To examine these questions, at the end of the 2014–2015 academic year the Massachusetts Executive Office of Education, Department of Elementary and Secondary Education, and Department of Higher Education worked together to coordinate the administration of the MCAS and PARCC assessments for a sample of first-year college students at 11 public higher-education institutions in the state.[1] By administering the exams to first-year students who are already in college, this design generates immediate evidence regarding the college outcomes of students relative to their performance on the MCAS or PARCC exams.

However, this approach also has limitations. The study sample is limited to enrolled college students at public institutions in the state, who might not be representative of the statewide population of high school students. One reason for this is that testing students who are already in college misses the students who did not enroll in college or who dropped out of college before the spring semester. Another reason is that even for the test-takers in the study, students' academic growth since 10th grade might differentially affect performance on the PARCC or MCAS tests. In addition, due to the time burdens of completing these exams in full, the study could recruit students to take only one component of the MCAS (which has two components of interest) or the PARCC exam (which has five components of interest). As a result, our analysis depends on additional assumptions to predict the combined validity of multiple test components at the same time.

Addressing these methodological concerns would require a longitudinal study that tracks the outcomes of students over three years from the point when they complete each exam in 10th grade through the end of their first year in college. Policymakers choosing between the two exams in 2015 cannot wait that long to make decisions about the tests. This study provides important, timely, and useful information regarding the validity of these two examinations and represents a valuable piece of evidence for Massachusetts to consider in deliberations about which assessment system to use in coming years. This is also the first study of its kind. To date, no reliable evidence demonstrates whether the new Common Core-aligned assessments provide accurate information about which students are prepared for success in college. By examining rigorous evidence about the validity of these tests, Massachusetts provides a model for other states considering difficult choices about whether to change their current statewide assessment systems.

[1] The Massachusetts Executive Office of Education commissioned this study. The authors of this report are employees of Mathematica Policy Research, with whom the Executive Office of Education contracted to develop and implement the study's methods and analyses. The Mathematica staff who conducted the project do not have financial interests that could be affected by the content in this report.

II. STUDY DESIGN AND SAMPLE

This study uses a random assignment design to measure whether scores on the MCAS and PARCC tests can identify students who succeed in college, a concept we refer to as the predictive validity of each exam. The study sample consists of 866 first-year college students who graduated from a Massachusetts high school and subsequently enrolled at one of the 11 public in-state campuses participating in the study.[2] These colleges and universities include institutions from across the state (Figure II.1), and participating institutions consist of 6 community colleges, 3 state universities, and 2 University of Massachusetts institutions. This breakdown roughly mirrors public higher education in the state, which consists of 15 community colleges, 9 state universities, and 4 University of Massachusetts campuses. Campuses in the study enroll almost half of the college students in the state's public system.

Figure II.1. Locations of the 11 study campuses
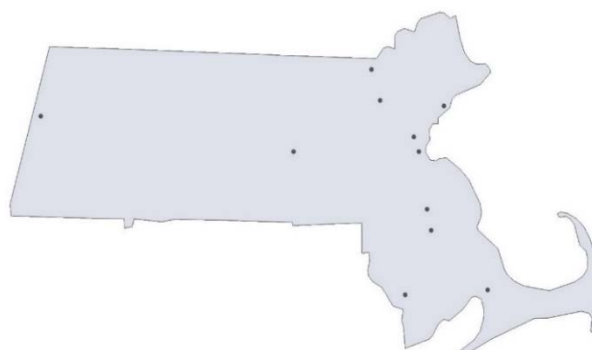
Students included in the study may differ somewhat from the statewide population of all high school students in Massachusetts. The study sample does not include students who did not enroll in college, and the sample also omits students who attended college in the fall semester but left college before or during the spring semester. The study includes only a subset of the state's public institutions of higher education and none of the private colleges and universities. Nonetheless, the data suggest that the students participating in this study do not differ greatly from the state as a whole. In our study sample, students' average scaled score from their 10th grade MCAS exams (256 in ELA and 258 in math) is between two and five points higher than the statewide average.[3] We also examined if the distribution of students' scores on the study-administered MCAS test was representative of the statewide distribution for high school students who took the same version of the exam. The statewide standard deviation of MCAS scores for all 10th graders (12.0 in ELA and 16.6 in math) is very similar to the standard deviation of study-administered MCAS scores in this sample (12.4 in ELA and 16.9 in math). In other words, the variation in MCAS scores found in our study data is very similar to the variation found in the state as a whole.

Within each of these 11 campuses (Table II.1), all eligible students received notices about the study and those who volunteered to participate were randomly assigned to complete one component of either the MCAS or PARCC exam. The great advantage of this random assignment procedure is that it ensures that students taking PARCC assessments were not

---

[2] The Massachusetts Executive Office of Education and Department of Higher Education recruited these 11 schools to provide a representative sample of public education institutions in the state.

[3] In our sample students took the 10th grade MCAS exam either in 2011 or 2012. In 2011, the statewide average 10th grade MCAS score was 252 in ELA and 253 in math, and in 2012 the average was 254 in ELA and 253 in math. The average 10th grade score in the study sample does differ from each of these averages by a statistically significant margin (two-tailed test), even though the magnitude of the difference is not large.

systematically different from students taking MCAS components. Because equivalent groups of students completed each test component (as we demonstrate in Appendix A), the study can rigorously compare these groups to determine which test has the strongest association with students' success in college, as measured by GPA and the need for first-year remedial coursework.

Table II.1. Overview of participating colleges and universities

Institution	Institution type	College enrollment	Study sample
Berkshire Community College	Community College	3,045	30
Bristol Community College	Community College	12,514	34
Bunker Hill Community College	Community College	19,871	48
Massasoit Community College	Community College	11,950	47
Middlesex Community College	Community College	12,999	55
Quinsigamond Community College	Community College	11,926	40
Bridgewater State University	State University	10,807	209
Massachusetts Maritime Academy	State University	1,535	104
Salem State University	State University	9,397	122
University of Massachusetts-Boston	University of Massachusetts	12,700	82
University of Massachusetts-Lowell	University of Massachusetts	12,993	95

Sources: Massachusetts Department of Higher Education Data Center, "Annual Unduplicated Student Headcount: FY 2014." University of Massachusetts Office of the President, "Facts 2014–2015."

In each campus, the study recruited eligible students (first-year enrollees who were Massachusetts residents when they graduated from high school) to participate in the study on a voluntary basis. In total, the study recruited a sample of 866 students to participate. As part of the recruitment process, students were offered gift cards as a participation incentive plus an additional incentive designed to encourage effort on the exams.[4] All participating students signed consent statements agreeing to release their test scores, college transcripts, and high school records for analysis as part of the study.

Within each campus, students were randomly assigned to one of seven different testing groups (Table II.2) using a simple sign-in sheet that randomly placed each student in a different test depending on the order they appeared at the testing center. The study examined 2015 versions of the PARCC tests, and selected the PARCC exams that best align with the content of the 10th grade MCAS tests.[5] In addition, the study included the PARCC's end-of-course Algebra II exam, to explore whether an exam that is more aligned with 11th grade content has more predictive validity than other PARCC tests. The PARCC system also includes an additional set of tests designed for grade 11 (both in ELA and in integrated math) and other end-of-course math

---

[4] In addition to the initial participation incentive, students were told that one difficult and one easy question from the exam would be selected at random and scored after they completed the test, and that participants would receive a $25 gift card for answering one of these questions correctly and a $50 gift card for answering two of these questions correctly.

[5] In the 2014-2015 school year, some Massachusetts high schools chose on a voluntary basis to participate in the PARCC exam, in addition to participating in the (required) 10th grade MCAS exam. However, none of these schools took PARCC's grade 10 ELA or Math II tests, to avoid burdening the same 10th grade students with both exams in the same year.

exams (including Algebra I and Geometry); this study did not examine the predictive validity of those other PARCC examinations. The study used the 2014 version of each MCAS test.

Table II.2. List of MCAS and PARCC test components in the study

Test component	Description
MCAS Math (2014 test form)	Grade 10 MCAS exam in math (paper-based test mode)
MCAS ELA (2014 test form)	Grade 10 MCAS exam in ELA (paper-based test mode)
PARCC Integrated Math II performance-based assessment (2015 test form)	End-of-course exam in math (paper-based test mode)
PARCC Integrated Math II end-of-year assessment (2015 test form)	End-of-course exam in math (paper-based test mode)
PARCC ELA performance-based assessment (2015 test form)	Grade 10 exam in ELA (paper-based test mode)
PARCC ELA end-of-year assessment (2015 test form)	Grade 10 exam in ELA (paper-based test mode)
PARCC Algebra II end-of-year assessment (2015 test form)	Advanced algebra end-of-course assessment (paper-based test mode)

The study uses two primary data sources: study-administered MCAS and PARCC test score data for the study sample and a data file recording students' college courses, grades, and Accuplacer scores. In addition, the study also received data on these students' high school test scores, including 10th grade ELA and Math MCAS scores and SAT scores for students who opted to take the SAT in high school. We received college transcript data for all students in the sample (93 percent of the students took at least one English course, and 79 percent took at least one math course), and we received data on prior 10th grade MCAS scores for 96 percent of the sample. We cleaned and prepared each individual data set separately and then merged the two data sets using student-level study ID numbers provided in each file. All of the study's data files were fully de-identified, meaning the data did not include students' names, dates of birth, or demographic information.

A. Data preparation for 2015 MCAS and PARCC scores

To prepare the 2015 MCAS and PARCC data file, it was necessary to identify and eliminate student outliers who did not complete the exam or did not make a good-faith effort to answer the questions correctly. In total, we removed the data of 19 students due to evidence of low effort on the exam. First, we eliminated students who did not complete their assigned MCAS or PARCC exams (these students were flagged by the MCAS and PARCC test-scoring entities and did not receive final scores), as well as students who took more than one exam.[6] After eliminating these cases, we performed a series of analyses to identify additional outliers. We examined the distribution of raw MCAS and PARCC scores in histograms, looking for spikes in scores around the lower end of the distribution, which could indicate that a group of students did not finish the test or randomly completed the test. Next, we assessed students' performance on the 2015 exams relative to their prior test scores to identify students who performed much worse relative to their

---

[6] Seven students took a study exam more than once, and we observed a pattern of low effort in their study-administered test scores; see Appendix A for further details.

peers on the 2015 test, in comparison to their performance on high school assessments.[7] Finally, we used information on the guessing thresholds for each test, meaning we identified whether students scored at or below the score that a student would be expected to achieve, on average, if he or she randomly guessed on all of the exam's multiple-choice questions.

This analysis revealed that patterns of low effort and guessing were rare in this sample.[8] In all, only 19 students (2 percent) were removed from the analytic sample for these reasons, reducing the original sample of 866 students to a final sample size of 847 students. A more detailed description of the outlier analysis, including summary statistics for the dropped students, can be found in Appendix A.

All of the study's main analyses use "raw" MCAS or PARCC scores, meaning the scores represent the number of questions that each student answered correctly. We chose to use raw scores primarily because scaled PARCC scores were not made available in time for the study; the scaling procedure used by the MCAS exam does not affect our results.[9]

B. Standardizing college course grades

Another important step in preparing the data was to standardize course grades across institutions and course types. This study examines the relationship between test scores and students' readiness for college coursework, using course grades on a 0-to-4 scale as a measure of college performance. But along with college performance, course grades also reflect the difficulty of the course and rigor of the institution's grading standards. More demanding grading standards at some institutions, for example, could lead to lower overall grades among those

---

[7] To compare students' performance on the 2015 exams relative to their prior test scores, we first standardized the 2015 MCAS and PARCC scores, the grade 10 MCAS scores, and the SAT scores. Students with extreme differences in high school and 2015 scores (declines of more than two standard deviations) and students who appeared as outliers on scatter plots of 2015 standardized test scores versus high school standardized test scores were flagged.

[8] For the two MCAS testing groups in the study, we also examined their effort in an additional way: comparing their study-administered test results (on the 2014-version of the MCAS exam) to statewide results in 2014, when 91 percent of 10th grade test-takers in Massachusetts received scores rated as proficient or higher in ELA and 79 percent scored as proficient or higher in math. On the study-administered test the results were somewhat lower: 66 percent of the study's MCAS sample scored as proficient or higher in ELA, and 75 percent of study sample scored as proficient or higher in math. We investigated whether these differences are driving the study's results by comparing the predictive validity of high school MCAS scores to the study-administered MCAS scores in our data (see Appendix D); those results show that, despite these differences, the students' study-administered test scores remain an appropriate proxy for the prior scores they received in 10th grade.

[9] We examined the predictive validity of MCAS scaled scores, and found correlations that were nearly identical to the correlations with MCAS raw scores (see Appendix D). Also, the scaling procedures used by MCAS and PARCC do not change the relative ranks of students as measured by their total raw score; for example, a student scoring at the 50th percentile of raw scores would remain at the 50th percentile using scaled scores. As a result, the scaling process for PARCC and for MCAS is not likely to produce large changes in the relationships we observe between the raw scores on each exam and college outcomes (although it is possible that non-linear scaling on the PARCC exam could change the correlations we report in this study). This is also why it was possible for PARCC to identify which students meet the exam's "college and career ready" standard using raw scores, before the scaling process for these scores was completed: converting raw scores to scaled scores does not change students' proficiency rating on the PARCC exam.

schools' students, irrespective of the students' college preparedness. Similarly, particular subject areas might be more challenging, leading to lower grades for students who take more courses in those subjects.[10]

Failing to account for these differences could bias the study's findings. Grades in remedial courses, for instance, represent a different level of college readiness than grades in nonremedial courses. We therefore standardized students' course grades to establish more consistency across the study sample. To do this, we used a two-step process. First, we adjusted grades for whether the course was a remedial course. Second, we adjusted grades for the institution and course subject.

In our first step, we compared students' grades in remedial courses with grades those same students received in college-level courses. We examined this separately for math and ELA classes and found that, on average, students tended to receive higher grades in remedial courses than in college-level courses (as might be expected). For example, students' GPAs in remedial math courses were 0.71 points higher, on average, than those same students' grades in their first-year college-level math courses.[11] This indicates that it was easier for these students to obtain a higher grade in remedial courses, compared with nonremedial courses in the same subject. To account for this difference in difficulty, we reduced remedial course grades in math and English language arts by the average within-student difference between remedial and college-level course grades. Specifically, we reduced the observed grades in remedial math courses by 0.71 points and remedial ELA courses by 0.36 points.

In our second step, we adjusted for differences in students' grades across institutions and by subject area. To do so, we reduced grades at institutions and in academic subjects in which the average grade was high compared with the full study sample, and increased grades at institutions and in subjects in which the average grade was unusually low compared with the full sample. We then calculated students' GPAs by averaging their adjusted course grades across all courses, and by course subject (math, ELA, and other subjects). Appendix B provides a detailed description of both the original and adjusted course grades, and it provides more information about the study's grade standardization procedures. This approach to standardizing grades assumes that the academic ability of students is (on average) similar across the study's campuses and similar for students who choose to seek classes in each academic subject. Because these assumptions might not hold in some instances, we also examine the effect of adjusting grades on the study's outcomes by repeating our analyses with the original, unadjusted grades and (separately) using an alternative standardization procedure that accounts for the selectivity of different campuses; these results—which are not notably different from the main results discussed below—are described in Appendix D.

---

[10] The study was also limited to examining correlations with course grades in the first year of college, which may differ to some extent from correlations with grades in more advanced courses that occur in later years of college (although we consider this to be unlikely).

[11] Throughout the report, we define grades using a 0-to-4 GPA scale, with a grade of F equal to 0 points and a grade of A equal to 4 points.

This page has been left blank for double-sided copying.

III. ANALYSIS AND RESULTS

Our analyses assess the extent to which scores on the PARCC and MCAS assessments are correlated with students' college performance (as measured by GPA), college readiness (as measured by Accuplacer scores), and placement in remedial courses (as measured by students' course-enrollment data). To do so, we examine the *correlation coefficient* for each test, which is a statistical measure of the relationship between test scores and college outcomes. The correlation coefficient provides a common benchmark to summarize the relationship between two variables. Across all possible scores on a given test, this measure summarizes how well the values of the test predict students' performance on each college outcome. Correlations have a minimum value of -1 and a maximum possible value of 1; a correlation coefficient equal to 1 would imply that a test perfectly predicts college outcomes, in the sense that variation in test scores (across all observed test scores) perfectly explains all of the variation observed in students' college performance. A correlation coefficient equal to 0 would imply that test scores provide no predictive information about students' future performance in college. As a benchmark, in our study data students' high school SAT scores (the sum of their scores on the SAT math, reading, and writing components) has a correlation of 0.27 with students' standardized grades in college.

Students in the study had all completed or nearly completed their first year of college when they took the MCAS or PARCC exam, so in a literal sense the study is only able to measure the *concurrent* validity of MCAS or PARCC test scores relative to college grades. However, because our data includes students' 10th grade MCAS scores we can directly compare high school MCAS scores to study-administered MCAS scores in the sample, examining whether the two sets of scores are similar. The correlation between MCAS scores at these two points in time is reasonably strong (0.71 in math and 0.51 in ELA).[12] In addition, the correlation between 10th grade MCAS scores and college GPA in the tested subject (0.31 in math and 0.20 in ELA) is very similar to the correlation we observe between study-administered MCAS scores and GPA (0.32 in math and 0.19 in ELA) in our data. This suggests that study-administered MCAS scores are providing a reasonably good proxy for MCAS scores in high school.

Another important consideration in the analysis pertains to the study's limited sample size. The study has a final analytic sample of 847 students, spread approximately evenly across seven different tests (two MCAS testing groups and five PARCC testing groups). With only a limited number of students taking a given test, the analysis does not have enough statistical power to detect small differences between the predictive validity of these exams. In our sample, there must

---

[12] While these correlations are reasonably strong, they are not perfect: on average, students tended to perform somewhat worse on the study-administered MCAS exam than they did taking the exam in 10th grade. We investigated directly whether these differences are driving the study's results by comparing the predictive validity of high school MCAS scores to the study-administered MCAS scores in our data (discussed here, and in Appendix D). Despite these differences, the students' study-administered test scores appear to be highly representative of the prior scores they received in 10th grade, in terms of the relationship between test scores and GPA.

be a difference of roughly 0.2 or greater between an estimated correlation for MCAS and an estimated correlation for PARCC to statistically distinguish between the two exams.[13]

Because each student in our sample took only one component of the MCAS or PARCC exams, we begin by comparing each individual component of PARCC against the relevant component of MCAS; for example, we separately contrast the predictive validity of the MCAS math component against the three PARCC math components in our sample: the Math II end-of-year (EOY) assessment; the Math II PBA); and the Algebra II EOY assessment. However, the PARCC exam is designed to *combine* scores from PBA and EOY assessments in each subject (the PBA tests are administered before the end of the school year, because they use more open-ended test questions and therefore take more time to score). To the extent that PBA and EOY test forms measure different aspects of college preparation in math or ELA, it is possible that the combination of these forms will have a substantially different amount of predictive validity than each component does in isolation.

Because we observe a score on only one test component per student, we cannot observe students' combined scores across all of the MCAS or all of the PARCC. To examine this issue, the study instead uses additional data on the correlations between these exam components as observed among test-takers outside the study sample. The Department of Elementary and Secondary Education provided MCAS between-component correlations, and PARCC between-component correlations were provided by the exam's publisher, Pearson, Inc. This analysis requires an important assumption—namely, that the between-component correlations observed outside the study sample (which are based on all high school test-takers, regardless of whether they are on track to graduate or enroll in college) apply equally well to the students in the study sample (all of whom are college enrollees). A more detailed explanation of the statistical procedures and assumptions used in this multicomponent analysis can be found in Appendix C.

A. Correlations with college GPA

The study's primary indicator of success in college is students' standardized GPA. We begin by separately analyzing correlations with college GPA for each of the seven different MCAS and PARCC test components in the study. We then examine the combined predictive validity of multiple related components for the same test—for example, by examining the combined predictive validity of the PARCC's integrated math performance-based and EOY assessments. These combined correlations may be most relevant to the state's decision, because they compare all of the information provided by the PARCC exams in the relevant subject with the complete information provided by MCAS in that subject. Finally, we describe the college outcomes of students falling into the various performance levels that the MCAS and PARCC define using cut-scores selected for each exam.

---

[13] We calculated the minimum difference in correlations that would be statistically significant at the 0.05 level, using a two-tailed test in our data (i.e., the minimum detectable difference with 50 percent power). Here we report this minimum detectable difference for a comparison of the combined correlation with GPA of two PARCC components (EOY and PBA) with the relevant correlation for the MCAS test in the same subject. The minimum detectable difference for the comparison of math subject tests is 0.20, and the minimum detectable difference for the comparison of ELA subject tests is 0.21.

1. Correlations between single test components and GPA

Our analysis measures the correlation between each of the seven MCAS and PARCC test components and standardized college GPA, and then compares each of these PARCC test correlations with MCAS test correlations in the same subject. For example, we compare the correlation between MCAS ELA raw scores and GPA in college English courses to the correlations for PARCC ELA scores (separately for the EOY and PBA test components). For each pair of correlations, we test whether the difference between the two correlations is statistically significant.[14] In Table III.1, we present 10 pairs of correlations in total: 5 pairs with respect to total GPA, 2 for ELA GPA, and 3 for math GPA.

Table III.1. How well do MCAS and PARCC components predict first-year college GPA?

GPA	MCAS	PARCC PBA	PARCC EOY	PARCC Algebra II
ELA				
Total GPA correlation	0.25	0.17	0.26	n.a.
[number of students]	[120]	[113]	[126]	
ELA GPA correlation	0.23	0.13	0.26	n.a.
[number of students]	[109]	[110]	[116]	
Mathematics				
Total GPA correlation	0.36	0.34	0.21	0.18
[number of students]	[129]	[116]	[121]	[122]
Math GPA correlation	0.36	0.37	0.40	0.24
[number of students]	[100]	[88]	[98]	[101]

Sources: College data from the Massachusetts Department of Higher Education and MCAS and PARCC test score data.

Note: We compared the MCAS correlation to each PARCC correlation, separately for ELA and math, and none of the differences were statistically significant at a 0.05 level using a two-tailed test.

n.a. = not applicable.

Overall, our results suggest that both exams predict college GPA and that they perform equally well for this sample of students. Across the 10 pairwise comparisons in our main analysis, none of the differences are statistically significant. We found no significant differences between the MCAS and PARCC correlations with adjusted total GPA, adjusted ELA GPA, or adjusted math GPA.

These correlations between test scores and GPA are modest in size, ranging from 0.07 to 0.40. The highest correlations are found among the math tests; for instance, the correlation between MCAS math scores and math GPA is 0.36, and the correlations between math GPA and PARCC Math PBA and PARCC Math EOY are 0.37 and 0.40, respectively. The correlations between the ELA tests and adjusted ELA GPA are lower, ranging from 0.13 to 0.26.

---

[14] The individual correlations were calculated by regressing each GPA type on each of the seven test forms. The correlations represent the square root of each model's R^2 statistic. To assess whether the correlations in each MCAS-PARCC pair were statistically significantly different from each other, we used Fisher's r-to-*z* transformation and then calculated the *p*-value for the test statistic, which is the difference in the resulting *z*-scores.

The study included an analysis of the PARCC Algebra II EOY test to examine whether a PARCC exam that is designed to measure learning in 11th grade has a higher correlation with college performance than PARCC's Math II EOY or PBA exam. We do not observe any evidence suggesting that this is the case: with respect to math GPA, the correlation for PARCC Algebra II (0.24) is actually somewhat lower than the correlation for the Math II EOY test (0.40) and the Math II PBA test (0.37). Note, however, that there is also an Algebra II PBA component that was not administered to students as part of this study, so we cannot determine whether the overall Algebra II exam has higher or lower predictive validity than PARCC's overall Math II exam. In addition, we found that a substantial number of students received minimal scores on the Algebra II test in this sample (see Appendix A). This may have occurred because students were given the Algebra II exam (an end-of-course test) regardless of whether they had taken algebra coursework in the recent past. If there is a less pronounced floor effect among high school students (due to the fact that they would be taking the Algebra II test immediately after finishing an algebra course), it is possible that high school students' scores on this exam would have a higher correlation with college GPA than the correlation we observe in our data.

2.   Correlations between multiple test components and GPA

In addition to examining the relationship between college GPA and scores on individual PARCC and MCAS test components, we completed a combined analysis that pools the PARCC PBA and EOY scores in each subject. These comparisons summarize the predictive validity of the entire PARCC exam in math and (separately) ELA and therefore may be most relevant to the state's choice of assessments. Because PARCC plans to combine the EOY and PBA components into a single test in 2016, these results may also be the most relevant to future administrations of the assessments (though it is impossible to be sure how the 2016 changes to PARCC will affect its predictive validity). Appendix C provides a detailed description of our process for calculating the relationship between GPA and these test combinations.[15]

In addition, we also examined PARCC PBA components (ELA and math) separately from the EOY components. Unlike the EOY tests, the PARCC PBA components feature more open-ended questions alongside multiple-choice questions, and the PBA exams are administered earlier in the school year due to the time needed to score students' exams. Because the PBA components and their scoring procedures are more complex than the EOY components, it might be of interest to test designers and policymakers to assess whether the PBA components predict college outcomes more or less successfully relative to PARCC's EOY components.

Consistent with these findings, this analysis did not reveal any statistically significant differences between the PARCC and MCAS exams (Table III.2). For example, the correlation between total GPA and the combined PARCC ELA tests, 0.25, is the same as the correlation

---

[15] Note that the correlations for the individual test components presented earlier in this chapter are not directly comparable with the correlations for combined test components shown here. This is because different samples are used in each analysis. The PARCC ELA EOY test, for example, has a correlation of .26 with ELA GPA, but that correlation declines to .25 when the ELA EOY test is combined with the PARCC ELA PBA test in our analysis. Typically, combining two tests would produce a higher correlation. That does not happen in this case because the standard deviation of GPA for students who took the PARCC ELA EOY test is lower than the standard deviation of GPA for the *pooled* sample of students who took either the EOY or PBA test in our sample. This issue is discussed in further detail in Appendix C.

between total GPA and the MCAS ELA test. Similarly, the correlation between the PARCC ELA tests and ELA GPA and between the MCAS ELA test and ELA GPA are both 0.23.

Table III.2. Do the MCAS and PARCC tests predict GPA when 2015 PARCC test components are combined?

	MCAS	PARCC PBA and EOY
ELA		
Total GPA correlation	0.25	0.25
[number of students]	[120]	[239]
ELA GPA correlation	0.23	0.23
[number of students]	[109]	[226]
Mathematics		
Total GPA correlation	0.36	0.37
[number of students]	[129]	[237]
Math GPA correlation	0.36	0.43
[number of students]	[100]	[186]

Sources: College data from the Massachusetts Department of Higher Education and MCAS and PARCC test score data.

Note: We compared the MCAS correlation to each PARCC correlation, separately for ELA and math, and none of the differences were statistically significant at a 0.05 level using a two-tailed test.

The story is similar for the PARCC and MCAS math tests. The correlation between the math tests and total GPA are essentially the same: 0.36 for the MCAS math test and 0.37 for the combined PARCC math tests. The difference between the exams is slightly larger when looking at math GPA. The MCAS math test has a correlation of 0.36 with math GPA, the same as its correlation with total GPA. But the PARCC math tests have a correlation of 0.43 with math GPA. These differences are not large enough to be statistically significant, however, meaning that the PARCC and MCAS math exams are statistically indistinguishable in our analysis.

These results and the results from the analysis of individual test scores both find that the PARCC and MCAS exams are similar in their relationship with college GPA. To help place the strength of these correlations in context, Figure III.1 shows the subject-specific correlations on a scale from zero, which indicates that exam scores predict none of the variation in students' GPA, to one, which indicates a perfect relationship in which exam scores predict 100 percent of the variation in students' GPA. A correlation of 0.5 is considered a strong correlation in social science research (Cohen 1988).

As an additional benchmark, we note the correlation in our sample between students' SAT scores, a known measure of college readiness, and GPA in the tested subject. The comparison with SAT scores is only illustrative and does not provide a true measure of the validity of the SAT in comparison with the other two exams. Because the SAT scores were recorded one to two years ago during high school, we would expect to see lower SAT correlations than what is observed with the 2015 administration of MCAS and PARCC tests. However, in our data study-administered MCAS scores have a very similar correlation with GPA compared with the correlation for students' prior MCAS scores in 10th grade. In fact, the correlation between SAT scores and college GPA in our data is reasonably similar to the results reported by the College

Board, once methodological differences have been taken in to account. The College Board reports a correlation of 0.33 between the SAT reading exam and first-year English grades, 0.37 between the SAT writing exam and first-year English grades, and 0.52 between the SAT math exam and first-year math grades (Mattern et al., 2012). These correlations differ from our results primarily because the College Board study uses a different set of statistical adjustments to analyze the variation in students' grades within individual courses; this approach is not possible in our study, due to the limited number of students taking the same course in our data.[16] Without these adjustments, the correlations with first-year grades in the College Board analysis are closer to those in our sample: 0.16 for the SAT reading exam, 0.22 for the SAT writing exam, and 0.28 for the SAT math exam.

As shown in Figure III.1, MCAS and PARCC correlations are similar to each other, and both exams are at least as correlated with college grades as the SAT scores of students in this sample.

Figure III.1. How large are the correlations between MCAS, PARCC, or SAT scores and GPA in each subject?
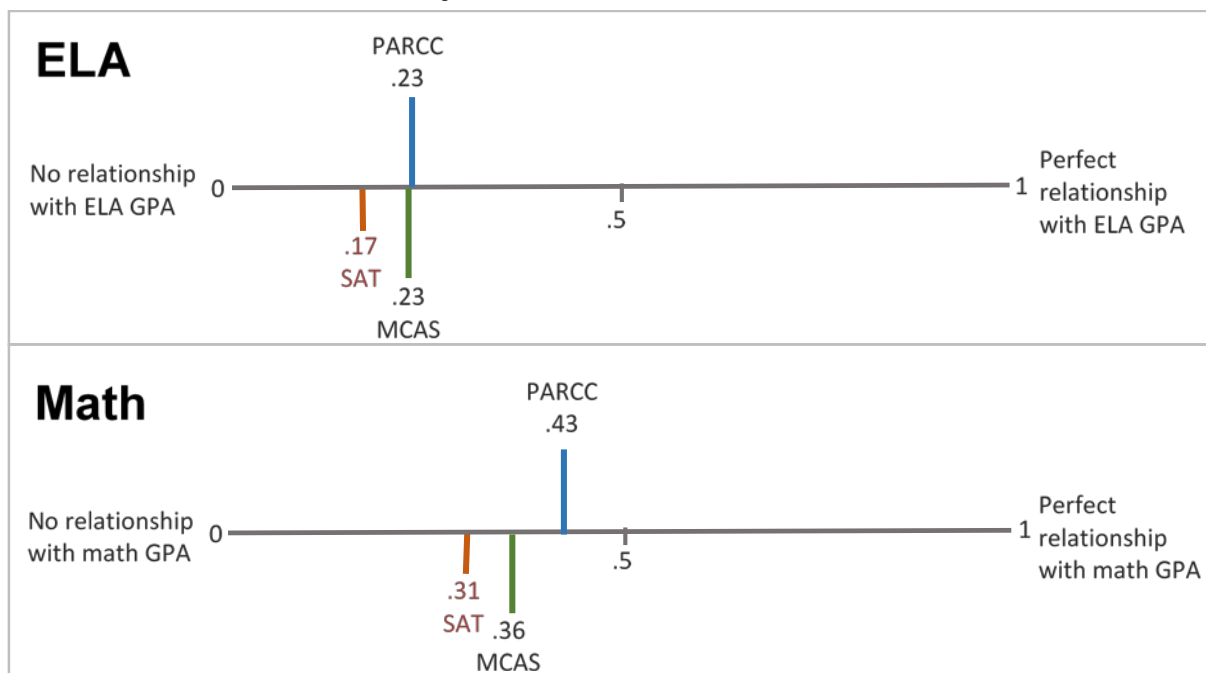
Note: The top panel illustrates the relative strength of correlations between standardized ELA GPA and the SAT (reading plus writing), MCAS (ELA), and PARCC (ELA PBA and EOY). The second panel illustrates correlations between standardized math GPA and the SAT (math), MCAS (math), and PARCC (math PBA and math EOY). Differences between the MCAS and PARCC correlations are not statistically significant.

---

[16] The College Board study calculates correlations using a weighted average of within-course correlations (meaning they calculate correlations between course grades and SAT scores separately for each college course). As part of that analysis, the College Board applies a "restriction of range" statistical adjustment to account for the limited range of data within courses, and this adjustment increases the overall magnitude of the correlations they report.

We also examined correlations between scores on the Accuplacer—an exam administered to students as they begin the first year of college—and college grades. Colleges use Accuplacer scores to determine whether students should be placed in remedial courses, so scores on this exam are designed specifically to measure students' readiness to perform college-level coursework. The correlation between Accuplacer Reading test scores and ELA grades (0.17, not shown in Figure III.1) is very similar to the corresponding correlation for the SAT. The relationships between math GPA and Accuplacer's College Math, Arithmetic, and Algebra scores (correlations of 0.28, 0.30, and 0.39, respectively) are somewhat higher than the correlations in ELA, and are also roughly similar in magnitude to the MCAS and PARCC correlations presented in the figure.

We found somewhat larger differences between the tests when we examined correlations for the PARCC separately by the type of PARCC component—the EOY component with fewer open-ended questions and the PBA component with more open-format items (Table III.3). The combined MCAS tests (math combined with ELA) have a correlation of 0.37 with total GPA, whereas the two PARCC EOY tests have a correlation of 0.28 and the two PARCC PBA tests have a correlation of 0.30. These differences between tests are not statistically significant, however. The cross-subject comparisons should be interpreted with caution because the MCAS results represent the entirety of the MCAS Math and ELA exams, but the PARCC results represent only half of the components (either EOY or PBA) for each academic subject.

Table III.3. Combined correlations for the PARCC PBA and EOY components

	MCAS ELA and Math	PARCC ELA and Math PBA	PARCC ELA and Math EOY
Total GPA correlation	0.37	0.30	0.28
[number of students]	[249]	[229]	[247]

Sources: College data from the Massachusetts Department of Higher Education and MCAS and PARCC test score data.

Note: We compared the correlation between the MCAS exams and GPA to the correlations between the combined PARCC exams and GPA, and none of the differences were statistically significant at a 0.05 level, using a two-tailed test.

3. College GPA across MCAS and PARCC performance levels

Although correlation coefficients summarize each test's relationship with GPA across the distribution of all possible scores, in practice the cut-scores used to define performance levels on each exam are likely to be important as well. In Massachusetts, high school students are required to achieve a "needs improvement" score in both math and ELA as a graduation requirement, and the percentage of students achieving proficiency thresholds on the state assessment has consequences for schools under federal and state accountability regimes. The MCAS exam has defined four different performance categories. In our sample, 75 percent of MCAS students received at least a proficient score in math and 66 percent scored as proficient or greater in ELA. The PARCC examination has defined five different performance categories relative to two parts of the test: (1) the combined score on the PBA and EOY grade 10 ELA exam; and (2) the combined PBA and EOY score on the integrated math II exam. The PARCC consortium specifies that students scoring in the two highest performance categories (category 4 or 5) should

be considered college and career ready in that subject.[17] In our study data, 60 percent of PARCC students scored as college and career ready in math and 66 percent scored as college and career ready in ELA.

The PARCC consortium has adopted a specific goal for the college and career readiness standard: the standard seeks to identify students who have at least a 75 percent chance of earning C-level grades in college. We examined if the PARCC standard meets this goal by modelling the relationship between PARCC scores and the likelihood of obtaining a GPA of 2.0 or better, and then calculating this likelihood at the PARCC cut-score for college and career readiness (that is, the lowest possible score in performance category 4).[18]

We find that the PARCC exam's college-readiness standard meets its stated target in both subjects. In ELA, students at the college-ready cut score have an 89 percent probability of earning at least a C average, and in math, students at the PARCC college-ready cut score have an 85 percent probability of earning a C average or better (Table III.4). For comparison, students at the MCAS cut score for proficiency have an 89 percent probability of earning at least a C average are in ELA and a 62 percent probability of earning at least a C average in math.

Table III.4. Percentage of students predicted to achieve a C or better in college at key MCAS and PARCC performance thresholds

Test	Students at the MCAS "proficiency" cut score	Students at the PARCC "college and career ready" cut score
ELA		
Predicted percentage of students earning at least a 2.0 GPA	89.4	89.3
[number of students]	[249]	[476]
Mathematics		
Predicted percentage of students earning at least a 2.0 GPA	62.1	84.8**
[number of students]	[249]	[476]

Source: College data from the MA Department of Higher Education and MCAS and PARCC test score data.

Note: Local polynomial regression models were used to obtain predicted values at the cut score for each test.

  * PARCC percentage is statistically distinguishable from the MCAS probability at the .05 level, two-tailed test.

** PARCC percentage is statistically distinguishable from the MCAS probability at the .01 level, two-tailed test.

Differences between the MCAS and PARCC performance levels are more readily observed by examining students' average college GPAs and the percentage of students earning at least a "B" average. In Figures III.2 (English) and III.3 (math) we present students' average (standardized) English GPA and math GPA for each designated level of performance, together

---

[17] In September 2015, the PARCC consortium named these performance categories as follows: category 1 scores did not yet meet expectations, category 2 scores partially met expectations, category 3 scores approached expectations, category 4 scores met expectations, and category 5 scores exceeded expectations. Individual states may choose to add additional interpretations or labels to each of these performance categories as well.

[18] To make the results comparable for students across different college campuses, this analysis of the exams' performance categories uses the same standardized GPA values as the other correlational analyses in the study.

with bar charts showing the percentage of students achieving an average GPA of at least 3.0 in the first year of college.[19]

Students in the proficient category on the MCAS ELA assessment earned an average GPA of 2.66 in their first-year college English classes, which was not statistically distinguishable from the 2.76 GPA earned by students in the college-ready category on the PARCC ELA assessment (Figure III.2). In contrast, students who were rated proficient on the MCAS Math exam had a lower math GPA (2.39) than students scoring in the college and career ready group for PARCC in math (2.81); this difference is statistically significant and is equivalent to the difference between a grade of C-plus and B-minus (Figure III.3).

There is a similar pattern in the percentage of students achieving "B" level grades in each subject. In ELA, students in PARCC's college ready performance category were about 8 percentage points more likely to achieve a 3.0 GPA compared to students rated as proficient on the MCAS, but the difference is not statistically significant (Figure III.2). In math, the difference is larger: in the PARCC college-ready group, students are 24 percentage points more likely to achieve B grades than students rated as proficient on the MCAS math test, and the difference is statistically significant (Figure III.3).

Figure III.2. English GPA, by MCAS and PARCC performance levels
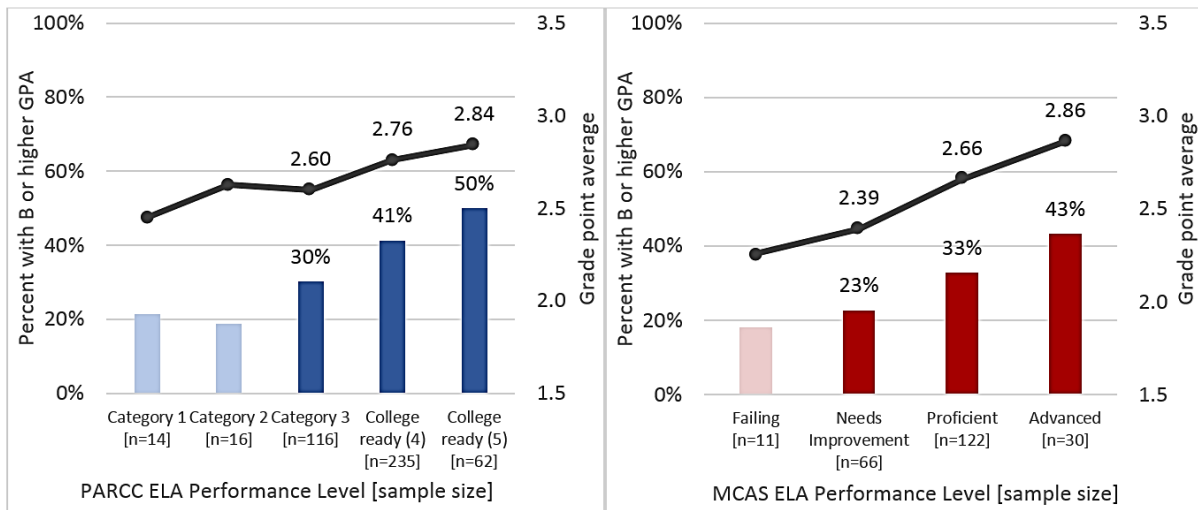
Note: Line indicates GPA and bars indicate percent with a "B" or higher. Columns are shown in a lighter shade without value labels if the sample size for a given performance category was smaller than 20 students.

---

[19] For PARCC, we defined proficiency levels for each student using the sum of raw scores on the PBA and EOY test components in each subject. We used the between-component correlations provided by Pearson to predict scores of individual students on the PARCC component they did not take.

Figure III.3. Math GPA, by MCAS and PARCC performance levels

Note: Line indicates GPA and bars indicate percent with a "B" or higher. Columns are shown in a lighter shade without value labels if the sample size for a given performance category was smaller than 20 students.

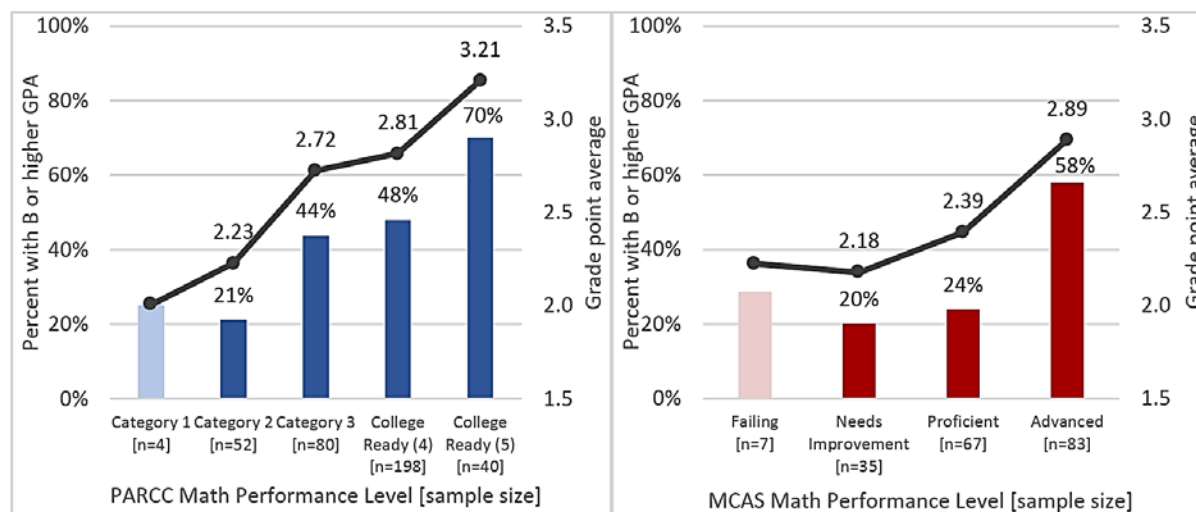Another way of comparing these performance categories is to examine the percentage of students who needed remedial coursework in their first year of college, despite meeting each test's key performance threshold (Table III.5).[20] This reveals a similar pattern to our results for college GPA: in math, the percentage of "proficient" MCAS students who enrolled in remedial courses (23.9 percent) is higher than the percentage of "college ready" PARCC students who took remedial courses (12.6 percent). This difference in remediation rates is statistically significant. For the ELA performance threshold, the remediation rate for "proficient" MCAS students (22.5 percent) is also higher than the remediation rate for "college ready" PARCC students (15.0 percent), but the difference between those two rates is only marginally statistically significant (with a p-value of 0.06).

To interpret these findings, it is helpful to remember that the definitions of the PARCC and MCAS performance categories are not directly comparable: the PARCC exam explicitly seeks to define categories (groups 4 and 5) that represent students who are prepared for college, whereas the MCAS performance levels are more narrowly targeted to measure proficiency in each high school subject relative to state curriculum standards. In addition, the differences between these exams' performance groupings do not reflect the tests' underlying ability to predict college outcomes using scores rather than performance categories. In fact, given the similar predictive validity represented by the correlations presented earlier, the differences between the two exams could be remedied by raising the threshold for proficiency on the MCAS Math exam (making it

---

[20] Table III.5 shows the percentage of "false positives," in the sense that these are students deemed proficient or college-ready even though they took first-year remedial courses. A different way of examining the data would be to consider the percentage of "false negatives," students scoring below the proficient or college-ready threshold who nevertheless did *not* take remedial courses. There is a tradeoff between reducing false positives and reducing false negatives: decreasing one type of error necessarily means increasing the other type of error. For instance, in math the MCAS proficiency standard produces a higher rate of false positives than the PARCC college-ready standard (24 percent for MCAS compared to 13 percent for PARCC), but the MCAS standard has a correspondingly *lower* rate of false negatives than PARCC (40 percent for MCAS versus 66 percent for PARCC).

more difficult to achieve a proficient MCAS rating). In other words, if performance levels were defined in a comparable way for the MCAS and PARCC, the cut-scores for both exams could be made equally predictive of college outcomes.

Table III.5. College remediation rates for students meeting MCAS and PARCC performance standards

	Students meeting MCAS "proficiency" standard	Students meeting PARCC "college and career ready" standard
ELA		
Percentage of students taking one or more remedial college courses [number of students]	22.5 [160]	15.0 [314]
Mathematics		
Percentage of students taking one or more remedial college courses [number of students]	23.9 [184]	12.6** [285]

Source: College data from the MA Department of Higher Education and MCAS and PARCC test score data.

  * PARCC remediation rate is statistically distinguishable from the MCAS rate at the .05 level, two-tailed test.

** PARCC remediation rate is statistically distinguishable from the MCAS rate at the .01 level, two-tailed test.

B. Correlations with Accuplacer scores and remedial coursework

The Accuplacer, a College Board standardized placement test for new college students, is generally used to determine whether students need remediation in reading or math to prepare for college coursework; it provides a useful alternative measure of a students' college preparedness. Accuplacer data are inconsistently available for students in the study sample. Different students took different Accuplacer subject exams (for example, College Math, Algebra, or Arithmetic), two of the study campuses do not administer the Accuplacer to new students at all, and one campus administered the Accuplacer to fewer than 30 percent of the study sample. The sample sizes for the Accuplacer analysis are therefore smaller than the full analytic sample, and they are potentially less representative than our primary analyses of college GPA.

The correlations between MCAS and PARCC scores and Accuplacer scores vary widely, but only one of 11 differences between the MCAS and PARCC correlations was statistically significant (Table III.6). Note, however, that the sample sizes in this analysis are small and vary substantially by Accuplacer exam. The small sample sizes of many of these correlations could play a role in making the estimates misleadingly high or low. For example, the sample sizes range from 26 students in the estimate for PARCC Algebra II EOY scores versus the Accuplacer Arithmetic test to 83 students for the comparison of MCAS ELA scores versus the Accuplacer Reading test.

To help address this issue, we also use the study's student-level course data to examine actual remedial course enrollment. Because we observe remedial course enrollment for all students in the data, we can analyze correlations between test scores on each exam and a variable indicating whether each student enrolled in at least one remedial course in their first year of college (controlling for average differences between college campuses in remediation rates).

Table III.6. Comparisons of correlations between MCAS/PARCC tests and Accuplacer scores

	MCAS	PARCC PBA	PARCC EOY	PARCC Algebra II
ELA				
Accuplacer Reading correlation	0.41	0.28	0.61	n.a.
[number of students]	[83]	[73]	[77]	
Mathematics				
Accuplacer College Math correlation	0.56	0.53	0.59	0.74
[number of students]	[36]	[37]	[39]	[38]
Accuplacer Algebra correlation	0.56	0.67	0.48	0.50
[number of students]	[63]	[51]	[63]	[53]
Accuplacer Arithmetic correlation	0.76	0.55	0.57	0.42*
[number of students]	[46]	[27]	[34]	[26]

Sources: College data from the Massachusetts Department of Higher Education and MCAS and PARCC test score data.

*PARCC correlation is statistically distinguishable from the MCAS correlation at the .05 level, two-tailed test.

** PARCC correlation is statistically distinguishable from the MCAS correlation at the .01 level, two-tailed test.

We begin by examining whether each individual component of PARCC and MCAS can predict which students do not enroll in remedial coursework, after controlling for differences between the colleges in our sample (Table III.7).[21] The correlations between individual components of the tests and avoiding remedial course-taking varied substantially, from 0.06 to 0.36. On average, for both exams, students who had higher MCAS or PARCC scores were less likely to have taken a remedial course in any subject. However, the MCAS ELA exam (with a correlation of 0.36) outperforms the PARCC ELA EOY exam (with a correlation of 0.06) by a statistically significant margin. In addition, the PARCC math EOY exam has a smaller correlation (0.12) than the MCAS Math exam (0.35), although that difference is only marginally statistically significant (with a p-value of 0.06). We did not observe any statistically significant differences between the validity of PARCC's components and MCAS components in predicting remediation rates in the tested subject: in math the three PARCC components all have similar correlations to MCAS (with no significant differences) in predicting math remediation. In ELA, the PARCC PBA component (0.07) and EOY component (0.22) have somewhat smaller correlations than the MCAS ELA test (0.31) in predicting ELA remediation, but these differences are not statistically significant.

---

[21] We controlled for campus differences when calculating the remedial course correlations because the colleges in the sample had different criteria for assigning students to remedial courses; each college determined its own Accuplacer cut-scores for remedial assignment and three colleges in the sample were piloting a program to use high school GPAs for placement, rather than Accuplacer scores.

Table III.7. Correlations between MCAS/PARCC test components and avoiding enrollment in a remedial course

	MCAS	PARCC PBA	PARCC EOY	PARCC Algebra II
ELA				
Remedial course enrollment correlation – any subject	0.36	0.26	0.06*	n.a.
[number of students]	[120]	[113]	[126]	
ELA remedial course enrollment correlation	0.31	0.07	0.22	n.a.
[number of students]	[120]	[113]	[126]	
Mathematics				
Remedial course enrollment correlation – any subject	0.35	0.24	0.12	0.21
[number of students]	[129]	[116]	[121]	[122]
Math remedial course enrollment correlation - tested subject	0.26	0.24	0.24	0.17
[number of students]	[129]	[116]	[121]	[122]

Sources: College data from the Massachusetts Department of Higher Education and MCAS and PARCC test score data.

*PARCC correlation is statistically distinguishable from the MCAS correlation at the .05 level, two-tailed test.

** PARCC correlation is statistically distinguishable from the MCAS correlation at the .01 level, two-tailed test.

The predictive value of PARCC's combined scores in each subject is of greater interest than the predictive value of each individual EOY and PBA test, however, since the EOY and PBA tests are intended to be used together. In addition to examining the relationship between remediation rates and scores on individual PARCC and MCAS test components, we completed a combined analysis that pools the PARCC PBA and EOY scores in each subject. These comparisons summarize the predictive validity of the entire PARCC exam in integrated math and (separately) ELA, using the same methods we used for the combined analyses of GPA (see Appendix C for additional details).

Combining PARCC's EOY and PBA components reveals that the PARCC exam and the MCAS exam provide an equivalent amount of validity in predicting which students need remedial coursework in college (Table III.8). In ELA, the correlation between MCAS scores and avoiding remedial coursework in any subject (0.36) is very similar to the combined correlation between remediation and scores on PARCC's PBA and EOY components (0.35). Likewise, in math we do not find a statistically significant difference between the MCAS correlation (0.35) and PARCC's PBA and EOY components (0.28) in predicting which students do not enroll in remedial courses (in any subject) during their first year of college. There are also no statistically significant differences between the tests with regards to predicting remediation in the tested subject.

Table III.8. Correlations with remediation rates when 2015 PARCC test components are combined

	MCAS	PARCC PBA and EOY
ELA		
Remedial course enrollment correlation – any subject	0.36	0.35
[number of students]	[120]	[239]
ELA remedial course enrollment correlation	0.31	0.20
[number of students]	[120]	[239]
Mathematics		
Remedial course enrollment correlation – any subject	0.35	0.28
[number of students]	[129]	[237]
Math remedial course enrollment correlation	0.26	0.25
[number of students]	[129]	[237]

Sources: College data from the Massachusetts Department of Higher Education and MCAS and PARCC test score data.

Note: In each subject, we compared the correlation between the MCAS exam and avoiding remediation to the correlation between the combined PARCC exams and avoiding remediation, and none of the differences were statistically significant at a 0.05 level, using a two-tailed test.

C. Correlations with SAT scores

The SAT, designed and administered by the College Board, provides another measure of college preparedness. The SAT consists of three tests: SAT Reading, SAT Math, and SAT Writing. In total, 737 of the 847 students in the analytic sample (87 percent) had scores in all three components of the exam.

We find no clear pattern of differences between the MCAS and PARCC tests with respect to their relationship with SAT scores (Table III.9). Even though the MCAS and PARCC have similar results, the magnitude of the relationships between MCAS and PARCC scores and SAT scores varies widely by subject. We find the highest correlations between the MCAS and PARCC math exams and the SAT Math. These correlations range from 0.68 to 0.80. The correlations between the PARCC and MCAS ELA tests and SAT Reading scores vary much more widely, ranging from 0.33 to 0.71. Both MCAS and PARCC tests have similarly low correlations with the writing component of the SAT exam.

When comparing the PARCC versus MCAS correlations with SAT scores, we find only two statistically significant differences among the seven pairs of correlations that we tested. One case favors the PARCC exam and the other case favors the MCAS exam. First, PARCC ELA EOY test scores are more highly correlated with SAT Reading scores than MCAS ELA test scores (0.71 versus 0.48, respectively). On the other hand, MCAS Math test scores are more highly correlated with SAT Math scores than the PARCC Math EOY component (with a correlation of 0.80 for MCAS and 0.68 for PARCC).

Table III.9. Correlations between MCAS/PARCC tests and SAT scores

	MCAS	PARCC PBA	PARCC EOY	PARCC Algebra II
ELA				
SAT Reading correlation	0.48	0.33	0.71**	n.a.
[number of students]	[102]	[96]	[114]	
SAT Writing correlation	0.21	0.14	0.29	n.a.
[number of students]	[117]	[112]	[125]	
Mathematics				
SAT Math correlation	0.80	0.70	0.68*	0.71
[number of students]	[105]	[109]	[105]	[106]

Sources: High school data from the Massachusetts Department of Elementary and Secondary Education and MCAS and PARCC test score data.

*PARCC correlation is statistically distinguishable from the MCAS correlation at the .05 level, two-tailed test.

** PARCC correlation is statistically distinguishable from the MCAS correlation at the .01 level, two-tailed test.

This page has been left blank for double-sided copying.

V. CONCLUSION

This study provides the first empirical assessment of PARCC exams' ability to predict students' success in college, and the first comparison of the predictive validity of PARCC to an existing state assessment. To be sure, the study cannot definitively answer all relevant questions about the predictive validity of PARCC assessments: it could not examine all of the various PARCC exams designed for high-school students, including assessments administered by computer rather than paper, and other math assessments that are course-specific. Moreover, the PARCC consortium has decided to make substantial changes to the exam in 2016 (including combining the PBA and EOY components into a single, shorter test) which could affect the predictive validity of PARCC scores. Despite these limitations, the report provides important, timely evidence to decision makers in Massachusetts seeking to choose an examination system to adopt for the 2015–2016 school year and beyond.

We find that scores on the PARCC and the MCAS do equally well at predicting students' success in college, as measured by first-year grades and by the probability that a student needs remediation after entering college. Scores on both tests, in both subjects, are significantly and positively correlated with students' college outcomes, and the differences between the predictive validity of the PARCC and MCAS scores are small.

When examining the predictive value of meeting each test's performance standard (defined by PARCC as college and career ready and by MCAS as proficient), the two tests produce results that are not statistically distinguishable in English language arts but that differ in mathematics. In mathematics, the PARCC exam has defined a higher performance standard for college and career readiness than the current MCAS standard for proficiency, making the PARCC performance standards better aligned with college grades and remediation needs.

Because the underlying scores on the MCAS and PARCC assessments are equally predictive, Massachusetts policymakers have more than one option to align high-school mathematics test standards with college readiness: one possibility would be to adopt the PARCC exam, but another option would be to continue using MCAS while simply setting a higher score threshold for college readiness. Either of these options would achieve the goal of ensuring that the state's high-school assessments provide better information about college readiness to students, parents, educators, and policymakers.

In many other states, the difference between existing proficiency standards and PARCC's college-ready standard is likely to be substantially larger than in Massachusetts, in English language arts as well as math. The MCAS is considered to have set higher standards for proficiency than the exam systems currently in use in most other states (Bandeira de Mello et al. 2015). States using substantially lower proficiency standards on their own assessments would see more of a divergence from PARCC results. If the current Massachusetts proficiency standards fall somewhat short of identifying students who are fully prepared to succeed at college-level math coursework, it is likely that the proficiency standards used in other state assessment systems fall far short of identifying college readiness. Assessing whether the underlying scores on assessments from other states would do better or worse than PARCC at predicting college outcomes would require studies like this one to be conducted in those states.

Even though the PARCC and MCAS examinations are similar in terms of their predictive validity, they may differ on a variety of other dimensions that are relevant to the state's choice. For example, the content knowledge and problem-solving skills measured across the tests' various components are not identical, and it is possible that these exams could differ in the extent to which they align with specific high school curricular reform goals or teaching standards. These two examinations may also differ in terms of the amount of time needed for administration, the expense of scoring and scaling, and the effort needed to understand and use data from multiple different tests. Finally, there may be other benefits (and drawbacks) to consider related to adopting the PARCC assessment because it is simultaneously being adopted by a consortium of other states. A full examination of these costs and benefits falls outside the scope of this study, but these other factors merit consideration in the state's decision.

REFERENCES

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

Benjamini, Yoav, and Yosef Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society*, vol. 57, no. 1, 1995, pp 289-300.

Carmichael, Sheila Byrd, Gabrielle Martino, Kathleen Porter-Magee, and W. Stephen Wilson. "The State of State Standards—and the Common Core—in 2010." Washington, DC: Thomas B. Fordham Institute, July 2010.

Chang, Kenneth. "Expecting the Best Yields Results in Massachusetts." *New York Times*, September 2, 1013.

Cohen, Jacob. *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1988.

Massachusetts Business Alliance for Education and the Center for Assessment. "Educating Students for Success." February 2015.

Massachusetts Department of Elementary and Secondary Education. "District Analysis, Review & Assistance Tools: Success After High School." Available at http://www.doe.mass.edu/apa/dart/. Accessed July 1, 2015.

Mattern, Krista D., Brian F. Patterson, and Jennifer L. Kobrin. "The Validity of SAT Scores in Predicting First-year Mathematics and English Grades." New York: The College Board, 2012.

National Center for Education Statistics. "NAEP State Profiles." Available at http://nces.ed.gov/nationsreportcard/states/. Accessed July 1, 2015.

This page has been left blank for double-sided copying.

APPENDIX A

DETAILED DESCRIPTION OF THE MCAS AND PARCC SAMPLE

This page has been left blank for double-sided copying.

The Massachusetts Department of Elementary and Secondary Education prepared the MCAS and PARCC data file and made the data available to the study team. The original MCAS and PARCC data file contained records for 866 students' 2015 MCAS or PARCC raw test scores (including scaled scores for the 2015 MCAS), 10th grade MCAS scaled scores, SAT scores, and a unique student identification number generated for the study. The data file was fully de-identified, meaning the data did not include student names, dates of birth, demographic information, or any government identification numbers. Our comparative analyses used the MCAS and PARCC raw test scores, since scaled scores were not provided for the PARCC tests in time for this study.

The final analytic sample contains 847 students, after eliminating students who had incomplete tests, took more than one exam, or performed below the guessing floor threshold. In this appendix we present information about the data preparation process and the study sample. First, we describe the preparation of the data, focusing on the identification and elimination of outliers. Next, we explain the baseline equivalence analysis. Finally, we present descriptive statistics of the final analytic sample used in the study.

A. Cases dropped from the 2015 MCAS and PARCC Scores

To prepare the 2015 MCAS and PARCC data file, it was necessary to identify and eliminate student outliers who did not complete the exam or did not make a good-faith effort to answer the questions correctly. First, we eliminated students who did not complete their assigned MCAS or PARCC exam, as well as students who took more than one exam. After eliminating these cases, we performed a series of analyses to identify additional outliers. We examined the distribution of raw MCAS and PARCC scores in histograms, assessed students' performance on the 2015 exams relative to their prior test scores, and examined cases that fell below the guessing thresholds for each test. Cases were eliminated only if there was a preponderance of evidence to suggest that they were problematic. In other words, we did not eliminate cases that appeared as outliers in only one analysis; only the cases that were flagged in more than one analysis were dropped.

These analyses revealed that patterns of low-effort and guessing were relatively rare in this sample. In all, only 19 students were removed from the analytic sample for the reasons outlined above, dropping the original sample of 866 students to a final sample size of 847 students. Of these 19 students, eight were removed because they did not complete their assigned exam; seven were removed because they took two exams; and four were removed because they appeared as outliers in multiple analyses. We interpret this to mean that the study's incentive scheme to encourage effort on the tests was successful for over 97 percent of the participating students in the sample. In Table A.1 we present a summary of dropped cases by test form.

Non-completers and multiple responders

In the study sample, eight students did not complete their assigned exam, five MCAS and three PARCC. These students were flagged by the MCAS and PARCC test-scoring entities and did not receive final scores. In the case of the PARCC exams, we received only the student ID, and in one case, the PARCC test taken; all other fields were missing. In the case of the MCAS exams, we received the raw scores and baseline student data, but we were not provided the scaled scores for students who did not complete their assigned exam. For consistency across the

two sets of exams, we eliminated all eight observations with exams that were known to be incomplete. These eight observations were excluded from all subsequent analyses, including the outlier analysis and the baseline equivalence analysis.

We also examined the scores of students who took more than one exam; there were seven students in the data who chose to participate in the study twice. There were no data on the dates that the students took the PARCC and MCAS exams, so we could not determine which test was taken first, which is the one we would have preferred to retain. This is problematic because these students could have learned about the tests and test-taking process between each instance of taking exam (an advantage that wasn't available for the vast majority of the students in our sample). In addition, we also found evidence that some of these students did not put forth a meaningful amount of effort on at least one of their exams. For instance, one of these students received a raw score of 60 on one occasion and a raw score of 3 on a second occasion taking the same exam. We did not use the higher score because we could not say for sure that the student did not score a 3 the first time and then, use knowledge of the questions to achieve a higher score the second time to earn the incentive.

Outliers in the raw score distributions

After eliminating the 15 students who did not complete their assigned tests or took more than one test, we examined the distribution of raw MCAS and PARCC scores in the remaining cases. We created histograms of raw scores for each of the seven test forms and looked for evidence of bunching, or a spike in scores, among the lowest possible scores in the distribution. This could indicate that a group of students did not complete the test or guessed on the test. There is some evidence of bunching in the histograms of the raw scores of the PARCC ELA EOY and the PARCC Algebra II EOY exams.[22] This could be an indication that some students did not complete the test or randomly completed multiple choice questions. However, it could also be a sign that the exam itself was difficult and produced "floor effects" at the lower end of the distribution. We flagged these for further investigation. The PARCC Math assessments had negatively skewed distributions, which is likely a sign of overall exam difficulty rather than low effort of some students. The remaining histograms appeared to have fairly normal distributions.

Scores out of line with high school exam scores

We also systematically assessed all of the remaining students' performance on the 2015 exams relative to their prior test scores to look for evidence of low effort. The goal of this analysis was to identify students who performed much worse relative to their peers on the 2015 test, in comparison to their performance on the grade 10 MCAS. For example, if a student scored one standard deviation above the mean on the grade 10 MCAS ELA exam and two standard deviations below the mean on the 2015 MCAS ELA exam or the PARCC EOY ELA exam, we might suspect that he/she did not put forth a meaningful effort on the 2015 exam. (Alternatively, differences in the other direction could mean that a student performed uncharacteristically poorly on the 10th grade assessment, learned a lot more than his/her peers in the three intervening years, or cheated on the 2015 exam, for instance.)

---

[22] Detailed histograms showing these floor effects are available from authors by request.

To do this, we standardized students' 2015 PARCC/MCAS exams into z-scores (defined relative to other test-takers in the study sample), and performed the same standardization procedure for students' 10th grade MCAS scores. We then computed the difference between the 2015 exam z-score and the relevant grade 10 MCAS z-score. In Table A.2 we present the percentage of the sample with z-score differences at select thresholds. If many students in the sample did not try on the 2015 exams, we would expect to see a distribution with larger percentages of students with negative z-score differences. Instead, we observe that the distribution is fairly balanced between high and low z-score differences.

In addition, we used the z-scores to create scatter plots of the 2015 exams against the relevant grade 10 MCAS exams, as well as the 2015 exams against the relevant SAT tests. In Figures A.1 and A.2, we present the scatter plots for PARCC ELA PBA, PARCC ELA EOY, and the PARCC Algebra II EOY tests versus the MCAS and SAT tests. In these plots, the students with z-score differences that are below –2 (those whose high school z-scores declined by more than 2 standard deviations) are shown in red. We used the red points to look for students whose scores were consistent outliers relative to both SAT scores and 10th grade MCAS scores. Nine of the 13 students with a z-score difference of less than a –2 were in either the PARCC ELA PBA or EOY tests (shown in Figure A.1). However, given the large variability of both positive and negative z-scores differences in the data, we concluded that there was not enough evidence to exclude these students from the analysis as outliers.

In contrast, the PARCC Algebra II EOY exam does appear to have clear outliers both when compared to the Grade 10 MCAS math z-scores and high school SAT Math z-scores (Figure A.2). This is consistent with the bunching that we observed in the lower end of the distribution in the histogram of raw PARCC Algebra II EOY scores. This analysis flagged four students from the PARCC Algebra II EOY test for further analysis; we eventually dropped these cases from the analytic sample.

Scores below the guessing floor

For each exam, we were provided with information about the score that the average student would receive if he/she randomly guessed on all of the exam's multiple-choice questions. The guessing floor was calculated based on the total number of questions in which guessing was possible, divided by the number of options. For example, if there were 20 multiple choice questions with four options each, the guessing floor would be "5" because a student who randomly selected answers would be expected, on average, to receive a score of "5."

For each exam, we examined the provided 'guessing floor' and also calculated a conservative guessing floor equal to 80 percent of the guessing floor provided. (For instance, a guessing floor of 5 would conservatively be a 4.) Next, we counted the number of students whose scores fell at or below both the standard and conservative guessing floors. In the case of the PARCC tests, four students' scores fell below the standard guessing floor and three students' scores fell below the conservative guessing standard in the PARCC Algebra II EOY exam. The four students coincide with the outliers observed in the PARCC Algebra II EOY histogram and scatter plots. Therefore, we eliminated all four of these students from the analysis. In the case of the MCAS tests, no students fell below the conservative or the standard guessing threshold for either the MCAS ELA test or the MCAS Math test. Therefore, we maintained all of the remaining students who took the MCAS in the analytic sample.

Table A.1. Summary of cases dropped from analytic sample, by test form

Test	Number dropped who had incomplete exams	Number dropped who took two exams	Number dropped who had scores below guessing threshold	Total dropped from analytic sample	Total original sample size by category	Percentage of original sample dropped
MCAS						
ELA	3	0	0	3	123	2.4
Math	2	0	0	2	131	1.5
PARCC						
ELA (PBA)	0	2	0	2	115	1.7
ELA (EOY)	0	2	0	2	128	1.6
Math (PBA)	0	0	0	0	116	0.0
Math (EOY)	0	1	0	1	122	0.8
Algebra II (EOY)	1	2	4	7	128	5.5
Missing[a]	3	0	0	2	2	100.0
Total	**8**	**7**	**4**	**19**	**866**	**2.2**

Source: High school data from the MA Department of Elementary and Secondary Education, college data from the MA Department of Higher Education, and MCAS and PARCC data.

[a] In the case of two students with incomplete PARCC exams, the test administrators did not provide scores for these students or the name of the specific PARCC test form administered.

Table A.2. Percentage of sample[a] with z-score differences less than or greater than specified threshold

Threshold	MCAS ELA	MCAS Math	PARCC ELA PBA	PARCC Math PBA	PARCC ELA EOY	PARCC Math EOY	PARCC Algebra II EOY	Total
Less than -2	0.9	0.0	4.7	0.0	3.3	0.0	2.5	1.6
Less than -1.5	1.8	0.8	11.2	3.5	7.4	2.6	5.8	4.7
Less than -1	15.0	6.5	13.1	14.9	15.6	10.4	16.7	13.1
Greater than 1	14.2	11.4	19.6	13.2	9.8	13.9	15.0	13.8
Greater than 1.5	8.8	6.5	10.3	7.0	3.3	7.0	5.0	6.8
Greater than 2	4.4	3.3	7.5	3.5	1.6	3.5	0.8	3.4
N Total	**113**	**123**	**107**	**114**	**122**	**115**	**120**	**814**

Source: Authors' calculations, based on study-administered MCAS and PARCC scores and 10th grade baseline MCAS scores.

Note: Z-score differences were calculated based on the z-score of the exam taken minus the z-score of the students' relevant grade 10 MCAS exam. The threshold categories are cumulative (e.g. the "Less than -1" category represents the percentage of *all* students whose 2015 test scores were more than one standard deviation below their 10th grade MCAS scores, which includes the students in the "Less than -1.5" and the "Less than -2" categories).

[a] The sample consists of 814 students, which represents all students with baseline 10th grade MCAS scores, after removing 8 students who had incomplete exams and 7 students who took more than one exam. The denominator for each test is the total number of students who took that exam and had 10th grade MCAS scores.

Figure A.1. Relationships between PARCC scores and high school exam scores

Figure A.2. Relationships between PARCC Algebra II scores and high school exam scores
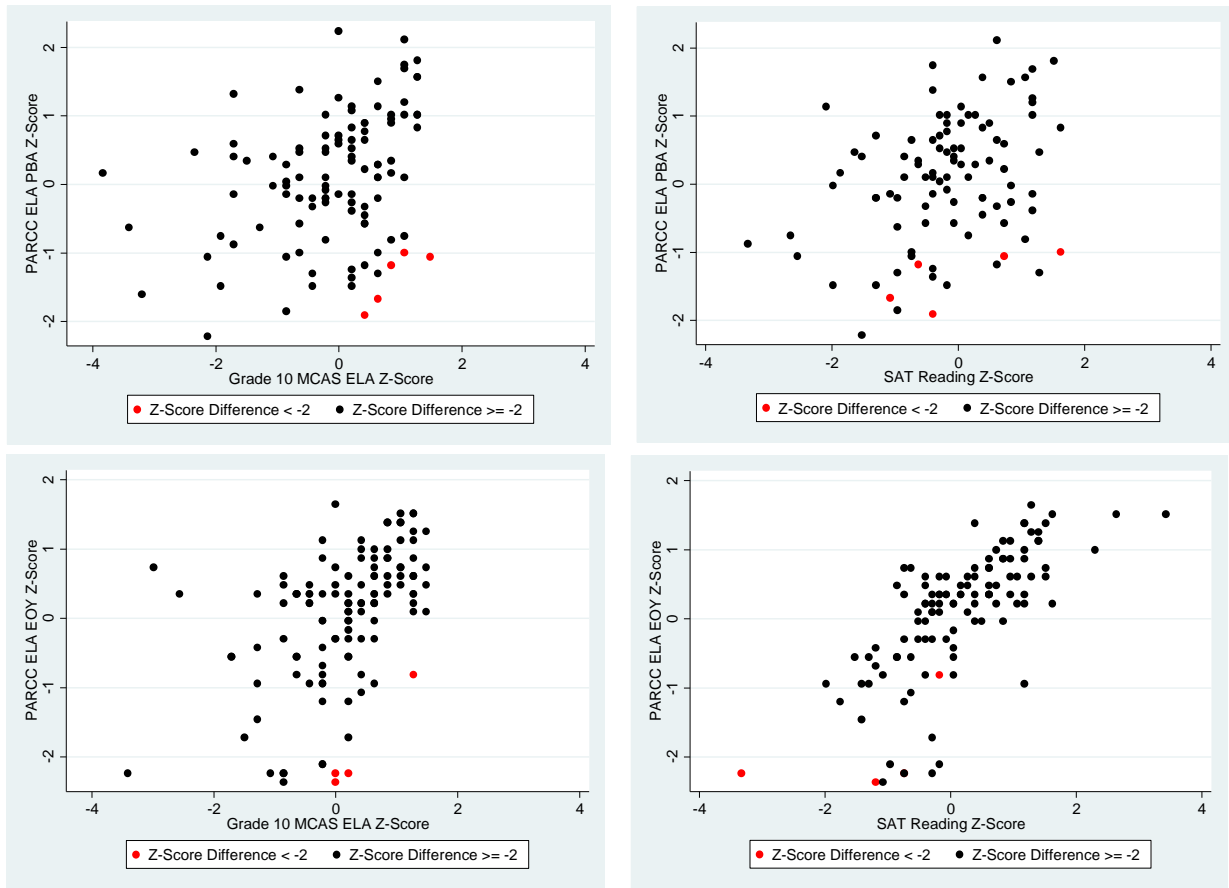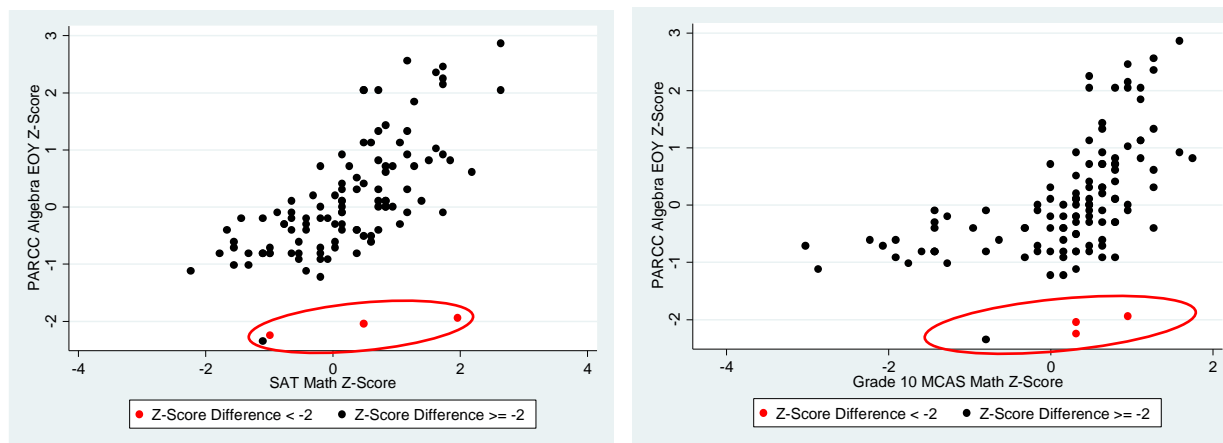
Source: High school data from the MA Department of Elementary and Secondary Education and 2015 MCAS and PARCC data.

B.   Baseline equivalence

We compared the baseline (pre-test) characteristics of students taking the different assessments to ensure that the random assignment process succeeded in creating equivalent groups of students. If the random assignment procedure worked correctly, the characteristics of students should be the same across groups, on average. If the groups are equivalent, we can be confident that any differences in outcomes across groups are attributable to the test forms and not to background differences among the students.

We assessed the baseline equivalence of the study sample in two ways: by comparing all seven test groups and by comparing the MCAS versus the PARCC test groups. First, we tested for the equivalence of student characteristics across each of the seven study groups (two MCAS test groups and five PARCC test groups), by examining indicators including students' 10th grade MCAS test scores, SAT scores, first-year unadjusted college GPAs (both total GPA and subject-specific GPAs), and enrollment in at least one remedial course during the first year of college (Table A.3). Second, we compared the characteristics of the group of students who took a component of the MCAS exam to the group students who took a component of the PARCC exam. We performed these analyses both with and without the PARCC Algebra II EOY (Tables A.4 and A.5, respectively), since the Algebra II test scores were not part of the study's multi-component analyses that compared multiple PARCC test-forms to the MCAS.. The final analytic sample of 847 students was used to conduct the equivalence analyses (although some of these students were missing baseline data for one or more indicators).

We tested the equivalence of ten high school and college characteristics across all seven groups[23] and found no statistically significant differences in 10th grade MCAS scores, SAT writing scores, college GPAs, or enrollment in remedial courses (see Table A.3). Across the study groups, we found some differences in SAT Reading scores, which ranged from an average of 475 to 512 per group, and in SAT Math scores, which ranged from an average of 501 to 537 per group. These differences were approximately one-third of a standard deviation of SAT scores in the study sample. After accounting for the number of different tests in the equivalence analysis, these differences were not statistically significant.

To account for multiple comparison groups and multiple outcomes, which increases the likelihood that a chance difference is incorrectly labeled as statistically significant, we applied the Benjamini-Hochberg (B-H) correction method (Benjamini & Hochberg, 1995). This method reduces the probability of false positives, or incorrectly concluding that there are statistically significant differences when variation simply occurred by chance. With this adjustment, none of the differences in student characteristics were statistically significant.

When comparing the MCAS test takers to PARCC test takers,[24] we found no statistically significant differences in any high school indicators, but we did find notable differences in college performance indicators; the differences in total GPA, ELA GPA, and math GPA were all statistically significant (see Table A.5). On average, students in the PARCC group had slightly higher GPAs than the students in the MCAS group. The PARCC group's average total GPA was 0.15 points higher than the MCAS group's average total GPA, and the differences in the subject-specific GPAs ranged from 0.19 to 0.21, on average. By comparison, the difference between a B and a B+ represents 0.30 GPA points. When the PARCC Algebra II EOY exam was excluded from the analysis, only the differences in ELA GPA and math GPA were statistically significant (see Table A.5). However, as with our analysis of equivalence across the seven testing groups, after adjusting for the number of different indicators we examined none of the differences in either analysis (including or excluding the PARCC Algebra II exam) remain statistically significant.

---

[23] We regressed each indicator on dummy variables for six of the seven "treatment" groups (omitting one exam category). Then, we conducted a post-regression F-test to test whether the coefficients characterizing the exams' relationships with each of the outcomes were jointly statistically significantly different from zero.

[24] We regressed each indicator on a dummy variable for PARCC versus non-PARCC, in other words, PARCC versus MCAS. Since there were only two groups, we used the T-results from the regression coefficient of the "treatment" dummy to determine whether the difference between the two groups were statistically significantly different than 0.

Table A.3. Baseline test scores and college performance of students in the seven testing groups

	MCAS ELA	MCAS Math	PARCC EOY Algebra	PARCC EOY ELA	PARCC EOY Math	PARCC PBA ELA	PARCC PBA Math	All students	P-value of F-test	Significant after adjusting for multiple comparisons?
High School MCAS ELA Score										
Mean	256	255	257	257	255	255	257	256	0.26	No
SD	9	11	9	9	9	10	9	9		
N	113	123	116	122	115	107	114	810		
High School MCAS Math Score										
Mean	257	256	260	259	258	257	259	258	0.46	No
SD	13	14	12	12	12	13	12	13		
N	113	123	116	122	115	107	114	810		
SAT Reading Score										
Mean	495	495	512	507	475	479	507	496	0.03*	No
SD	74	94	94	89	86	88	88	88		
N	102	105	106	114	105	96	109	737		
SAT Math Score										
Mean	505	510	537	533	513	501	517	517	0.04*	No
SD	81	92	87	84	81	95	92	88		
N	102	105	106	114	105	96	109	737		
SAT Writing Score										
Mean	475	477	497	490	468	470	486	481	0.25	No
SD	87	94	85	88	90	87	87	89		
N	102	105	106	114	105	96	109	737		
Total GPA (Original)										
Mean	2.90	2.92	3.12	3.07	3.00	3.00	3.08	3.01	0.25	No
SD	0.81	0.83	0.63	0.74	0.68	0.80	0.64	0.74		
N	120	129	122	126	121	113	116	847		
ELA GPA (Original)										
Mean	3.11	2.92	3.25	3.24	3.21	3.10	3.20	3.15	0.15	No
SD	0.97	0.98	0.70	0.77	0.74	0.98	0.80	0.86		
N	109	120	119	116	109	110	108	791		
Math GPA (Original)										
Mean	2.59	2.47	2.77	2.74	2.67	2.83	2.69	2.68	0.15	No
SD	1.23	1.19	1.16	1.17	1.17	1.07	1.08	1.16		
N	92	100	101	101	98	87	88	667		

	MCAS ELA	MCAS Math	PARCC EOY Algebra	PARCC EOY ELA	PARCC EOY Math	PARCC PBA ELA	PARCC PBA Math	All students	P-value of F-test	Significant after adjusting for multiple comparisons?
Other Subject GPA (Original)										
Mean	2.94	2.99	3.19	3.10	3.01	3.00	3.12	3.05	0.12	No
SD	0.78	0.87	0.55	0.73	0.68	0.82	0.64	0.73		
N	117	127	121	125	119	112	115	836		
Enrollment in Remedial Course										
Proportion	0.37	0.30	0.20	0.17	0.28	0.24	0.17	0.25	0.17	No
SD	0.48	0.46	0.41	0.37	0.45	0.43	0.38	0.43		
N	120	129	122	126	121	113	116	847		

Source: High school data from the MA Department of Elementary and Secondary Education, college data from the MA Department of Higher Education, and MCAS and PARCC data.

Note: The sample of each test form excludes test scores of students who had incomplete tests, took more than one test, and performed below the guessing floor threshold. The total sample is 847 students, but baseline data were not available for all students for each indicator.

*Significantly different from zero at the .05 level, two-tailed test.

**Significantly different from zero at the .01 level, two-tailed test.

Table A.4. Baseline test scores and college performance of students who took the MCAS versus PARCC tests

High school baseline scores					College performance				
	MCAS	PARCC	P-value of T-test	Significant after adjusting for multiple comparisons?		MCAS	PARCC	P-value of T-test	Significant after adjusting for multiple comparisons?
High School MCAS ELA Score					**Total GPA (Original)**				
Mean	256	256	0.65	No	Mean	2.91	3.06	0.03*	No
SD	10	9			SD	0.82	0.70		
N	236	574			N	249	598		
High School MCAS Math Score					**ELA GPA (Original)**				
Mean	257	259	0.23	No	Mean	3.01	3.20	0.02*	No
SD	14	12			SD	0.98	0.80		
N	236	574			N	229	562		
SAT Reading Score					**Math GPA (Original)**				
Mean	495	497	0.88	No	Mean	2.53	2.74	0.01**	No
SD	84	90			SD	1.21	1.13		
N	207	530			N	192	475		
SAT Math Score					**Other Subject GPA (Original)**				
Mean	508	521	0.27	No	Mean	2.96	3.09	0.08	No
SD	87	88			SD	0.83	0.69		
N	207	530			N	244	592		
SAT Writing Score					**Enrollment in Remedial Course**				
Mean	476	483	0.57	No	Proportion	0.33	0.21	0.09	No
SD	91	88			SD	0.47	0.41		
N	207	530			N	249	598		

Source: High school data from the MA Department of Elementary and Secondary Education, college data from the MA Department of Higher Education, and MCAS and PARCC test score data.

Note: The sample of each test form excludes test scores of students who had incomplete tests, took more than one test, and performed below the guessing floor threshold. The total sample is 847 students, but baseline data were not available for all students for each indicator.

*Significantly different from zero at the .05 level, two-tailed test.

**Significantly different from zero at the .01 level, two-tailed test.

Table A.5. Baseline test scores and college performance of students who took the MCAS versus PARCC tests, excluding PARCC Algebra II EOY

High school baseline scores					College performance				
	MCAS	PARCC	P-value of T-test	Significant after adjusting for multiple comparisons?		MCAS	PARCC	P-value of T-test	Significant after adjusting for multiple comparisons?
High School MCAS ELA Score					**Total GPA (Original)**				
Mean	256	256	0.72	No	Mean	2.91	3.04	0.06	No
SD	10	9			SD	0.82	0.72		
N	236	458			N	249	476		
High School MCAS Math Score					**ELA GPA (Original)**				
Mean	257	258	0.36	No	Mean	3.01	3.19	0.04*	No
SD	14	12			SD	0.98	0.83		
N	236	458			N	229	443		
SAT Reading Score					**Math GPA (Original)**				
Mean	495	493	0.69	No	Mean	2.53	2.73	0.02*	No
SD	84	89			SD	1.21	1.12		
N	207	424			N	192	374		
SAT Math Score					**Other Subject GPA (Original)**				
Mean	508	517	0.42	No	Mean	2.96	3.06	0.17	No
SD	87	88			SD	0.83	0.72		
N	207	424			N	244	471		
SAT Writing Score					**Enrollment in Remedial Course**				
Mean	476	479	0.77	No	Proportion	0.33	0.21	0.08	No
SD	91	88			SD	0.47	0.41		
N	207	424			N	249	476		

Source: High school data from the MA Department of Elementary and Secondary Education, college data from the MA Department of Higher Education, and MCAS and PARCC test score data.

Note: The sample of each test form excludes test scores of students who had incomplete tests, took more than one test, and performed below the guessing floor threshold. The total sample is 847 students, but baseline data were not available for all students for each indicator.

*Significantly different from zero at the .05 level, two-tailed test.

**Significantly different from zero at the .01 level, two-tailed test.

C. Descriptive statistics for the MCAS and PARCC analytic sample

In Table A.6 we present the number of students who took each MCAS or PARCC test form, as well as the mean raw score, the standard deviation, and the range of scores for each test. These descriptive statistics are drawn from the final analytic sample of 847 students. Each test group has roughly the same number of students; the sample size for test form ranges from 113 to 129. The scores presented in this table are raw scores; because each test form has a different number of questions, the mean, standard deviations, and ranges differ from one test form to the next.

In Figures A.3 to A.6 we present histograms of the distributions of the 2015 MCAS and PARCC raw scores.

Table A.6. Summary statistics of 2015 MCAS and PARCC raw test score results

Test	N	Mean	Standard deviation	Minimum	Maximum
MCAS ELA	120	44.60	10.46	18	64
MCAS Math	129	35.83	12.56	9	58
PARCC ELA PBA	113	43.87	15.97	7	80
PARCC Math PBA	116	12.40	7.95	2	33
PARCC ELA EOY	126	29.35	7.67	11	42
PARCC Math EOY	121	21.83	8.82	7	51
PARCC Algebra EOY	122	23.84	9.09	9	51

Source: Scores from study-administered MCAS and PARCC tests.

Note: The sample of each test form excludes test scores of students who had incomplete tests, took more than one test, and performed below the guessing floor threshold.

Figure A.3. Histogram of 2015 MCAS ELA and Math raw score distributions
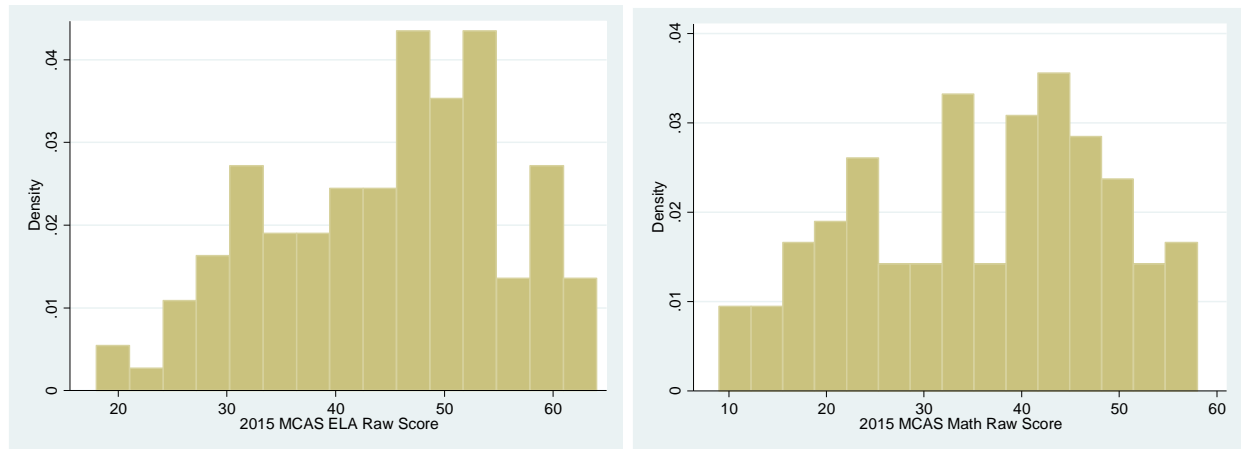
Source: 2015 MCAS ELA raw scores from sample of 120 students. 2015 MCAS Math raw scores from sample of 129 students.

Note: The samples represented in all histogram distributions in this section exclude test scores of students who had incomplete tests, took more than one test, and performed below the guessing floor threshold.

Figure A.4. Histogram of PARCC ELA and Math PBA raw score distributions
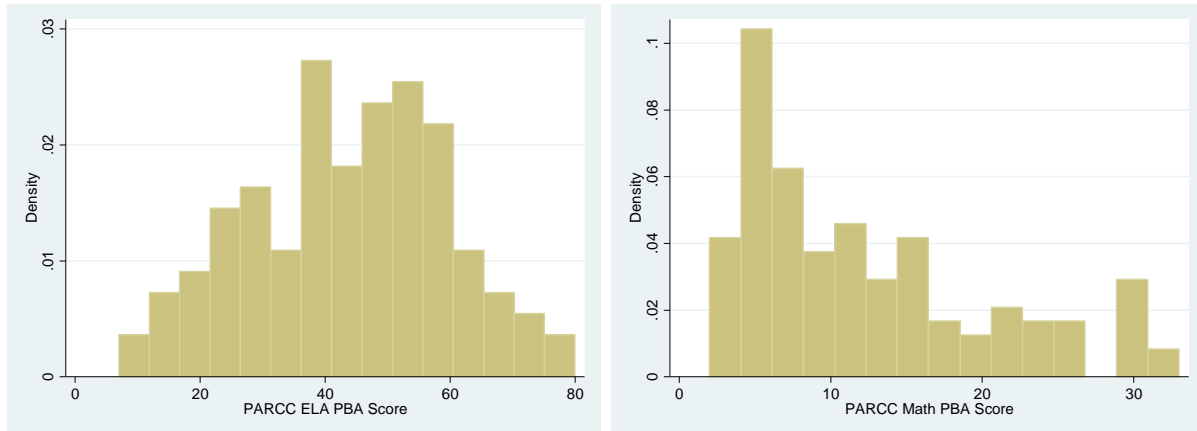

Source: 2015 PARCC ELA PBA raw scores from sample of 113 students. 2015 PARCC Math PBA raw scores from sample of 116 students.

Figure A.5. Histogram of PARCC ELA and Math EOY raw score distributions
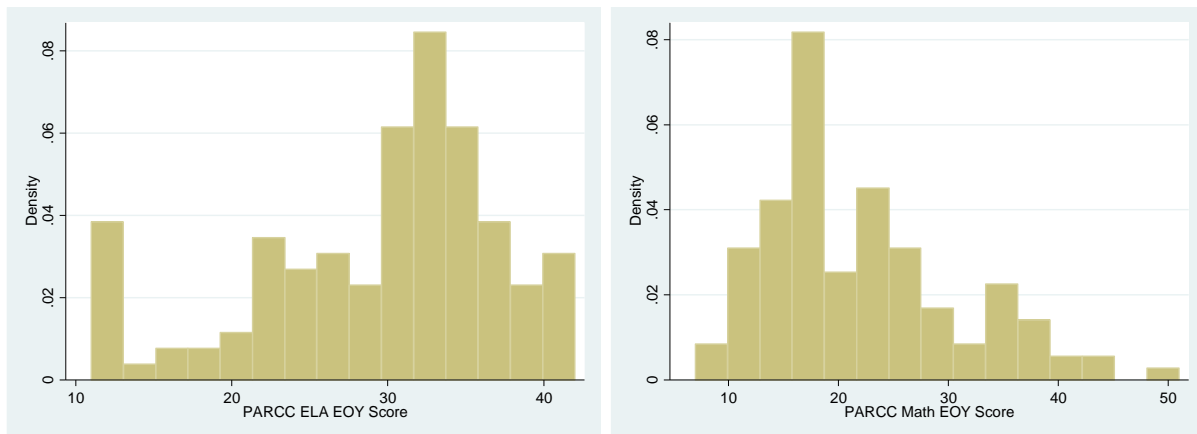
Source: 2015 PARCC ELA EOY raw scores from sample of 126 students. 2015 PARCC Math EOY raw scores from a sample of 121 students.

Figure A.6. Histogram of PARCC Algebra II EOY raw score distributions
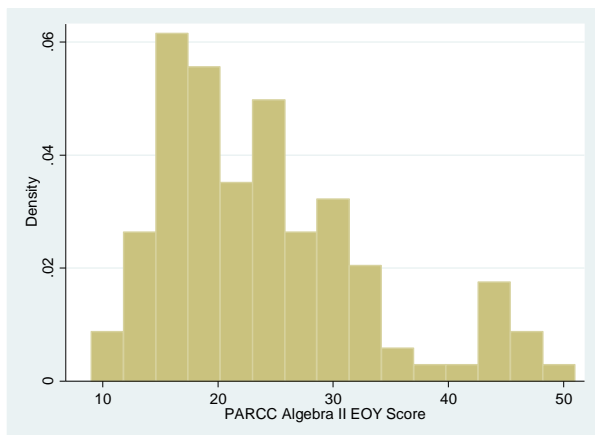
Source: 2015 PARCC Algebra II EOY scores from a sample of 122 students

APPENDIX B

COLLEGE COURSE GRADE DATA

This page has been left blank for double-sided copying.

Participating colleges and universities submitted data on college courses and grades for students who participated in the study. Because study eligibility was limited to students who were first-time freshmen, all course data were from students' first year in college. Course data included the course ID, section ID, term, course title, available credit hours, a remedial course indicator (remedial math, remedial reading, and remedial writing), and the Classification of Instructional Programs (CIP) code, a standardized code used to categorize courses by subject area and topic. Course outcome data included the outcome status (completed, withdrew, failed, incomplete, and in progress) and the course grade, either a numerical grade on a zero to four scale or a pass/fail indicator.

A. Data preparation

We assigned subject codes to all the courses in our sample, categorizing courses as English language arts, math, or 'other subjects' using the two-digit CIP code, which groups courses by their general subject area. We categorized some remedial courses using the subject identified in the remedial course indicator.[25] In our data, six courses did not have CIP codes or remedial indicators. We categorized these courses based on their course title and course descriptions from online course catalogs. Next, we identified all remedial courses, coding as remedial all courses with a remedial course indicator and an additional three courses that did not have a remedial indicator but were categorized as remedial based on their two-digit CIP code.

The course grade variable included either numerical grades, on a zero to four scale, or pass/fail grades reported as 'P' or 'F.' Additionally, some course entries gave a grade 'zero' to students who withdrew or did not complete the course, while other entries provided no grade value in cases of withdrawals. To create a consistent numerical grade variable across all courses, we coded grades as missing for all course showing a status of 'withdrew,' 'incomplete,' or 'in progress.' In addition, we coded all "P" grades as missing because we did not have a consistent numerical grade to associate with these entries across subjects and campuses. Finally, we coded grades to have a value of zero for all courses with an outcome of 'failed' or 'F.'

B. Summary of unadjusted course grades

Table B.1 summarizes the unadjusted course grades for students included in our analyses. The average course grade was 3.08, but varied by course level, subject, and institution. The average grade in English language arts courses, for example, was 3.18, while the average grade in math courses was 2.68. By campus, average grades ranged from a low of 2.77 at Quinsigamond Community College to a high of 3.42 at the University of Massachusetts, Lowell.

---

[25] For courses coded with the two-digit CIP code of "32," indicating a remedial course, we categorized them based on the full six-digit CIP code, which provides a more detailed subject grouping. When the six-digit CIP code did not indicate a specific subject, we used the subject identified in the remedial course indicator.

## Table B.1. Summary of unadjusted course grades

	Number of courses	Average grade	Standard deviation
All courses	7,698	3.08	0.99
By remedial status			
Remedial	254	2.44	1.42
Non-remedial	7,444	3.10	0.96
By course subject			
English language arts	1,529	3.18	0.91
Math	1,014	2.68	1.23
Other subjects	5,155	3.13	0.94
By institution			
Berkshire Community College	209	2.95	1.26
Bristol Community College	255	3.17	1.00
Bunker Hill Community College	282	3.14	1.10
Massasoit Community College	355	2.98	1.20
Middlesex Community College	347	2.83	1.11
Quinsigamond Community College	290	2.77	1.20
Bridgewater State University	1,930	3.11	0.94
Massachusetts Maritime Academy	1,162	2.84	0.88
Salem State University	1,081	3.12	0.99
University of Massachusetts-Boston	716	3.10	1.03
University of Massachusetts-Lowell	1,071	3.42	0.73

Source: College data from the MA Department of Higher Education.

Note: Excludes grades of students who had incomplete tests, took more than one test, or performed below the guessing floor threshold on a test.

Figure B.1 shows the overall distribution of grades. The unadjusted course grade distribution skews to the right, with the grade range of 3.6 to 4.0 containing the highest density of courses.

Figure B.1. Distribution of unadjusted course grades across all courses
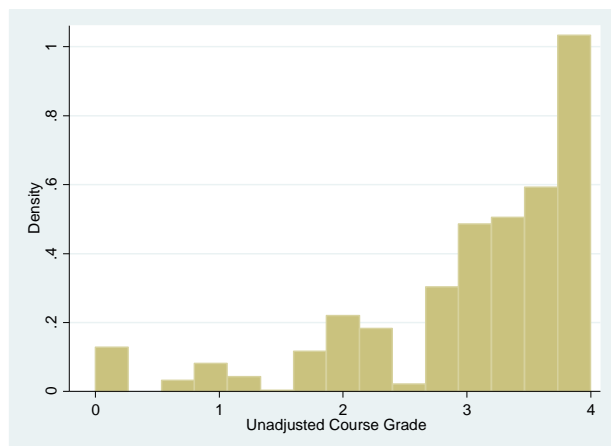
Figures B.2, B.3, and B.4 compare the distribution of grades by remedial status, subject, and institution. As these figures show, remedial courses have a much larger concentration of failing grades compared with non-remedial courses, and math courses have a larger concentration of

grades below 3.0 compared with English language arts courses and courses in other subjects.[26] The distribution of course grades also differs by institution. Some institutions, like Massasoit and Quinsigamond Community Colleges, have a larger concentration of course grades below 2.0, while other institutions, like the University of Massachusetts, Lowell, have a higher concentration of grades above 2.0.

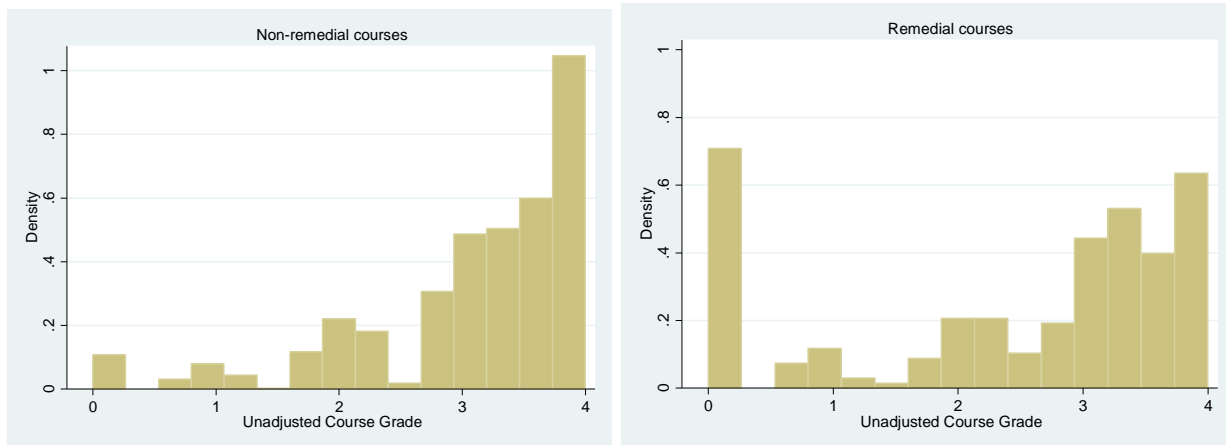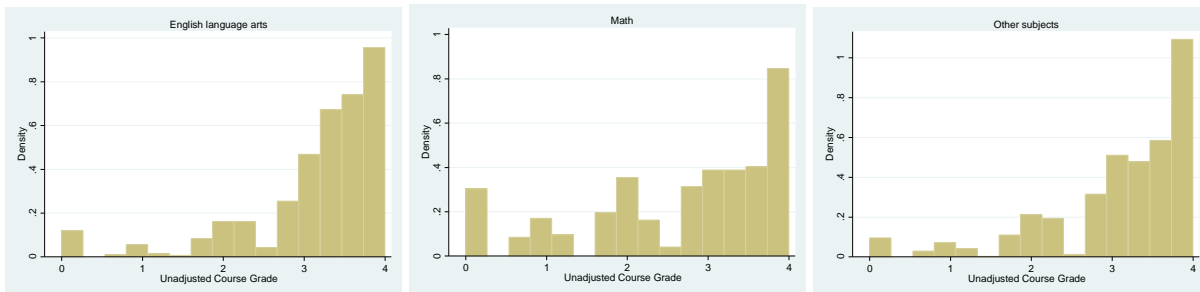## Figure B.2. Distribution of unadjusted course grades by remedial status

Figure B.3. Distribution of unadjusted course grades by subject

---

[26] As we describe below, when remedial grades in the same subject are examined within-student, that is comparing the grades a student receives in remedial courses with the grades that same student receives in non-remedial courses, remedial grades are *higher* on average. This indicates that remedial courses are less difficult than non-remedial courses, and that grades in remedial courses do not represent the grades students would receive in a college-level course.

Figure B.4. Distribution of unadjusted course grades by institution
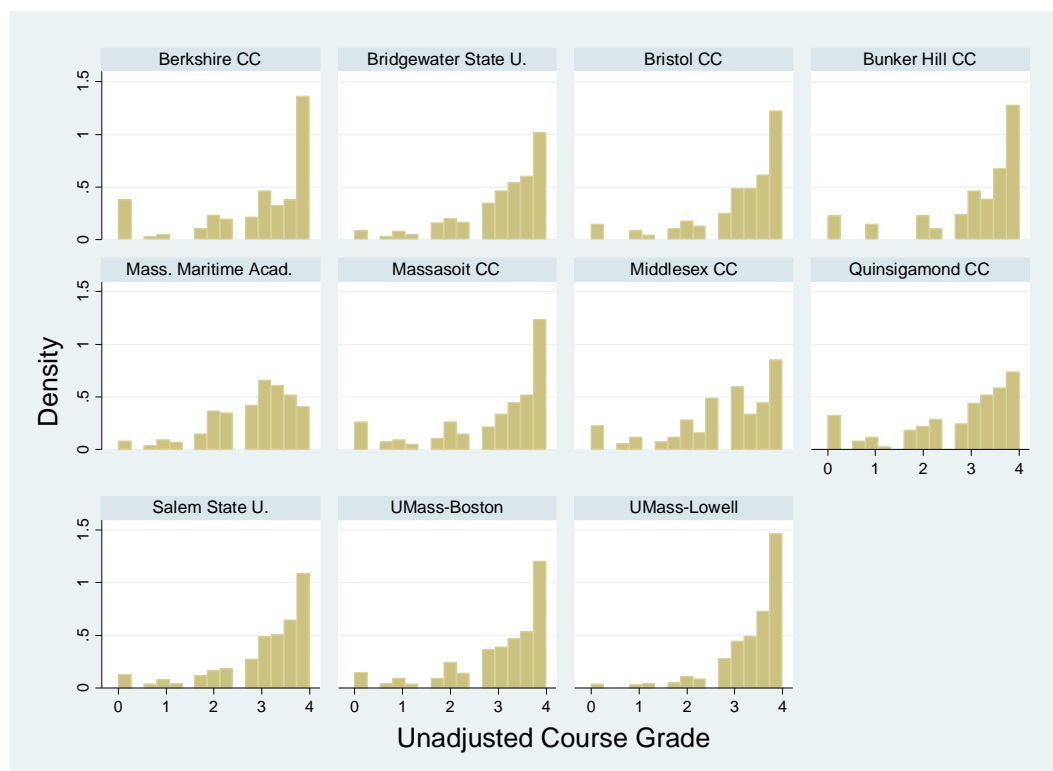
C. Standardizing course grades

As the above tables and figures show, the distribution of grades vary by institution, course type, and subject. This variation is likely produced by a combination of two different factors: (1) differences in the academic preparation of students enrolling in each campus; and (2) differences in the rigor of grading standards and difficulty of the courses selected by students across campuses. Because this study uses course grades as a measure of college readiness, it is important to isolate the variation in grades that is due to students' college preparedness. Failing to do this could bias the study's findings by allowing some students to appear more "college ready" by taking easier courses or courses with more lax grading standards. To standardize course grades, we used a two-step process: first, we adjusted grades for whether the course was a remedial course and second, we adjusted grades for the institution and course subject.

Remedial course grade adjustments. Grades in remedial courses represent a different level of college readiness than grades in non-remedial courses. A student who receives an "A" in a remedial math course, for example, may have only received a "C" if that student were taking a college-level math course. Because this study uses course grades as a measure of college readiness, it is important to adjust remedial course grades to reflect how students would perform in a college-level course in the same subject.

To do this, we conducted a within-student analysis, comparing the remedial and non-remedial course grades of students who took *both* a remedial and non-remedial course in the same subject area during their first year of college. We used a linear regression with unadjusted numerical grades as the dependent variable and a variable indicating if a course was a remedial

course as the independent variable. This allowed us to calculate the average difference within students between remedial and non-remedial grades in the same subject (we repeated this analysis separately for English language arts and math courses).

We found that, on average, students' grades in remedial math courses were 0.71 points higher than the same students' grades in non-remedial math courses. Therefore, we adjusted grades in remedial math courses downward by 0.71 points. For English language arts, students' grades were 0.36 points higher in remedial courses, so we adjusted grades in remedial ELA courses downward by 0.36 points.[27]

Adjustments for course subject and institution. In our second step, we adjusted grades for the course subject area and the institution in which the student was enrolled. We used a linear regression model with students' numerical grades (after the adjustment for remedial courses) as the dependent variable and dummy variables for each institution and subject area (math, English language arts, and other) as the independent variables. We saved the residuals from this regression model, rescaled them to a zero to four grade point scale, and defined the resulting variable as the "adjusted GPA" outcome used for the study's core analyses of MCAS and PARCC.

The following statistical equation describes the regression formula used to adjust grades for course subject and institution:

$$(1) \qquad GPA_{i,j,c} = \beta_1 ELA_j + \beta_2 Math_j + \boldsymbol{\delta_c Campus_c} + e_{i,j,c}$$

In the model, $GPA_{i,j,c}$ is a grade for course i in subject j (English language arts, math, or other subjects) at campus c. *ELA* is an indicator variable for whether a course was an ELA course, *Math* is an indicator variable for whether a course was a math course, and **Campus** represents a vector of indicator variables for each campus. The two β coefficients represent the estimated relationship between a student's course grade and whether that course was an ELA or math course, controlling for the campus in which the student was enrolled. The vector of coefficients $\boldsymbol{\delta}$ show the fixed effect on course grades for each campus (controlling for the course subject). The residuals from this regression, e, represent the adjusted grades that form the basis of the study's standardized GPA variables for each student.

We also tested an alternative grade standardization procedure that adjusts our estimates of the campus fixed effects to account for campus selectivity (i.e. estimating these fixed effects while controlling for students' prior high school MCAS scores), thereby allowing the average course grade to vary across institutions and subjects. Using this other approach had no effect on our overall findings. Additional information on this alternative grade adjustment can be found in Appendix D.

---

[27] The order of remedial and non-remedial course taking differed by subject. Students took remedial math courses prior to taking college-level math courses, while half of students who took both remedial and non-remedial ELA courses took them during the same academic term. This may influence the remedial/non-remedial grade differences, with a smaller difference between remedial and non-remedial ELA grades because the courses were taken simultaneously.

D. Summary of adjusted course grades

Table B.2 summarizes the adjusted course grades for students included in our analyses. The average adjusted course grade was 2.71, with no remaining variation by subject and institution due to the grade standardization procedures described above. After standardization, grades in remedial courses did remain lower on average than grades in non-remedial courses; this occurs because our adjustment procedure for remedial courses involved a separate within-student analysis.

Table B.2. Summary of adjusted course grades

	Number of courses	Average grade	Standard deviation
All courses	7,698	2.71	0.74
By remedial status			
Remedial	254	2.18	0.93
Non-remedial	7,444	2.73	0.72
By course subjects			
English language arts	1,529	2.71	0.69
Math	1,014	2.71	0.91
Other subjects	5,155	2.71	0.71
By institution			
Berkshire Community College	209	2.71	0.93
Bristol Community College	255	2.71	0.76
Bunker Hill Community College	282	2.71	0.85
Massasoit Community College	355	2.71	0.91
Middlesex Community College	347	2.71	0.84
Quinsigamond Community College	290	2.71	0.90
Bridgewater State University	1,930	2.71	0.71
Massachusetts Maritime Academy	1,162	2.71	0.66
Salem State University	1,081	2.71	0.76
University of Massachusetts-Boston	716	2.71	0.77
University of Massachusetts-Lowell	1,071	2.71	0.57

Source: College data from the MA Department of Higher Education and authors' calculations.

Note: Excludes grades of students who had incomplete tests, took more than one test, or performed below the guessing floor threshold on a test.

Figure B.5 shows the distribution of adjusted grades. While the adjusted course grades still skew to the right, they are more evenly distributed across the grade point scale. Figures B.6, B.7, and B.8 compare the distribution of grades by remedial status, subject, and institution. As with the overall grades, the grade adjustments resulted in a more even distribution, with a smaller concentration of courses with high grades and a larger concentration with middle-to-low grades.

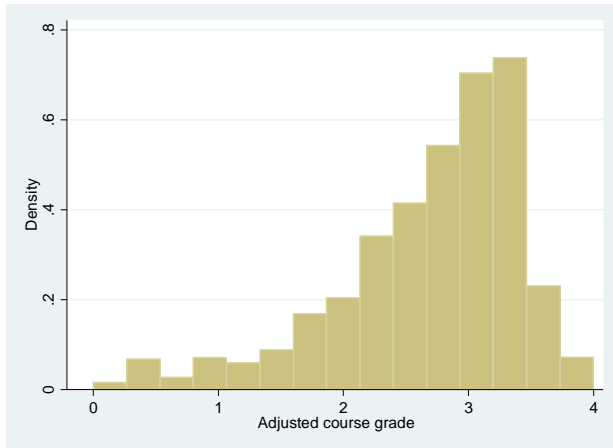## Figure B.5. Distribution of adjusted course grades across all courses

Figure B.6. Distribution of adjusted course grades by remedial status
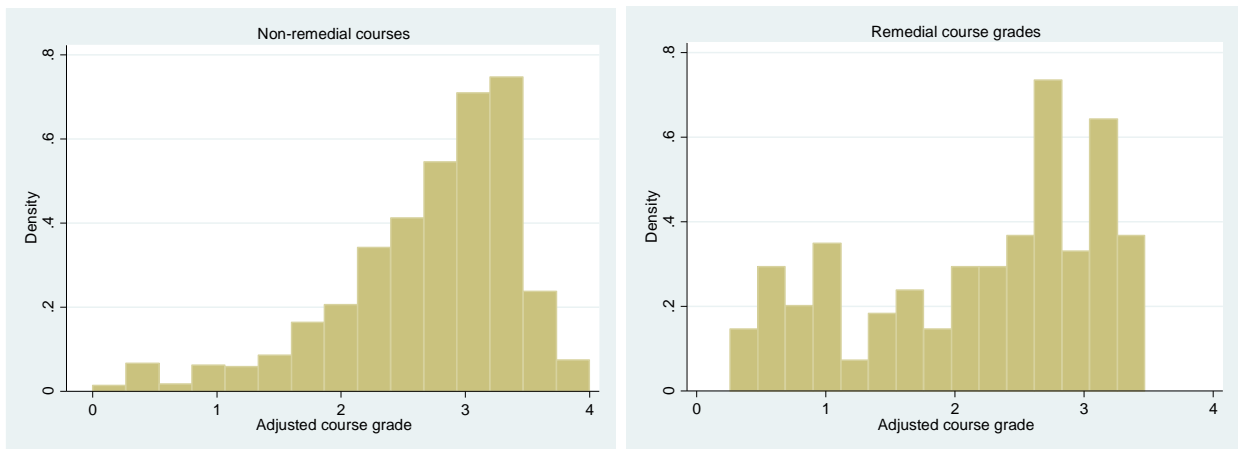

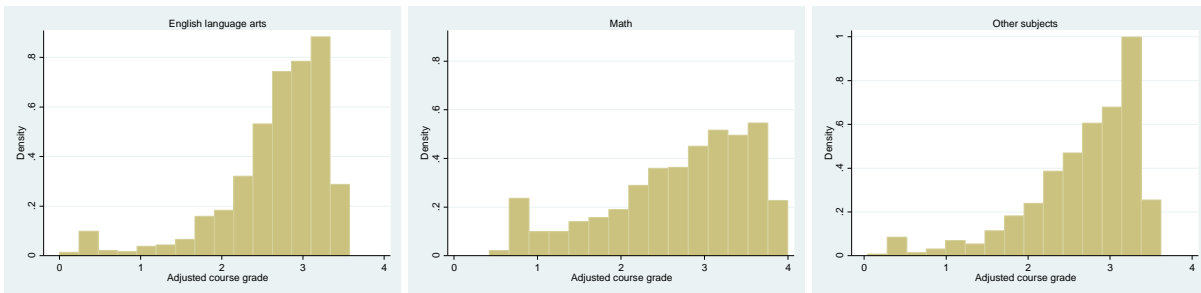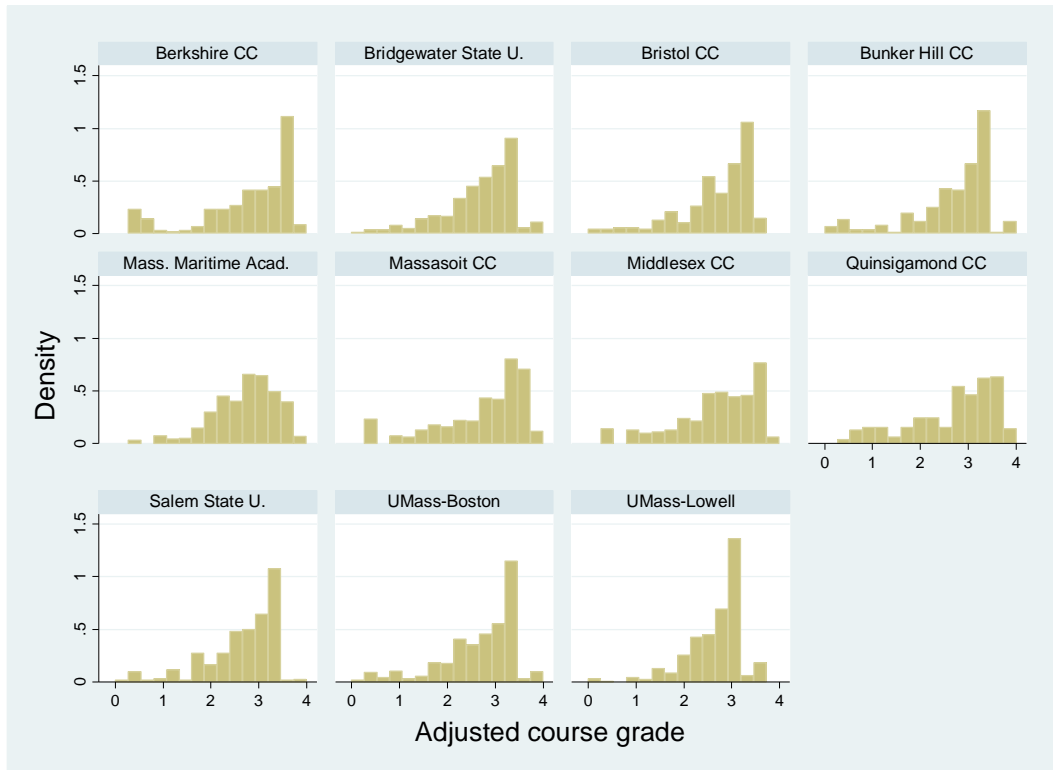## Figure B.7. Distribution of adjusted course grades by course subject

Figure B.8. Distribution of adjusted course grades by institution

APPENDIX C

ANALYSIS METHODS FOR MULTIPLE TEST COMPONENTS

This page has been left blank for double-sided copying.

Policymakers are interested in learning about the total predictive validity of the PARCC examination, which consists of both PBA and EOY components in math and ELA. However, we cannot directly examine scores on the entire PARCC test in our study data; to minimize the burden on participating students, the state only tested study participants in a single test component (either the PBA or EOY component of either the math or ELA exam). This is potentially problematic, because if the PBA and EOY test forms measure different aspects of college preparation in math or ELA, the *combination* of these forms will have a substantially different amount of predictive validity than each component does in isolation.

To address this, we used an approach that is able to combine the results from different test-takers who completed different sub-tests, using information about how these test components are related to each other. By analyzing the relationship between scores on two different test components, we can infer what the combined relationship should be between scores on those two components and a third variable of interest (i.e., college GPA or enrollment in remedial college courses).

Specifically, we obtained data on the correlations between exam components that were observed among test-takers who are outside the study sample. MCAS between-component correlations were provided by the Department of Elementary and Secondary Education (with the correlations based on all 2014 test-takers in the state), and PARCC between-component correlations were provided by Pearson (based on a much smaller sample of 2015 test-takers involved with pilot-testing the PARCC exam). We used the following between-component correlations:

1.   10th grade MCAS ELA versus 10th grade MCAS math

2.   PARCC integrated math II PBA versus PARCC integrated math II EOY

3.   PARCC grade 10 ELA PBA versus PARCC grade 10 ELA EOY

4.   PARCC integrated math II PBA versus PARCC grade 10 ELA EOY

5.   PARCC integrated math II PBA versus PARCC grade 10 ELA PBA

6.   PARCC integrated math II EOY versus PARCC grade 10 ELA EOY

7.   PARCC integrated math II EOY versus PARCC grade 10 ELA PBA

Using these correlations requires an important assumption, namely that the between-component correlations observed outside the study sample (which are based on all high school test-takers, regardless of whether they are on track to graduate or enroll in college) apply equally well to the students in the study sample (all of whom are college enrollees). If in fact the correlations between test components among the general high school population do not hold for the study sample, the study's inferences regarding the predictive validity of multiple test components may be biased in some way. However, we do not have any reason to suspect that such a bias would be systematically more favorable or less favorable to either exam.

To demonstrate how we applied these between-component correlations in the analysis, consider an analysis of the combined predictive validity of the PARCC EOY and PBA

components in a single subject (mathematics) relative to a single outcome (GPA).[28] We begin by estimating the following two equations using the student-level test score data collected for the study:

(1) $GPA_i = a_1 + \beta_{pba} * PBA_i + \varepsilon_1$

(2) $GPA_i = a_2 + \beta_{eoy} * EOY_i + \varepsilon_2$

where GPA represents the (standardized) grade point average of student *i*, *a* is an intercept term, PBA_i and EOY_i are the student scores on each exam, ε is an error term, and β_{pba} and β_{eoy} are regression coefficients of interest representing the relationship between student scores and GPA separately for the two test components.

Next, we inserted the between-component correlation provided by Pearson into the conventional equation for omitted variable bias, shown below, to extrapolate an additional type of value: the relationship between each test component and GPA, adjusting for scores on the missing component of each exam. The omitted variable bias equations account for the fact that EOY scores were left out of equation (1) and PBA scores were left out of equation (2). The equations are:

(3) $\beta_{pba} = \delta_{pba} + \left(\delta_{eoy} * corr(PBA, EOY)\right)\left(\dfrac{StdDev(EOY)}{StdDev(PBA)}\right)$

(4) $\beta_{eoy} = \delta_{eoy} + \left(\delta_{pba} * corr(PBA, EOY)\right)\left(\dfrac{StdDev(PBA)}{StdDev(EOY)}\right)$

where β_{pba} and β_{eoy} are observed in equations (1) and (2), *corr(PBA,EOY)* is the correlation between test components provided by Pearson, *StdDev(PBA)* is the standard deviation of PBA scores in the study data, and *StdDev(EOY)* is the standard deviation of EOY scores in the study data. Because we are left with only two unknown values in equations (3) and (4), we can then rearrange terms to solve for the values of δ_{pba} and δ_{eoy}. Solving these equations allows us to calculate the coefficients for a regression of GPA on both test components:

(5) $GPA_i = a + \delta_{pba} * PBA_i + \delta_{eoy} * EOY_i + \varepsilon$

The final step in the analysis procedure is to solve for the correlation coefficient provided by this regression. We can calculate that correlation using the following equation, which is derived from the standard formula for the coefficient of determination (R^2) of regression (5):

---

[28] This procedure is designed solely for two examination components; the study's analyses examined the combined predictive validity of only two MCAS or PARCC examination components at a time.

$$(6) \qquad r_{pba,eoy} = \sqrt{\frac{\mathrm{var}(\delta_{pba}*PBA_i + \delta_{eoy}*EOY_i)}{\mathrm{var}(GPA_i)}}$$

$$= \sqrt{\frac{\delta_{pba}^2\,\mathrm{var}(PBA_i) + \delta_{eoy}^2\,\mathrm{var}(EOY_i) + 2\delta_{pba}\delta_{eoy}\,\mathrm{cov}(PBA,EOY)}{\mathrm{var}(GPA_i)}}$$

In this equation every term is either observed directly in the data or derived in steps (1) through (4). The items directly observed in the study's student-level data are the variance of PBA scores, var(*PBA*), the variance of EOY scores, var(*EOY*), the covariance of PBA and EOY scores, cov(*PBA,EOY*), and the variance of GPA for students who took either the PBA or EOY exam, var(*GPA*). The values of δ_{pba} and δ_{eoy} are calculated in steps (1) through (4). Using these values we are able to calculate $r_{pba,eoy}$, which represents the combined correlation of the PBA and EOY test components with college GPA.

This page has been left blank for double-sided copying.

APPENDIX D

SENSITIVITY ANALYSES

This page has been left blank for double-sided copying.

We conducted a series of sensitivity analyses to test the robustness of the findings presented in the main report. First, we re-estimated the correlations for each test using students' original (unadjusted) GPAs and also tested if our findings were robust to using an alternative approach to standardizing GPAs. Second, we included students' baseline SAT scores as controls in the models to test whether the small baseline differences in SAT scores had effects on the results. Next, we estimated correlations using models that specified alternative functional forms, since descriptive analyses revealed some evidence of non-linearity in the relationship between college GPA and scores on some components of the PARCC and MCAS exams. We also compared the correlations between raw MCAS scores and college GPA (the correlations in our main analysis) to the correlations for scaled MCAS scores. Finally, we compared the correlations between 10th grade MCAS scores and college GPA to the correlations between study-administered MCAS scores and college GPA.

A. Correlations between MCAS/PARCC exams and original grades

We also analyzed the correlation between each of the seven MCAS and PARCC test components and original (unstandardized) college GPA. The purpose of this sensitivity analysis is to assess whether the conclusions we drew were the result of our adjustment procedure for college GPA. This is a check to see whether the standardization of GPA is masking differences between the MCAS and PARCC exams and biasing results. As we observe in Table D.1, none of the differences between MCAS and PARCC correlations are statistically significant, and the magnitudes of these correlations are similar to the study's primary results that use our standardized measure of GPA. We therefore conclude that the results are robust to the GPA adjustment.

B. Correlations between MCAS/PARCC exams and adjusted grades using an alternative grade standardization approach

We tested an alternative approach to standardizing college GPA to determine if our findings are robust to the method of grade standardization. In this approach, we standardized grades using a four step process. First, we adjusted grades for remedial courses as described in Appendix B. Second, we used a linear regression model with students' numerical grades (after the adjustment for remedial courses) as the dependent variable, and students' high school MCAS scores in ELA and math, dummy variables for each institution, and dummy variables for each subject (ELA and math) as the independent variables. By adding high school MCAS scores to the model, we control for students' ability prior to entering college. As a result, the coefficients for the institutions and subject area terms are adjusted for differences in the underlying ability of students in those campuses and courses; this model attempts to better isolate institutional or course effects (such as different grading standards or varying levels of difficulty). In our third step, we calculate a "predicted" course grade using these adjusted coefficients for each institution and subject area. Finally, we subtract the predicted grades from students' numerical grades (after adjusting for remedial courses). We defined the resulting residual grade as the basis for students' alternative adjusted GPA.

We then analyzed the correlations between each of the individual MCAS and PARCC exams and this new GPA variable. As shown in Table D.2, the results are very similar to our primary results, with no significant differences between the exams.

C. Correlations between MCAS/PARCC exams and adjusted grades, controlling for baseline SAT scores

We re-estimated the correlation between each of the seven MCAS and PARCC test components and adjusted college GPA, accounting for baseline SAT scores. Since we observed small baseline differences in SAT scores between testing groups, we controlled for this indicator to ensure that this baseline difference did not bias our results. In Table D.3 we present the results from this analysis. The findings are very similar to the findings in our primary analysis (which did not control for SAT scores): across 15 pairwise comparisons between PARCC and MCAS, 14 of the differences were not statistically significant. As we found in the analysis above (see table D.1) the correlation between MCAS math and other GPA (0.41) is greater than the correlation between PARCC Math EOY and other GPA (0.15); this difference is statistically significant (p=0.04). Thus, we conclude that controlling for SAT scores does not change the study's substantive conclusions about MCAS and PARCC.

D. Correlations between MCAS/PARCC exams and adjusted grades, using non-linear models

We performed two tests to examine whether linear models provide an appropriate specification for the relationships between GPA and test scores. First, we estimated the correlations between the MCAS/PARCC exams and standardized grades using quadratic models (Table D.4). Second, we conducted a visual analysis of nonparametric models using local polynomial regression approach (Figures D.1-D.7). Both of these sensitivity tests show that using non-linear models does not meaningfully affect the study's results and findings.

We performed a sensitivity test using curvilinear models because the descriptive analyses of test scores (see Appendix A) revealed some evidence of non-linearity in the relationships between GPA and the distribution of PARCC Math PBA, PARCC Math EOY, and PARCC Algebra II EOY test scores. To assess whether this had an effect on the main results, we examined whether a quadratic model that allows the slope of regression line to curve provides better fits for the data and increases the magnitude of the correlations. We found that the magnitude of the correlations between each test and measures of GPA are nearly the same for the linear and the curvilinear models: the estimate for each correlation changes by between 0.01 to 0.06. As with our primary analyses that use a linear model, out of 15 pairwise comparisons between PARCC and MCAS 14 of the differences are not statistically significant using this new model specification.

We also performed a visual analysis of scatter plots of the tested-subject GPA and total GPA versus the seven test forms, and compared the data to local polynomial regression lines (14 comparisons in total, as shown in Figures D.1-D.7). A local polynomial regression fits a different curved regression line across small subsets of the data; this allows the functional form to vary from one part of the distribution of test scores to the next. We inspected these scatter plots to look for non-linearity in the local polynomial regression lines, and found that in nearly all cases a linear model is approximately accurate. In addition, in each of the four cases that show some curvature (one MCAS component and three PARCC components), the correlation (r-value) for the local polynomial regression model is only 0.02 to 0.06 higher than the linear model, and the difference between the linear model and the local polynomial model is not statistically

significant. In addition, there is no evidence that using a local polynomial model has a differential effect for MCAS (0.04 increase) than for PARCC (0.02, 0.04, 0.06 increases).

E. Correlations between raw/scaled MCAS exams and adjusted grades

We compared the correlations between raw MCAS scores and standardized grades to the scaled MCAS score correlations, and found that these two sets of correlations are almost identical (Table D.5). The purpose of this is to assess whether our findings are robust to the type of MCAS scoring system used in our main analyses. This is important because we used the raw MCAS scores to conduct the primary analyses, but the scaled scores are the ones presented to districts, schools, parents, and the public (the main report findings are based on raw MCAS scores because scaled scores were unavailable for the PARCC tests, and there was a need for consistency and comparability across tests). We found that the differences between raw and scaled MCAS correlations with college GPA range from 0.01 to 0.04, and none of these differences are statistically significant. In other words, the study's results are robust to the type of MCAS scores (raw versus scaled) used in the analyses.

F. External validity of the study results, relative to high school students in Massachusetts

The study's sample was restricted to first-year college students in Massachusetts public institutions who graduated from a Massachusetts public high school and were still enrolled in college in spring of their first year. This limited sample is potentially problematic for two reasons: 1) the correlations between college students' exam scores and their concurrent grades may not be representative of the correlations we would observe if we tested 10th grade students and then observed their college grades three years later, and 2) the correlations for this limited sample may not be representative of the correlations we would find if we had test scores and grades for all Massachusetts students. We conducted two additional analyses to address these issues.

1.   Correlations between 10th grade/2015 MCAS exams and adjusted grades

For the students in the study sample, we compared the correlations between the scaled 10th grade MCAS scores and standardized college grades to the correlations for scaled MCAS scores on the study-administered test, and found that these two sets of correlations are almost identical (Table D.6). The purpose of this is to assess whether the concurrent validity of the MCAS exam (how well a student's score on the study-administered MCAS exam predicts course grades that same school year) is a good proxy for the predictive validity of the exam (how well that student's 10th grade MCAS performance predicted his or her first-year college grades). We found that the correlation between 10th grade MCAS scores and college GPA in the tested subject (0.31 in math and 0.20 in ELA) is statistically indistinguishable from the correlation between 2015 scores and GPA (0.32 in math and 0.19 in ELA). This suggests that in our sample study-administered test scores are providing a reasonably good proxy for students' high school test scores.

2.   Examining if there is a need for restriction of range adjustments

Because our correlations are based on the range of scores of students currently enrolled in their spring semester in college, they may not represent the correlations we would find if the

study could observe scores from the full population of students in Massachusetts, such as those enrolled in private colleges and those who enrolled in college initially but left before their second semester. To test for this, we examined the statewide standard deviation of MCAS scaled scores for 10th graders in 2014 (12.0 in ELA and 16.6 in math) and compared those values to the standard deviation of MCAS scaled scores in our study sample (12.4 in ELA and 16.9 in math). Because the variation in scores in our study data is very similar to the variation found in the state as a whole, this suggests that there were no notable restricted-range problems in the analysis of MCAS scores. It was not possible to examine the population-level standard deviation of PARCC scores, however, because no Massachusetts high school students took the ELA grade 10 or Math II exams in the 2014-2015 school year (due to the fact that MCAS was still required for all 10th graders in that year).

Table D.1. Comparisons of correlations between MCAS/PARCC tests and original first-year college GPA

Test	Original total GPA		Original GPA in tested subject[a]	
	N	Correlation (r)	N	Correlation (r)
MCAS ELA	120	0.24	109	0.23
PARCC ELA PBA	113	0.18	110	0.14
MCAS ELA	120	0.24	109	0.23
PARCC ELA EOY	126	0.27	116	0.28
MCAS Math	129	0.36	100	0.31
PARCC Math PBA	116	0.36	88	0.35
MCAS Math	129	0.36	100	0.31
PARCC Math EOY	121	0.25	98	0.38
MCAS Math	129	0.36	100	0.31
PARCC Algebra EOY	122	0.26	101	0.26

Source: College data from the MA Department of Higher Education and MCAS and PARCC test score data.

[a] The tested subject is ELA for the MCAS and PARCC ELA tests, and math for the MCAS and PARCC math tests.

[b] Non-tested subjects include all subjects other than ELA and math.

  *Correlations are statistically distinguishable at the .05 level, two-tailed test.

**Correlations are statistically distinguishable at the .01 level, two-tailed test.

Table D.2. Comparisons of correlations between MCAS/PARCC tests and GPA using an alternative GPA standardization procedure

Test	Alternate adjusted total GPA		Alternate adjusted GPA in tested subject[a]	
	N	Correlation (r)	N	Correlation (r)
MCAS ELA	120	0.25	109	0.24
PARCC ELA PBA	113	0.16	110	0.12
MCAS ELA	120	0.25	109	0.24
PARCC ELA EOY	126	0.28	116	0.27
MCAS Math	129	0.41	100	0.40
PARCC Math PBA	116	0.39	88	0.39
MCAS Math	129	0.41	100	0.40
PARCC Math EOY	121	0.29	98	0.45
MCAS Math	129	0.41	100	0.40
PARCC Algebra EOY	122	0.26	101	0.28

Source: College data from the MA Department of Higher Education and MCAS and PARCC test score data.

[a] The tested subject is ELA for the MCAS and PARCC ELA tests, and math for the MCAS and PARCC math tests.

[b] Non-tested subjects include all subjects other than ELA and math.

  *Correlations are statistically distinguishable at the .05 level, two-tailed test.

**Correlations are statistically distinguishable at the .01 level, two-tailed test

Table D.3. Comparisons of correlations between MCAS/PARCC tests and adjusted first-year college GPA, controlling for baseline SAT scores

	Adjusted total GPA		Adjusted GPA in tested subject[a]	
Test	N	Correlation (r)	N	Correlation (r)
MCAS ELA	102	0.30	94	0.19
PARCC ELA PBA	96	0.23	94	0.11
MCAS ELA	102	0.30	94	0.19
PARCC ELA EOY	114	0.31	106	0.28
MCAS Math	105	0.41	83	0.48
PARCC Math PBA	109	0.45	82	0.41
MCAS Math	105	0.41	83	0.48
PARCC Math EOY	105	0.20	86	0.39
MCAS Math	105	0.41	83	0.48
PARCC Algebra EOY	106	0.39	90	0.47

Source: College data from the MA Department of Higher Education and MCAS and PARCC test score data.

[a] The tested subject is ELA for the MCAS and PARCC ELA tests, and math for the MCAS and PARCC math tests.

[b] Non-tested subjects include all subjects other than ELA and math.

  *Correlations are statistically distinguishable at the .05 level, two-tailed test.

**Correlations are statistically distinguishable at the .01 level, two-tailed test.

Table D.4. Comparisons of correlations between MCAS/PARCC tests and adjusted first-year college GPA, using a quadratic regression model

	Adjusted total GPA		Adjusted GPA in tested subject[a]	
Test	N	Correlation (r)	N	Correlation (r)
MCAS ELA	120	0.25	109	0.24
PARCC ELA PBA	113	0.23	110	0.17
MCAS ELA	120	0.25	109	0.24
PARCC ELA EOY	126	0.28	116	0.26
MCAS Math	129	0.37	100	0.38
PARCC Math PBA	116	0.38	88	0.41
MCAS Math	129	0.37	100	0.38
PARCC Math EOY	121	0.21	98	0.41
MCAS Math	129	0.37	100	0.38
PARCC Algebra EOY	122	0.18	101	0.24

Source: College data from the MA Department of Higher Education and MCAS and PARCC test score data.

[a] The tested subject is ELA for the MCAS and PARCC ELA tests, and math for the MCAS and PARCC math tests.

[b] Non-tested subjects include all subjects other than ELA and math.

  *Correlations are statistically distinguishable at the .05 level, two-tailed test.

**Correlations are statistically distinguishable at the .01 level, two-tailed test.

Figure D.1. Scatter plots of average adjusted ELA GPA (left) and average adjusted total GPA (right) versus MCAS ELA scores, with local polynomial regression line
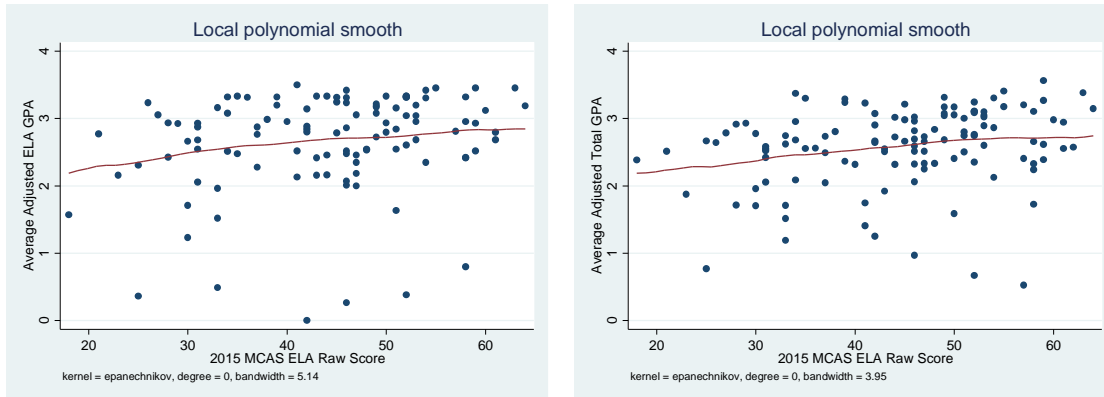

Figure D.2. Scatter plots of average adjusted math GPA (left) and average adjusted total GPA (right) versus MCAS Math scores, with local polynomial regression line
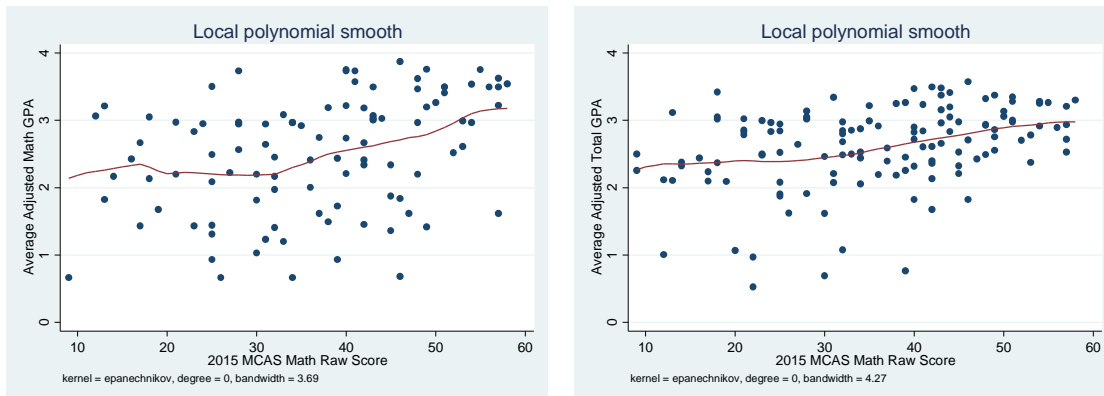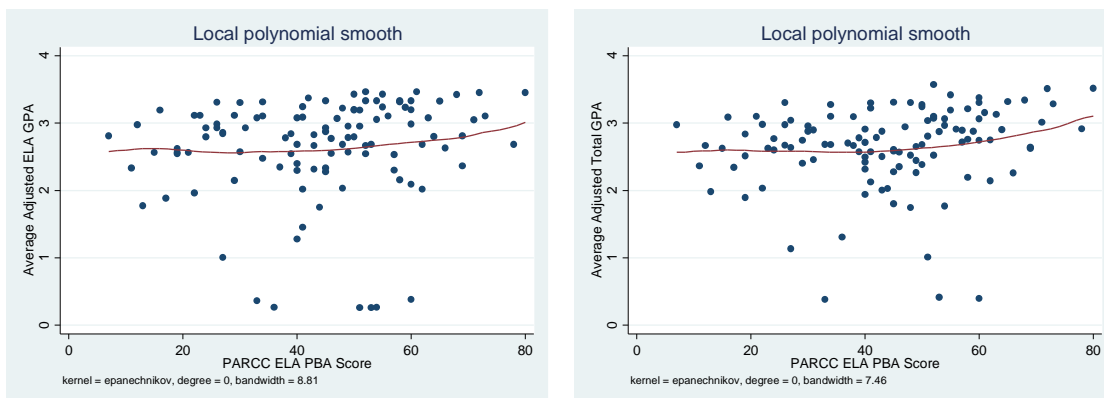

Figure D.3. Scatter plots of average adjusted ELA GPA (left) and average adjusted total GPA (right) versus PARCC ELA PBA scores, with local polynomial regression line

Figure D.4. Scatter plots of average adjusted ELA GPA (left) and average adjusted total GPA (right) versus PARCC ELA EOY scores, with local polynomial regression line
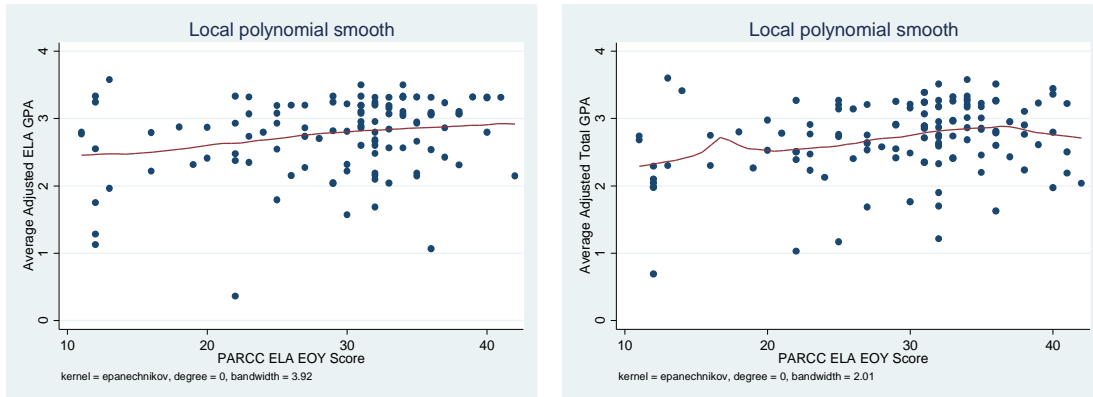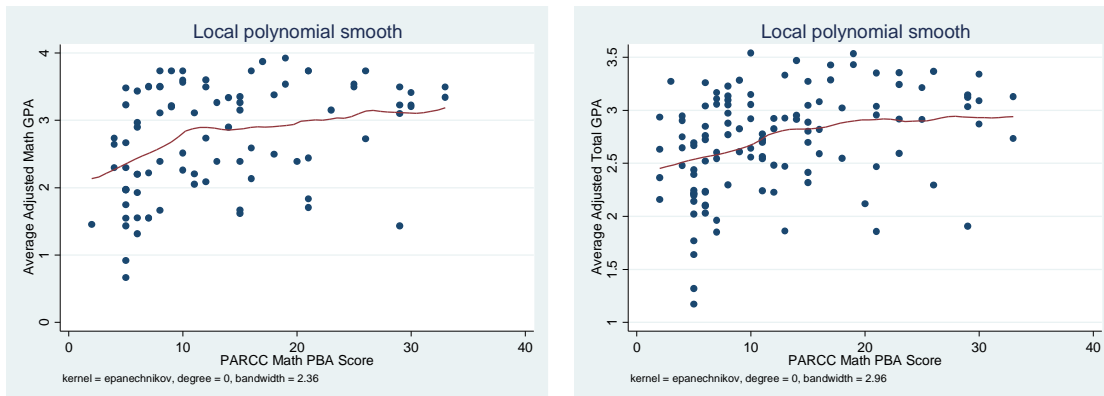

Figure D.5. Scatter plots of average adjusted math GPA (left) and average adjusted total GPA (right) versus PARCC Math PBA scores, with local polynomial regression line

Figure D.6. Scatter plots of average adjusted math GPA (left) and average adjusted total GPA (right) versus PARCC Math EOY scores, with local polynomial regression line
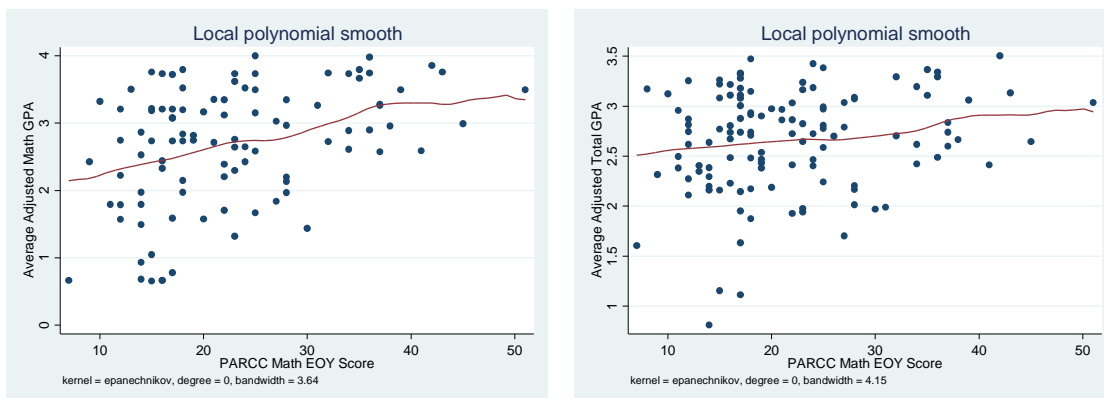
Figure D.7. Scatter plots of average adjusted math GPA (left) and average adjusted total GPA (right) versus PARCC Algebra II EOY scores, with local polynomial regression line
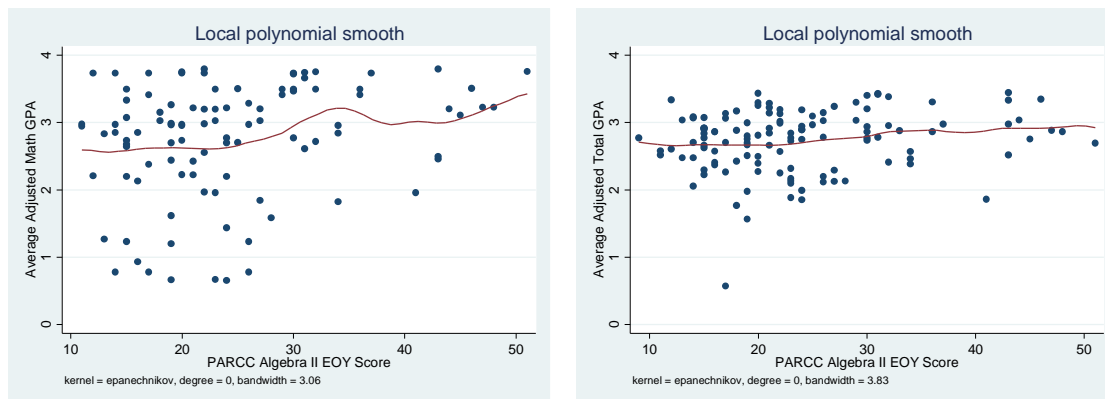

Table D.5. Comparisons of correlations between raw/scaled MCAS tests and adjusted first-year college GPA

	Adjusted total GPA		Adjusted GPA in tested subject[a]	
Test	N	Correlation (r)	N	Correlation (r)
MCAS ELA raw	120	0.25	109	0.23
MCAS ELA scaled	120	0.25	109	0.23
MCAS Math raw	129	0.36	100	0.36
MCAS Math scaled	129	0.34	100	0.32

Source: College data from the MA Department of Higher Education and MCAS and PARCC test score data.

[a] The tested subject is ELA for the MCAS ELA test, and math for the MCAS math test.

[b] Non-tested subjects include all subjects other than ELA and math.

  *Correlations are statistically distinguishable at the .05 level, two-tailed test.

**Correlations are statistically distinguishable at the .01 level, two-tailed test.

Table D.6. Comparisons of correlations between 10th grade MCAS tests or study-administered MCAS tests and adjusted first-year college GPA

	Adjusted total GPA		Adjusted GPA in tested subject[a]	
Test	N	Correlation (r)	N	Correlation (r)
Grade 10 MCAS ELA	113	0.22	104	0.20
Study-Administered MCAS ELA	113	0.19	104	0.19
Grade 10 MCAS Math	123	0.30	95	0.31
Study-Administered MCAS Math	123	0.35	95	0.32

Source: College data from the MA Department of Higher Education and MCAS and PARCC test score data.

Note: The sample for each test pair includes students who took the study-administered MCAS test in that subject, took the MCAS test in 10th grade, and had a GPA in the specified subject.

[a] The tested subject is ELA for the MCAS ELA tests, and math for the MCAS math tests.

  *Correlations are statistically distinguishable at the .05 level, two-tailed test.

**Correlations are statistically distinguishable at the .01 level, two-tailed test.

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research