EdWorkingPaper No. 22-645

Increasing High School Students' Preparation and Interest in STEM Fields: Does a Graduation Requirement Make a Difference?

Ann Mantil
Brown University

John Papay
Brown University

Preeya Pandya Mbekeani
Brown University

Richard J. Murnane
Harvard University

Preparing K-12 students for careers in science, technology, engineering and mathematics (STEM) fields is an ongoing challenge confronting state policymakers. We examine the implementation of a science graduation testing requirement for high-school students in Massachusetts, beginning with the graduating class of 2010. We find that the design of the new requirement was quite complicated, reflecting the state's previous experiences with test-based accountability, a broad consensus on policy goals among key stakeholders, and the desire to afford flexibility to local schools and districts. The consequences for both students and schools, while largely consistent with the goals of increasing students' skills and interest in STEM fields, were in many cases unexpected. We find large differences by demographic subgroup in the probabilities of passing the first science exam and of succeeding on retest, even when conditioning on previous test-score performance. Our results also show impacts of science exit-exam performance for students scoring near the passing threshold, particularly on the high-school graduation rates of females and on college outcomes for higher-income students. These findings demonstrate the importance of equity considerations in designing and evaluating ambitious new policy initiatives.

VERSION: September 2022

Suggested citation: Mantil, Ann, John Papay, Preeya Pandya Mbekeani, and Richard J. Murnane. (2022). Increasing High School Students' Preparation and Interest in STEM Fields: Does a Graduation Requirement Make a Difference?. (EdWorkingPaper: 22-645). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/9e56-3074

Increasing High-School Students' Preparation and Interest in STEM Fields:

Does a Graduation Requirement Make a Difference?

July 26, 2022

ABSTRACT

Preparing K-12 students for careers in science, technology, engineering and mathematics (STEM) fields is an ongoing challenge confronting state policymakers. We examine the implementation of a science graduation testing requirement for high-school students in Massachusetts, beginning with the graduating class of 2010. We find that the design of the new requirement was quite complicated, reflecting the state's previous experiences with test-based accountability, a broad consensus on policy goals among key stakeholders, and the desire to afford flexibility to local schools and districts. The consequences for both students and schools, while largely consistent with the goals of increasing students' skills and interest in STEM fields, were in many cases unexpected. We find large differences by demographic subgroup in the probabilities of passing the first science exam and of succeeding on retest, even when conditioning on previous test-score performance. Our results also show impacts of science exit-exam performance for students scoring near the passing threshold, particularly on the high-school graduation rates of females and on college outcomes for higher-income students. These findings demonstrate the importance of equity considerations in designing and evaluating ambitious new policy initiatives.

INTRODUCTION

An emergent policy challenge of the early 21st century is the development of a labor force skilled in science, technology, engineering and mathematics, known as STEM fields. A number of highly publicized reports have drawn public attention to this imperative in the United States. For example, in 2007, the National Academies of Science pushed for “the need for world-class science and engineering—not simply as an end in itself but as the principal means of creating new jobs for our citizenry as a whole as it seeks to prosper in the global marketplace of the 21st century” (Institute of Medicine, p. 40). This report, carrying the dramatic title *Rising Above a Gathering Storm*, drew a through line from K-12 education in STEM fields to “high-technology jobs in our knowledge economy” (p. 134).

This sense of urgency came at a time when measures of American secondary-school science skills and knowledge were provoking concerns. On the science portion of the 2006 Programme for International Student Assessment (PISA) test, 15-year-olds attending US public schools scored below the average of 30 participating OECD nations (OECD, 2007). American 8th graders showed no significant improvement in science on the Trends in International Mathematics and Science Study (TIMSS) between 1995 and 2007. These dispiriting results echoed those from the National Assessment of Educational Progress (NAEP), on which there was a significant decline in 12th-grade science performance from 1996 to 2005 – both overall and in the domains of Earth, physical, and life sciences (Grigg et al, 2006).

For state policymakers aiming to boost students’ science skills and knowledge, one potential policy lever is adding science requirements for high school graduation, such as an examination in science that students must pass. An exit-exam requirement is consistent with evidence that high-stakes testing in mathematics and English/language arts (ELA) has shifted resources and attention away from untested subjects (Au, 2007; Dee, Jacob & Schwartz, 2013; Murnane & Papay, 2010). Indeed, 10 of the 13 states with exit exams in the 2018-19 school year had a science assessment (Gewertz, n.d.); half of these states required students to pass science outright, while the other five offered more flexibility (e.g., students could pass science or social studies, students could attain a combined minimum score across multiple tests).

We studied the implementation of a science exit exam in Massachusetts, beginning with the graduating class of 2010, using data on student demographics, test performance, course-taking, high-school graduation, and college outcomes. As with most new high-stakes, standardized tests, initial failure rates were relatively high but dropped quickly as educators and students adapted to the new exam. Over time, performance by all demographic subgroups has improved, but we observe rising inequality as subgroup differences in failure rates and retest success have become increasingly pronounced. The state's growing population of English language learners (ELs), who are now almost 8% of all test-takers, is especially impacted. In the graduating cohorts of 2018 to 2020, ELs comprise 41% of the students who fail on their first attempt and 53% of those who never pass.

We also find that the science exam led to substantial shifts in local practice. Massachusetts is unusual among states with science exit exams in the flexibility it affords, with students permitted to test in 9th or 10th grade and in one of four science content areas. In the years following the introduction of the exit exam, science course-taking patterns shifted noticeably, as many schools and districts changed their course sequences to test students in 9th grade and in biology. While Massachusetts policymakers went to great lengths to afford local control, the state has instead seen a shift towards standardization of the science curricular sequence, particularly for students' initial year of high school.

After reviewing previous research on high-school exit exams and presenting the policy context in Massachusetts, we address four research questions. The first two concern the design and initial implementation of the policy:

(1) How did the state navigate the tensions inherent in the design of its high-school exit exam?

(2) How did students fare on the test, and for the subset of students who scored near the passing threshold in the early years of the test, what was the causal impact of barely passing instead of barely failing on their later educational outcomes?

The remaining questions relate to changes that occurred over time:

(3) How have schools/districts responded to the exit-exam requirement?

(4) How have students' testing outcomes changed over time, overall and for key demographic subgroups?

BACKGROUND AND CONTEXT

High-school exit exams have received a great deal of attention from researchers interested in evaluating their impacts on student outcomes, including high-school dropout and graduation, college enrollment and degree completion, and labor-market earnings. One line of research uses difference-in-differences or interrupted time series approaches to compare outcomes of students before and after the introduction of the exit exam (Baker & Lang, 2013; Hemelt & Marcotte, 2013; Holme et al., 2010; Dee & Jacob, 2006; Caves & Ballestra, 2018; Warren, Grodsky, & Lee, 2008). These studies seek to explore the impact of introducing the policy on educational outcomes. Most research has focused on high school graduation outcomes, but the few that have explored college-going have found limited impacts (Holme et al., 2010).

Another line of research employs regression-discontinuity designs to estimate causal impacts of barely passing exams in mathematics or English language arts on students who score near the cutoff (Papay, Mantil & Murnane, forthcoming; Papay, Murnane & Willett, 2010; Papay, Murnane, & Willett, 2014; Reardon et al, 2010; Ou, 2010; Polson, 2018; Clark & Martorell, 2014). This work largely finds that barely passing (as opposed to failing) reduces high school graduation outcomes, particularly for students from low-income families, with more modest impacts on longer-term outcomes.

How schools respond to exit-exam requirements has received less attention and study. There is some evidence that schools, especially those serving large numbers of children from low-income families, increase attention on students whose prior scores were "on the bubble" for passing exit exams (Neal & Schanzenbach, 2010; Booher-Jennings, 2005). The larger literature on test-based accountability in mathematics and English Language Arts has also established that schools reallocate instructional time away from untested subjects (Holme, 2008).

To our knowledge, our study is the first to focus specifically on an exit exam in science. For policymakers with an interest in boosting students' STEM skills, how might a science exit exam be designed, and what are the expected consequences of its implementation, for students overall and for equity across subgroups?

Massachusetts Context

In 1993, the state legislature passed the Massachusetts Education Reform Act (MERA), which mandated the development of high-school exit exams in core academic subjects, including science. The graduating cohort of 2003 was the first for which the passage of 10th grade mathematics and English Language Arts (ELA) tests was a requirement for a diploma; the state put off implementation of tests in science and social studies, the two other core-area subject tests specified in MERA. Students who failed to obtain a scaled score of at least 220, considered passing, on their first attempt at the math and ELA tests were afforded multiple retest opportunities in subsequent school years. Students who did not pass on retest had several avenues to appeal (described below). Students who completed local high-school requirements but never passed the test or had a successful appeal earned a Certificate of Attainment rather than a traditional diploma.

When the state's Board of Elementary and Secondary Education (BESE) returned in 2005 to discuss implementing a science exit exam, Massachusetts students already were scoring well in science relative to their peers around the country. In 2005, only four other states had higher average scores than Massachusetts on the NAEP 8th grade science test, and the state's average score had improved significantly since 2000. Its strong performance and highly rated science standards, which were the first in the nation to incorporate technology and engineering, seemed to position the state's students well for STEM success (Rennie Center, 2007).

Despite this relatively good news, Massachusetts policymakers worried that the supply of high-school graduates well-prepared for science careers would not meet projected demand. Thirty percent of employment growth in Massachusetts between 2004-2014 was projected to come from STEM fields, and those fields were experiencing high job vacancy rates relative to others in the state (Conaway, 2007).

In contrast to growing demand, the share of college-bound students in Massachusetts who were considering STEM majors was below the national average (MASSIP, 2009), and STEM majors made up a declining share of college graduates in the state. Concern over the interest level in STEM among students exiting the state's high schools was accompanied by rising alarm over the adequacy of their preparation for college-level work. According to a brief produced by the Massachusetts Department of Elementary and Secondary Education (DESE), "[r]ecent preliminary reports from the new statewide School-to-College Database, jointly produced by the Massachusetts Department of Education and the Board of Higher Education, indicate that among 2005 high school graduates who entered a Massachusetts public college or university as a first-time, full-time, degree-seeking candidate in fall 2005, more than one in four enrolled in a developmental (remedial) mathematics course. This included 5 percent of University of Massachusetts students, 16 percent of state college students, and 55 percent of community college students" (Conaway, 2007, pp. 6-7).

The impetus to proceed with adding the science exit exam as a graduation requirement came from multiple sources. The state's business community advocated for the need to develop human capital, given technological advances and the emerging role of biotechnology in creating jobs in Massachusetts (Sacchetti, 2005b; Reville, P., personal communication, March 16, 2022). Then-Governor Mitt Romney was a strong proponent of adding science to the required high-school exit exams and urged the BESE to do so; the new policy was formally adopted in June of 2005 (Sacchetti, 2005a).

The new requirement, which was binding beginning with the graduating class of 2010, was considered one part of an ambitious policy agenda to increase STEM proficiency. Other pieces were the development of a recommended high-school course of study that included three years of laboratory sciences and the creation of programs designed to improve the science content knowledge of the state's teachers (Conaway, 2007; Executive Office of Education, 2008). Additionally, the BESE voted to require students scoring between 220 and 238 on the mathematics and ELA exit exams (in the "Needs Improvement" category) to complete an Educational Proficiency Plan (EPP), in an attempt to raise the bar for performance in those subjects to "Proficient". This new EPP requirement also took effect for the graduating cohort of 2010.

DATA & METHODS

We combine annual student enrollment and test-score datasets from the Massachusetts DESE to construct a longitudinal database for all K-12 public school students in the state beginning with the 2002-03 school year. This database includes MCAS scores, information on student demographics and school attendance, and high-school completion outcomes. Beginning with the 2011-12 school year, state administrative data includes students' high-school course-taking. College enrollment and graduation data comes from the National Student Clearinghouse (NSC), which matches to state records using names and dates of birth. These NSC data include nearly all US colleges and universities. We have complete NSC records on students for nine years after they took the 10th grade MCAS exit exams.

In our analytic samples, we define cohorts of students by their expected year of high-school graduation, which is four years after their initial 9th-grade enrollment (or three years after initial 10th-grade enrollment, for the minority of students who were not enrolled in a Massachusetts public high school in 9th grade). Each cohort includes approximately 70,000 students. We retain in our sample students who took the science exam and subsequently dropped out or transferred out of the state's public school system.

We categorize students as low-income or higher-income based on whether they had *ever* been eligible for free or reduced-price lunch (FRPL) during their enrollment histories in the state. We do this to correct for the underreporting of FRPL eligibility among older students (Mirtcheva & Powell, 2009). In 2008, the maximum annual income for reduced-price lunch eligibility for a family with two adults and two children was $36,641 (this is equivalent to approximately $49,744 in 2021 dollars). Importantly, students whom we identify as “higher income” are quite heterogeneous: they come from families whose incomes fell just above this threshold up to those who were quite wealthy. Other demographic indicators, including race/ethnicity, gender, English proficiency, special-education status, and whether the student attended a high school in one of the state’s urban school districts or urban charter high schools, come from the year the student first took the science exit exam.

While many of the analyses we present are descriptive, we also employ the regression-discontinuity design used in our previous work, beginning with Papay et al. (2010) and extended in Papay et al. (2014). This method allows us to estimate the causal impact of barely passing the science exam, as opposed to barely failing it, on students’ later outcomes of interest. The forcing variable is a student’s raw score from their first attempt at the science examination, centered on the minimum passing score. This approach, described in detail in earlier studies, relies on the identifying assumption that students near the passing threshold are essentially equivalent in their underlying knowledge and skill. The strict adherence to an exogenously determined cut score divides these students into treatment and comparison groups, where a significant difference in their later outcomes represents the causal effect of barely passing, as compared to barely failing, the science exam. However, this estimate only applies to students who score quite near the cutoff on their first attempt.

As in Papay et al. (2022, footnote 7), we calculate optimal bandwidths (h^*) for each outcome of interest and test the robustness of the results to bandwidth selection. In each regression-discontinuity model, we include fixed effects for the year and subject of the student’s first high-school science test. We also include a vector of student-level covariates, including family income, race/ethnicity, gender, English language proficiency, special education status, and the student's attendance in the school year prior to the year they first took the science exam. In a series of models testing the heterogeneity of impacts, we interact the passing variable with the relevant demographic indicator of interest.

POLICY DESIGN AND INITIAL IMPLEMENTATION

Designing the Science Exit Exam

The Massachusetts science exit-exam policy is more nuanced than those in most other states with similar requirements. In Massachusetts, the June science exams are offered in four different 10th-grade content areas: introductory physics, biology, chemistry, and technology/engineering. Each exam includes a set of common items, which are used to calculate students' individual scores, and matrix-sampled field test items that differ across test forms. The common items include 40 multiple-choice and 5 open-response items that sample from the state's content standards and assess both factual recall and scientific reasoning.[1]

To fulfill the graduation requirement, a student must obtain a scaled score of at least 220 on one of the science exams; the test scale ranges from 200 to 280. If they fail on their first attempt, they can retake the exam in the same subject or test in a different science subject. Unlike in most other states, Massachusetts students cannot use their performance on an alternate assessment (e.g., an AP exam) to fulfill the science requirement; this is because MERA mandates that the exam assess the high-school science standards adopted by the state.

Also, students in other states usually take the science exit exam for the first time as 10th graders, in accordance with a provision of the No Child Left Behind Act. In Massachusetts, DESE intended the science exam to be a 10th grade test, but the US Department of Education allowed the state to administer the test in 9th grade provided that it assessed 10th-grade standards

[1] Released test items are available for review at https://www.doe.mass.edu/mcas/testitems.html.

(Nellhaus, J., personal communication, March 25, 2022). This flexibility was offered for multiple reasons, such as (1) to allay concerns over imposing an additional testing burden on 10th grade students, who would otherwise take three high-stakes tests during a relatively short window in the spring; (2) to align the science tests with existing high-school course offerings instead of administering a single integrated assessment of multiple domains; and (3) to encourage schools in the development of technology/engineering courses, which were primarily offered in 9th grade. Starting in 2007, 9th graders were given the option to test and could meet their science requirement if they scored at 220 or above. Those who passed the test in 9th grade had their score "banked" as their official 10th grade score. Students who did not take a 9th grade science test were required to take one in 10th grade.

A key decision point in the implementation of an exit exam is the setting of the passing threshold. Policymakers must navigate a tension inherent in the policy design: a higher cut score may incentivize larger gains in curricular rigor and student mastery, but also means that more students fail the exam and become at-risk for not graduating. In Massachusetts, over 18% of the first cohort of 9th-grade science test-takers scored below the passing threshold in 2007, indicating a willingness on the part of the state's education leaders to set the cut score relatively high in the performance distribution.

Though its passage is required for graduation, retest opportunities for the science exam are more limited than in math and ELA. In the first two years, students needing to retest had to wait until the following year's standard June administration. A February retest administration was added in 2009, but only in biology. Students who needed to retest and opted for one of the other three subjects were only able to test in June. This decision, which was due both to costs and the difficulties with field testing enough test items in the other subjects, proved consequential in how schools and districts responded to the policy.

The state also built in safeguards for students who retest but do not pass the science exam. Districts have the option to submit a cohort or portfolio appeal for such students (Massachusetts DESE, 2020).[2] In a cohort appeal, a student's grade point average and MCAS scores are compared to those of other students in the same course in the same school. The portfolio appeal involves the submission of student work samples. Only 4-5% of students who fail the exam on their first attempt successfully fulfill the requirement via an appeal.

Patterns in Initial Implementation

To analyze the implementation of the science requirement, we focus on the graduating cohorts of 2010-12. We do so for two reasons. First, these students were the first for whom a high-school diploma was contingent on passing the science exam, in addition to passing the math and ELA tests. As a result, they illustrate the impact of introducing the new policy. Second, they took the test sufficiently long ago that we can track longer-term outcomes. We limit our sample to students who tested for the first time as entering 9th graders or in the following year as 10th graders. We exclude the 5,578 students whose first science test occurred at a different time, because they had been retained in grade or they entered a Massachusetts public school later than 10th grade, along with 73 students missing on key demographic indicators. The final sample includes 209,544 first-time science test-takers, or 69,848 per year, on average.

In the following sections, we describe first the overall patterns in test performance, success on retest, and causal impacts on students scoring near the cutoff. We then investigate heterogeneity by demographic subgroup.

[2] A third type of appeal is the transcript appeal, which is only for students who transfer to a Massachusetts public high school late in their senior year.

Overall Patterns

Across all grades and subject tests, about 12% of students in the initial three cohorts failed the science exam on their first attempt. The failure rates for these cohorts in the other tested subjects, by comparison, were 6.7% for math and only 2.8% for ELA. Not surprisingly, the science failure rate was highest for the cohort of 2010 (14.3%), the first to be subject to the new exit-exam requirement. Of the students who initially failed science, about half of them passed both the other tests on their first attempt, meaning that the science exam was the only state requirement posing an obstacle to graduation.[3] Overall, only 83% of students statewide passed all three exit exams on the first try, compared to 91% of the 2009 cohort that had passed both required exams.

We see differences in performance when examining the timing and test selection in the initial three cohorts of science test-takers, as shown in Table 1. First, the failure rate of students who took a science exam in grade 9 was three percentage points higher than for students who first took a science exam as 10th graders. Second, there were pronounced differences by science content area, with a 21% failure rate on the chemistry test compared to 16% for physics and 11% for biology and technology/engineering.

<Insert Table 1 about here>

We explore these differences further by comparing students with similar academic performance on high-school entry. The left panel of Figure 1 displays the predicted probabilities of passing, conditional on a cubic function of students' 8th grade MCAS science scores, for the three largest test-taking groups: 9th and 10th grade biology and 9th grade physics.[4] These three

[3] Many of these students would have had to complete an EPP in math and/or ELA as well, if they passed the test but their score was below 240.
[4] Figure 1 excludes 17,892 students with missing 8th grade scores, or 8.5% of high-school test-takers. These students did not attend a Massachusetts public school for 8th grade.

groups included over 94% of test-takers in the initial years of the exam. Focusing on the lower half of the distribution, we see that students who took the biology test in 10th grade were more likely to pass than those with similar 8th grade MCAS science scores who tested in biology as 9th graders. The majority of students did test in 9th grade, probably with the goal of spreading the required exit exams out over two years, but these results indicate that testing in 10th grade may be more conducive to success, at least for students in the lower tail of the performance distribution, as might be expected given that the test was targeted at a 10th grade level.

<Insert Figure 1 about here>

There were even more pronounced differences in passing rates by subject test. In the right panel of Figure 1, we show that students who took the chemistry exam were far less likely to pass than students who took any other test after scoring similarly in 8th grade. Physics test-takers were also somewhat more likely to fail than those with the same 8th grade scores who took biology or the technology/engineering exam. At least two explanations could account for this: the tests were not of comparable difficulty, meaning that the required level of proficiency was different for different tests, or the variation in curricular materials, instructional quality, and other school-based factors made it harder for students in different subjects to reach the same level of proficiency. In choosing to offer four different science tests, Massachusetts policymakers had aimed to afford a great deal of flexibility and deference to local curricula, but the decision meant that students confronted different challenges in achieving the proficiency standard.

The subject test students initially took also influenced their retest behavior, as shown in Table 1. Among biology test-takers who failed and then retested, 92% took the biology test again. But students who took one of the other three subject tests generally retested in a different subject, almost always biology. This is perhaps because the state offered a February retest in biology only, so students who failed Biology could retake it. Students who initially failed chemistry and then retested in a different subject experienced the most success on retest, which serves as additional evidence of the difficulty of the chemistry test.

Across the four subjects, most of the students who failed on their first attempt did eventually pass the science exam on retest. Only 4% of students statewide never fulfilled the science exit-exam requirement, even in the cohorts first impacted by the policy; half of them had never attempted to retake the test, presumably because they left the Massachusetts public-school system.

Figure 2 depicts the different pathways that students took after failing their first science test, including participation and success with retest and appeals. The figure also includes the percentage of each group that graduated from high school "on time," meaning within two years of the science test (for first-time 10th grade test-takers) or three years (for 9th grade test-takers).[5] The overall on-time graduation rate for these cohorts was 85%, but for students who initially failed science, the rate was only 53%.

<Insert Figure 2 about here>

Most students (85%) who initially failed science did retest, and almost half passed on their first retest attempt; another third needed multiple retest attempts but eventually did achieve a passing score. Among students who failed but eventually passed on retest, the on-time graduation rate was 70%, which is quite comparable to the 73% rate for students who barely passed the test on their first attempt. The appeals process was activated for relatively few

[5] These percentages include students who transferred out of the state's public schools prior to their expected graduation date in the denominator but not in the numerator. In the initial cohorts, 347 students (less than 0.2% of test-takers) graduated from high school without passing a science test or successfully appealing, likely to confusion over which students were in the graduating cohorts bound by the new requirement.

students; districts filed an appeal for just over 3% of all students who initially failed, as shown in Figure 2. Most appeal attempts met with success, and students who were granted an appeal graduated from high school in similar numbers as their peers who passed on their first retest.

What impact did performance on the science test have on students' later educational attainments? We use a regression-discontinuity approach to estimate this for students scoring near the passing threshold, using outcomes related to secondary-school completion and college. The first, on-time high-school graduation, is an indicator taking the value of 1 for first-time 9th grade test-takers who graduated from high school within 3 years of the test and for first-time 10th grade test-takers who did so within 2 years. We also examine five-year high-school graduation, which gives students in each grade one additional year to fulfill the competency determination and other local graduation requirements.

We find overall impacts on both high-school outcomes for students near the cutoff in the first three cohorts affected by the science requirement, as shown in the first row of Table 2. Just passing the test on the first attempt increased the probability of graduating on time by approximately four percentage points. The impact on five-year graduation is substantially smaller – just over one percentage point – and marginally significant, suggesting that the new requirement may have slowed some students' progress through high school but did not prevent them from graduating after an extra year.[6] We report on whether these impacts vary by demographic subgroup in the following section.

<Insert Table 2 about here>

[6] The cross-validation procedure produced different optimal bandwidths for on-time high-school graduation (h*=2) and five-year graduation (h*=3). While the results shown in Table 2 are from models using these different bandwidths, estimates from the same bandwidth (either h=2 or h=3) follow the same pattern described here.

There is little evidence of an overall impact on college outcomes for students near the cutoff. We focus on whether the student enrolled within four years (or five years if the test was first taken in 9th grade). We also examine whether students graduated from college within nine years of taking the science exam in 10th grade (or 10 years for 9th-grade test-takers). While the point estimates for enrolling and graduating from any college are statistically significant in Table 2, there is no corroborating visual discontinuity at the passing cutoff in the associated plots of the sample mean probabilities. In the absence of such evidence, we conclude that there is a weak, if any, overall effect of barely passing the science exam, rather than barely failing it, on the later college outcomes of students who scored near the passing threshold on their first attempt.

Heterogeneity by Subgroup

A key consideration in assessing the implementation of the science exam is whether there are differences in failure rates, retest participation and success, and impact on later outcomes by demographic subgroups. While there are many dimensions on which to examine heterogeneity, we focus on three: gender, family income, and English language proficiency. We investigated differences by race/ethnicity, urbanicity, and disability status, but did not find clear-cut patterns using those indicators, particularly in contrast to the three detailed below.

Gender: In the first three cohorts of students subject to the science CD requirement, first-time test-takers were evenly divided by gender; as seen in Table 3, there was no difference between male and female students in their probability of passing on the first attempt. Among the 12% who failed, the female students had slightly higher rates of retest participation and success. The final column displays the on-time graduation rate of the students in each group who had failed on their first attempt. The difference of 10 percentage points, which favors female students, is much larger, due to additional factors beyond whether the science exit-exam requirement had been fulfilled.

<Insert Table 3 about here>

We find clear evidence of heterogeneous effects by gender in our regression-discontinuity analyses of the impact of barely passing vs. barely failing the exam. Returning to Table 2, barely passing the science exam increased the probability of graduating from high-school on time by seven percentage points for female students, with no corresponding impact for male students. While the impact on five-year graduation for females was somewhat smaller, it is still significantly different than the impact for males (p=.001).

In the top panel of Figure 3, we plot the sample mean probabilities of graduation for female and male students. There is confirmatory visual evidence of a discontinuity at the passing threshold for females but not for males. These discontinuities and the robustness of the parameter estimates in Table 2 to bandwidth selection strongly support the inference of a causal impact on high-school graduation for female students, but not males. However, we do not find heterogeneity in college impacts by gender. The few significant estimates involving the college outcomes of male students are not replicated at other bandwidths. Also, the plots of sample mean probabilities of college enrollment and graduation by gender (not shown) contain no clear evidence of discontinuities at the cutoff.

<Insert Figure 3 about here>

Previous studies using regression-discontinuity designs have not reported differential impacts of barely passing an exit exam by gender[7]; our own analysis of the introduction of the Massachusetts math exam found high-school graduation effects that were quite similar for males

[7] The exception is Reardon et al (2010), which finds that female students (but not males) in California who barely fail the math exam are less likely to take advanced math courses in 11th grade.

and females on the threshold of passing. Why are there impacts on female students scoring near the cutoff of the science exam, but not males? The two groups exhibit no notable differences in the grade and the subject in which they originally tested (with the exception of the tech/engineering exam, taken by less than 3% of students), nor on other demographics like race/ethnicity, English proficiency, and family income. Their passing rates on the 8th grade science test were similar.

One striking difference is that 39% of males near the passing threshold were receiving special-education services in high school, compared to only 26% of females. It may be that more male students in this group already perceived themselves as low-performing and were therefore less susceptible to the "signal" provided by their science test score. The on-time graduation rate of males near the passing cutoff was 65.8%, compared to 75.2% for females, indicating that factors beyond performance on the test were differentially impacting students' progression through high school.

Family Income: About 43% of science test-takers in the first three cohorts were from low-income families, as shown in Table 3. However, the vast majority of students who failed (82%) came from this group. Put differently, almost a quarter of low-income students failed the exam on their first attempt, compared to only 4% of their peers from more affluent families.

Of course, this pattern might be explained by preexisting group differences in academic skills at the time of high school entry. In the first plot displayed in Figure 4, we show that differences in 8th grade standardized test scores for low-income and higher-income students account for some but not all of the differences in their probability of passing the science exam. The graph shows the predicted probability of passing for the two groups, conditioning on a cubic function of students' standardized scores on the 8th grade science test. In the upper part of the distribution, predicted probabilities for all groups are nearly 1; almost all students who scored well in 8th grade passed the high-school science test the first time that they took it. But among those scoring below the state average as 8th graders, students from low-income families were less likely to pass than those from more advantaged backgrounds with the same 8th grade scores.

<Insert Figure 4 about here>

These pronounced differences carry troubling equity implications, given the high stakes of the science exit exam. Students’ experiences during grades 9 and 10 apparently produced gaps in performance by family income, even when comparing students who had scored at the same level as 8th graders. Differences in access to high-quality curricula and experienced, highly trained teachers are likely contributing factors.

Among those who failed, retest participation was quite similar among low-income and higher-income students (Table 3), but higher-income students were much more likely to pass on their first retest attempt (56%, compared to 39% of low-income students). Again, though, this difference might simply reflect their performance on the initial test. If more of the higher-income group was initially quite close to the passing threshold, then their differential success on retest would be expected. However, as illustrated in the top right panel of Figure 4, there continue to be pronounced differences by family income in the predicted probability of passing the first retest when comparing student with similar initial science scores. While students who initially scored near the passing threshold nearly all passed their first retest, the curves diverge lower in the distribution, indicating that students from low-income families were less likely to pass on retest. These differences, even between students with similar performance on the first test, suggest that students from low-income families received less preparation and support with retesting than did students from higher-income families, on average.

While many students who failed the first retest eventually passed, there remained a 10-percentage point difference in the proportions of low-income and higher-income students who never passed the science exam. We see an even larger disparity in high-school graduation by family income. Among low-income students who failed the science exam on their first attempt, only 49% graduated from high school on time, compared to 69% of higher-income students.

While these differences are purely descriptive, our regression-discontinuity estimates show an impact of barely passing the science exam, as opposed to barely failing it, on high-school graduation for students near the passing cutoff from *both* low-income and higher-income families. The coefficients for the two groups reported in Table 2 do not differ significantly (p=.113), and visual discontinuities are apparent in both plots of the sample mean probabilities of graduation for students around the cutoff. However, the larger impact estimate for low-income students is broadly consistent with many prior studies on the introduction of exit exams (Papay et al, forthcoming; Papay et al, 2015; Ou, 2010; Reardon et al, 2010) and indicates that family advantage, and its correlation with higher-quality instruction and resources in and out of school, may help to blunt the effect on high-school graduation for students scoring near the cutoff.

For college outcomes, though, we see larger impacts on higher-income students. For higher-income students, barely passing increased the probability of enrolling in college by about 6 percentage points and of graduating from college by about 3 percentage points. These effects are robust to bandwidth selection, and discontinuities are clearly visible in the associated plots, which appear in the bottom panel of Figure 3. Estimates for low-income students, on the other hand, are much smaller and not accompanied by a pronounced visual discontinuity.

Looking at graduation from four-year colleges and universities in particular, we see an effect of about a percentage point for higher-income students. However, given that only 20 percent of higher-income students at the science cutoff go on to complete a Bachelor's degree, this is a meaningful impact. This result is quite consistent with our study of the causal impact of just passing the Massachusetts math exit exam (Papay et al., forthcoming), which found small but significant impacts on four-year college outcomes for higher-income students, but not low-income students.

English Proficiency: While the share of ELs in the Massachusetts public-school population has been rising steadily, students who remained ELs in 10th grade comprised less than 4% of students in the first three cohorts subject to the science exit-exam requirement. However, they made up almost 17% of the students who failed, and the 56% failure rate for ELs was by far the highest of any demographic subgroup included in Table 3.

English learners also experienced markedly less success on retest than other groups. While their rate of retest participation was quite similar to other subgroups, less than 30% of the ELs who had failed the science exam passed their first retake. Nearly a quarter of the ELs who failed and retook the science test at least once never passed, which is again the largest share of any demographic subgroup reported in Table 3. In results not shown, we find that the intersection of limited English proficiency with other demographic variables does not account for this result. For example, among ELs who initially failed the science exam, the passage rate on first retest was 34% for those in non-urban districts, 33% for those from higher-income families, and 31% for those without a diagnosed disability. There is also virtually no difference by gender.

In the bottom panel of Figure 4, we plot the predicted probability of retest success for ELs and non-ELs who had the same scores on their first attempt at the science exam. This comparison reveals that ELs who had just missed the passing score were actually *more* likely to pass on retest than students fully proficient in English with the same initial score. However, ELs who initially scored below the 25th percentile were far less likely to succeed on retest than non-ELs who had received the same score on their first attempt. This pattern may reflect the ongoing process of language acquisition among ELs, which likely hampered many of them from fully demonstrating their science content knowledge and skills. ELs who barely failed on their first attempt may have been more likely to pass on retest than other students because their language skills improved in the interim, while those farther below the cutoff may not have had sufficient time and practice with English to fully understand and respond to the test items. The linguistic complexity of the biology test, which is the exam taken by over 90% of students when they retest and contains written passages that students must read and analyze, may be a contributing factor.

Our regression-discontinuity evidence suggest no impact of barely passing the exam on the later educational attainments of ELs (Table 2). It may be that ELs were less sensitive to the impact of barely passing or failing the science exam because so many of them had additional obstacles to contend with. For example, of the students near the science cutoff who also took the other tests, 37% of ELs had failed one or both of the other exit exams, compared to 19% of non-ELs. Alternatively, since ELs scoring just below the cutoff had higher-than-average success on the science retest, they may not have experienced longer-term effects from initially failing it.

CHANGE OVER TIME

Since the implementation of the science exit-exam requirement, Massachusetts has experienced substantial demographic shifts in its public-school student population (Papay et al, 2020). Over the last decade, the percentage of high-school test-takers who are ELs has increased from 4% to over 8%. Quantifying the trend in enrollment by family income is complicated by the state's 2015 shift from National Student Lunch Program (NSLP) eligibility to a different measure.[8] The new “economic disadvantage” measure captures students up to a lower household income threshold than did NSLP eligibility and is therefore not directly comparable. However, even though the new measure is an undercount compared to NSLP eligibility, the share of economically disadvantaged students (31% for the cohort of 2021) is now higher than the share of students eligible for free or reduced-price meals a decade ago (26% for the cohort of 2010).

Students’ Testing Outcomes

Despite the rising share of higher-needs students, the failure rate for the science exit exam has declined dramatically over time, as depicted in Figure 5. For the graduation cohorts of 2019 and 2020, about 92% of test-takers fulfilled their science requirement on their first attempt. While fewer students have failed the science exam in recent years, a larger share of those who did fail also failed at least one other test. The group who failed *only* science comprised 2% of all first-time test-takers in those cohorts, down from almost 6% in the initial years of policy implementation. However, the on-time high-school graduation rate of these students has remained quite stable over time at 53%.

<Insert Figure 5 about here>

As the passing rate has increased, the demographics of students who initially score below the cutoff have undergone major shifts. Male students now make up 59% of those who fail, whereas in the initial cohorts, students who failed were evenly split by gender. Nearly 90% of students who fail now fall into the “economically disadvantaged” category for family income,

[8] Prior to 2015, the income measure was whether a student qualified for free or reduced-price school meals through the NSLP. Because of changes in that program’s operation, Massachusetts shifted to a new measure of economic disadvantage based on student's participation in one or more of the following state-administered programs: the Supplemental Nutrition Assistance Program (SNAP); the Transitional Assistance for Families with Dependent Children (TAFDC); the Department of Children and Families' (DCF) foster care program; and MassHealth (Medicaid). (Massachusetts Department of Elementary and Secondary Education, 2015)

compared to 82% of NSLP-eligible students in the earlier years. Additionally, most of those are ELs and/or students with disabilities; only 21% of those who fail on their first attempt carry neither designation.

However, within each of these demographic subgroups, the percentage of students who fail has declined over time, mirroring the overall trend in the state. The highest failure rate by far continues to be for ELs; 43% of recent EL test-takers fail the science test on their first attempt. In the earliest cohorts, though, 56% of EL test-takers failed. In comparison, the current failure rate is 20% for students with disabilities (down from 33%) and only 13% for students who are classified as economically disadvantaged (down from 23% of low-income students in 2010-12).

As in the initial cohorts, about 85% of students who fail their first science exam attempt to pass on retest, and just under half of them succeed on their first try. However, the proportion of those students who retest but never pass has risen from 19% in the early years to 27% in recent graduating cohorts. While this is perhaps not surprising, given that the passing cutoff now falls in the 8^{th} or 9^{th} percentile of the student performance distribution (depending on the subject test), the on-time high-school graduation rate for these students is only 28%.

As the passing rate has increased, the students who now score below the cutoff are those with low skills and, in most cases, special needs, including ELs and students with disabilities. In fact, ELs are the only subgroup for which both the count and the proportion of students who never pass science have increased over time. Of the approximately 6,500 ELs in the 2018-2020 graduating cohorts who failed on their first attempt, 31% retested but never scored above the passing threshold, compared to 23% for students with disabilities and 24% for students from low-income families. Large discrepancies in retest success continue to exist for ELs even after accounting for differences in performance on the initial test. Over 60% of the recent ELs who failed science also failed the math and/or ELA exams, meaning that their teachers had to remediate multiple subject areas.

Students' Interest in STEM

Increasing students' academic preparation for post-secondary science coursework and careers was one motivation behind the adoption of the science exit exam, and spurring students' interest in pursuing further study or training in STEM fields was another. Using data from the Massachusetts Department of Higher Education, we calculate the proportion of students enrolled in the state's public two and four-year colleges and universities who elected to major in one of these fields. We do this separately for students who took a 10th-grade exit exam while enrolled in a Massachusetts public high school (about 35% of whom went on to a Massachusetts public college) and for those who did not. The results, displayed in Figure 6, show an increase in STEM majors over the period from 2005-20 for all students, but a faster rate of growth among those students who had attended public high schools in Massachusetts. The trends appear to diverge beginning around 2010, which is when the students bound by the new science exit-exam requirement would have been expected to enroll in college. While this evidence is descriptive and other state initiatives were also attempting to address this issue, the timing is at least consistent with the idea that the new policy spurred student interest in STEM fields, as it was intended to do.

<Insert Figure 6 about here>

Schools' Responses to the Exit-Exam Policy

In addition to changes in testing outcomes and students' pursuit of STEM-related careers, there have been pronounced changes in the timing and subject-test choices made by students and schools. In the initial three cohorts of students, the majority (58%) took advantage of the option to test early in 9th grade, while 42% tested for the first time as 10th graders (Table 4).[9] In the 2018-20 cohorts, 76% of students tested in 9th grade. There has been a more modest shift to biology. Overall, in the early cohorts, 71% of first-time science test takers took biology and 24% took physics. Today, 76% of students take biology, and 21% physics. Very few students take the chemistry and technology/engineering tests. In fact, the state has announced plans to discontinue these two tests, with the last administration planned for spring 2023.

These changes reflect a shift in high-school course-taking sequences in the state. Perhaps as a reaction to the testing policy, more schools prepared students to take the 9th grade biology test. In recent cohorts, 53% of students took this test (compared to 32% in earlier cohorts). This growth largely came from a shift away from 10th grade biology. Those students who take the biology test as 10th graders are now substantially lower-performing, on average, on high-school entry; their mean score on both the 8th grade science and mathematics MCAS tests was -0.15 SD, compared to +.07 SD for students who tested in biology as 9th graders. At least some of these students attend schools with a three-semester biology sequence, which culminates in students taking the MCAS biology test in the winter or spring of their 10th grade year. Others are enrolled in the state's career/technical education (CTE) programs and spend half of the school day on core academic instruction, so that they finish a full-year biology course in two years instead of one.

These differences raise questions about how students sort into these test-taking groups. Our evidence indicates that decisions about the timing and subject choice for the MCAS science exam are often made at the school rather than the student level. About 92% of schools, serving

[9] In the discussion in this section, we exclude those students who entered MA public schools as 10th graders, since they could not have tested as 9th graders. For the combined 2010-12 graduating cohorts, there were 5,389 students in this category, or 2.6% of first-time test-takers in those years.

about 90% of students in the first cohort affected by the policy, had nearly all students test in one or two grade/subject groups (e.g., 9^{th} grade physics and 10^{th} grade biology). For example, the share of high schools in which nearly all students took the biology exam as 9^{th} graders rose from 15% to 20% in the first three years after the introduction of the science exam and reached 34% for the 2020 and 2021 graduating cohorts. Almost half of students who did not take the 9^{th} grade science exam attended schools in which fewer than 5% of 9^{th} graders did take an exam. In most cases, then, it appears that the choice of which test to take and when to take it was less about students’ discretion and more about the curricular sequence offered at their high schools.

We use administrative course-taking data for later cohorts to look at the next science course taken by students who fail their first science test. A surprising share of these students (17% overall) do not take a science course at all. For those who do, most take a course in a different subject than the one in which they had failed the test; their schools do not appear to interpret their poor performance as a sign they lack a basic proficiency in the subject. The exception is the students who failed the biology exam, about 40% of whom take another biology course the next year. In fact, half of all students who initially failed the science test end up taking biology the next year.

DISCUSSION

Massachusetts policymakers were motivated by a clear goal in adopting a high-school exit exam in science: increasing students’ preparation for and interest in postsecondary coursework and careers in STEM fields. While their theory of change was straightforward, they confronted several key decision points when it came to implementing the requirement. Some of their decisions reflected an unusual amount of flexibility and deference to local jurisdictions, such as the design of four different subject exams and the option for students to test initially in 9^{th} or 10^{th} grades. In other aspects, however, the Massachusetts policy design was more rigid: Massachusetts did not allow students to demonstrate their proficiency using any other test (e.g., an Advanced Placement or IB exam) and the passing threshold proved quite challenging for many students in the initial cohorts to meet.

The first few years of implementation followed a similar pattern to what had occurred earlier with the state's mathematics and ELA exams: an initially high failure rate that dropped steadily over the first several years. However, the rapid decrease in the overall failure rate could reflect increases in students' science knowledge and skills due to improved instruction, higher student motivation, and greater emphasis on science in the state's high schools, all of which were explicit goals of the policy. But it could also reflect increased familiarity with the examination or other sources of score inflation, which refers to gains on a high-stakes assessment that do not transfer to other tests of the same content. As educators gained familiarity with the different subject tests, they likely developed strategies to help students near the cutoff perform better on them. With the available data, therefore, we cannot draw firm conclusions about the system-wide consequences of the introduction of the science requirement.

However, the test does serve as a measure of both improved outcomes and rising inequity. Failure rates for every key demographic subgroup declined over the years, consistent with the overall pattern in the state. However, disparities by gender, family income, and especially English proficiency widened over time. Students from low-income families were less likely to pass the science exam, even compared to students from higher-income backgrounds with the same level of academic performance on high-school entry. The subgroup of English learners continues to have the highest failure rates on their first attempt by far, and ELs are less likely to pass on retest than other students who scored at the same level on their initial test. Of course, disparate performance likely reflects underlying inequalities in preparation and instructional quality among different groups. This suggests a critical need for improvements in high-school instruction and access to opportunities for students from vulnerable subgroups.

For the initial cohorts of science test-takers, we have enough longitudinal data to shed light on the causal impact of exit-exam performance on the later outcomes of students scoring near the passing threshold. We interpret our findings that barely passing increased on-time high-school graduation (particularly for low-income students) and college outcomes (for higher-income students) as an unintended consequence of the new policy.

However, the largest impact of barely passing the exam was on high-school graduation for female test-takers near the cutoff. As with any regression-discontinuity design, it is difficult to distinguish between the encouragement effect of barely passing and the discouragement effect of barely failing. It could be that females who just passed the exam on their first attempt were encouraged by their performance; if this translated into increased motivation and perseverance as they continued through high school, it could explain their greater success in graduating. Alternatively, the experience of just failing the exam might discourage female students on the failing side of the cutoff, either through the imposition of an obstacle to graduating or by negatively impacting their effort and confidence in their remaining high-school courses (or both).

To shed light on this question, we follow the approach from Papay et al (2016) and use female students' 8th grade science test performance to differentiate between those who had passed or failed as middle-schoolers. In results available on request, we find that the high-school graduation impacts are concentrated in the group of female students who had failed the science exam as 8th graders. Among these students, barely passing in high school on their first attempt increased the probability of on-time graduation by 12 percentage points and the probability of five-year graduation by 9 points. We interpret this as suggestive evidence of an encouragement effect. These students had all failed in middle school and may have expected to fail again. By contrast, we find no evidence of impacts for females who had passed in 8^{th} grade.

Again, it is important to note that our estimates of these effects apply only to students near the passing cutoff in the initial years of the exit exam. Over time, as the pool of students who failed shrank dramatically, those who are left below the cutoff are students with the lowest academic skills on high-school entry, most of whom are ELs, have a diagnosed disability, or both. The majority of these students also fail one or both of the other exit exams, meaning that they face barriers to high-school graduation beyond the science requirement. We do not yet have data to investigate causal impacts on later outcomes for these more recent cohorts of students.

In addition to its impacts on students, the science exit exam also proved consequential for the scope and sequence of public high-school science instruction in Massachusetts, as schools responded to the new requirement. Despite the state's flexibility, over 95% of students now test in one of three groups: 9^{th} and 10^{th} grade biology and 9^{th} grade physics. The investment in developing the chemistry and technology/engineering exams never paid off in terms of student participation in those tests. Moreover, the challenges with equating tests of non-overlapping domains meant that the passing thresholds for different tests have been in different parts of the performance distribution. The exit exam may also have contributed to the standardization of high-school course sequences, particularly in 9^{th} and 10^{th} grades. In recent years, more than three-quarters of students test for the first time as 9^{th} graders, and the proportion of first-time test-takers in 9^{th} grade biology has steadily grown over time. This shift was perhaps motivated in part by the decision to offer a February testing opportunity in biology only; beginning in 2023, the state will add a similar administration in physics.

Almost 15 years after the rollout of the science exit exam, the state continues to adjust the details: eliminating the chemistry and technology/engineering exams, expanding testing opportunities in physics, adding Spanish and American Sign Language (ASL) versions of the biology and physics assessments, and transitioning to computer-based tests aligned to a new set of science standards. Our work, which has benefited from the availability of rich longitudinal administrative data, sheds light on the complexities of the policy's implementation and its consequences, in some cases unexpected, for students and schools. Evidence from Massachusetts suggests that a high-school exit exam can be an effective tool in increasing students' achievement and interest in science, but the equity implications of such a policy need to be carefully attended to.

REFERENCES

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. Educational Researcher, 36(5): 258-267.

Baker, O., & Lang, K. (2013). The effect of high school exit exams on graduation, employment, wages and incarceration. (Working Paper 19182). National Bureau of Economic Research.

Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. American Educational Research Journal, 42(2), 231–268. https://doi.org/10.3102/00028312042002231

Caves, K. & Balestra, S. (2018). The impact of high school exit exams on graduation rates and achievement, The Journal of Educational Research, 111(2), 186-200

Clark, D., & Martorell, P. (2014). The signaling value of a high school diploma. Journal of Political Economy, 122(2), 282–318.

Conaway, C. (2007). Supply and demand of STEM workers: STEM jobs are growing, but are enough Massachusetts students qualified? Education Research Brief, Massachusetts Department of Elementary and Secondary Education. https://files.eric.ed.gov/fulltext/ED508405.pdf.

Dee, T. S., & Jacob, B. A. (2006). Do high school exit exams influence educational attainment or labor market performance? (Working paper 12199). National Bureau of Economic Research. https://www.nber.org/papers/w12199.

Dee, T. S., Jacob, B., & Schwartz, N.L. (2013). The effects of NCLB on school resources and practices. Educational Evaluation & Policy Analysis, 35(2), 252-279.

Executive Office of Education, Commonwealth of Massachusetts. (2008). Ready for 21st century success: The new promise of public education. https://archives.lib.state.ma.us/handle/2452/36539?show=full.

Gewertz, C. (n.d.). Which states require an exam to graduate? Education Week. Retrieved May 26, 2022 from https://www.edweek.org/teaching-learning/which-states-require-an-exam-to-graduate.

Grigg, W., Lauko, M., and Brockway, D. (2006). The Nation's Report Card: Science 2005 (NCES 2006-466). U.S. Department of Education, National Center for Education Statistics. Washington, D.C.: U.S. Government Printing Office.

Hemelt, S. W., & Marcotte, D. E. (2013). High school exit exams and dropout in an era of increased accountability. Journal of Policy Analysis and Management, 32(2).

Holme, J. J. (2008). High stakes diplomas: Organizational responses of California's high schools to the state's exit examination requirement. Research in the Sociology of Education, 14, 157–188

Institute of Medicine. (2007). Rising above the gathering storm: Energizing and employing America for a brighter economic future. Washington, DC: The National Academies Press. https://doi.org/10.17226/11463.

Massachusetts Department of Elementary and Secondary Education. (2015). A changing metric: Low income vs. economically disadvantaged. https://www.doe.mass.edu/infoservices/data/ChangingMetric.pptx.

Massachusetts Department of Elementary and Secondary Education. (2020, October 16). Overview of the MCAS performance appeals process. https://www.doe.mass.edu/mcasappeals/.

Murnane, R. & Papay, J.P. (2010). Teachers' views on No Child Left Behind: Support for the principles, concerns about the practices. Journal of Economic Perspectives, 24(3), 151–166.

Neal, D. & Schanzenbach, D.W. (2010). Left behind by design: Proficiency counts and test-based accountability. The Review of Economics and Statistics, 92(2), 263–283.

Organization for Economic Cooperation and Development. (2007). PISA 2006: Science competencies for tomorrow's world. https://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/pisa2006results.htm#Vol_1_and_2.

Ou, D. (2010). To leave or not to leave? A regression discontinuity analysis of the impact of failing the high school exit exam. Economics of Education Review, 29(2), 171–186.

Papay, J.P., Mantil, A., Murnane, R.J., An, L., Donohue, K., & McDonough, A. (2020). Lifting all boats? Accomplishments and challenges from 20 years of education reform in Massachusetts. Providence, RI: Educational Opportunity in MA, Brown University.

Papay, J.P., Mantil, A., & Murnane, R.J. (2022). On the threshold: Impacts of barely passing high-school exit exams on post-secondary enrollment and completion. Educational Evaluation and Policy Analysis. https://doi.org/10.3102/01623737221090258.

Papay, J.P., Murnane, R.J., & Willett, J.B. (2016). The impact of test-score labels on human-capital investment decisions. Journal of Human Resources, 51(2), 357-388.

Papay, J.P., Murnane, R.J., & Willett, J.B. (2015). Income-based inequality in educational outcomes: Learning from state longitudinal data systems. Educational Evaluation and Policy Analysis, 37(1S), 29S-52S.

Papay, J.P., Murnane, R.J., & Willett, J.B. (2014). High-school exit examinations and the schooling decisions of teenagers: A multi-dimensional regression-discontinuity analysis. Journal of Research on Educational Effectiveness, 7(1), 1-27.

Papay, J.P., Murnane, R.J., & Willett, J.B. (2010). The consequences of high school exit examinations for low-performing urban students: Evidence from Massachusetts. Educational Evaluation and Policy Analysis, 32(1), 5-23.

Polson, C. (2018). TAKS-ing students? Texas exit exam effects on human capital formation. Economics Of Education Review, 62, 129-150.

Reardon, S.F., Arshan, N., Atteberry, A., & Kurlaender, M. (2010). Effects of failing a high school exit exam on course taking, achievement, persistence, and graduation. Educational Evaluation and Policy Analysis, 32, 498–520.

Rennie Center for Research and Education Policy. (2008). Opportunity to Learn audit: High school science. https://www.renniecenter.org/research/reports/opportunity-learn-audit-high-school-science.

Sacchetti, M. (2005a, January 26). Romney presses for MCAS science hurdle. Boston Globe, 22.

Sacchetti, M. (2005b, March 30). State closer to requiring students to pass science MCAS. Boston Globe, 20.

UMass Donahue Institute, Research and Evaluation Group. (2009). Massachusetts Statewide STEM Indicators Project (MASSIP) 2009 Report. https://www.mass.edu/stem/documents/MASSIP%202009%20Full%20Report.pdf.

Warren, J. R., Grodsky, E. S., & Lee, J. C. (2/008). State high school exit examinations and post-secondary labor market outcomes. *Sociology of Education*, 81, 77–107.

TABLES

Table 1

Student failure rates and retest behavior on the high-school science in the expected graduation cohorts of 2010-12, by grade and subject test

	Number failing	Proportion among test-takers who failed	Proportion failing	Retested in Same Subject	Retested in Different Subject	Did Not Retest
9th grade	15,962	0.623	0.135	0.413	0.433	0.154
10th grade	9,641	0.377	0.106	0.756	0.098	0.146
Biology	16,331	0.638	0.109	0.786	0.067	0.147
Chemistry	864	0.034	0.205	0.111	0.763	0.126
Physics	7,813	0.305	0.156	0.109	0.730	0.161
Tech/Engineering	595	0.023	0.106	0.165	0.681	0.155

Table 2

Estimated causal effects of passing the high-school exit examination in science, as opposed to failing it, on the probability of selected high-school and college outcomes for students at the margin of passing, for all students and by gender, family income, and English proficiency

Student group	On-time high-school graduation (MCAS+2) h*=2	Five-year high-school graduation (MCAS+3) h*=3	Enrollment in any college (MCAS+4) h*=2	Graduation from any college (MCAS+9) h*=2	Graduation from 4-yr college (MCAS+9) h*=2
All students	0.0368**	0.0129+	0.0240**	0.0136*	0.00207
	(0.00517)	(0.00593)	(0.00268)	(0.00379)	(0.00201)
Female	0.0734**	0.0397**	0.0126+	-0.00016	0.00017
	(0.00111)	(0.0103)	(0.00528)	(0.00160)	(0.0008)
Male	-0.00402	-0.0169**	0.0362**	0.0284*	0.00388
	(0.0104)	(0.0034)	(0.00117)	(0.00649)	(0.00483)
Low-income	0.0423**	0.0144	0.0116*	0.00899+	-0.00119
	(0.00845)	(0.0105)	(0.00295)	(0.00419)	(0.0026)
Higher-income	0.0189**	0.00763	0.0619**	0.0267**	0.0125**
	(0.00309)	(0.00946)	(0.00275)	(0.00218)	(0.00056)
English language learners	-0.00394	-0.00387	-0.0494	-0.0216+	-0.0148
	(0.0143)	(0.0259)	(0.0242)	(0.0078)	(0.00744)
Proficient in English	0.0413**	0.0148	0.0323**	0.0167*	0.00321
	(0.00691)	(0.00829)	(0.00122)	(0.00481)	(0.0016)
N	17102	23874	17102	17102	17102

Notes. MCAS is the Massachusetts Comprehensive Assessment System high-school science test; (MCAS+2) indicates that the outcome was measured two years after 10th-grade students took the test for the first time (or three years for students who initially tested in 9th grade). Standard errors clustered on raw score point are in parentheses; *h** indicates the optimal bandwidth used in the regression-discontinuity model for each outcome.
** *p*<0.01, * *p*<0.05, + *p*<0.1

Table 3

Initial test performance of students in the cohorts of 2010-12 who took the high-school science exam and retest behavior of those students who failed, by demographic subgroup

	Initial test (all students)			*Retest & graduation (students who failed)*			
	Proportion among all test-takers (n=209544)	Proportion among test-takers who failed (n=25603)	Proportion who failed	Ever retested	Passed on first retest	Never passed	HS Graduation Rate
Female	0.495	0.494	0.122	0.862	0.432	0.152	0.581
Male	0.505	0.506	0.122	0.837	0.404	0.17	0.476
Low-income	0.427	0.816	0.234	0.846	0.387	0.18	0.492
Higher-income	0.573	0.184	0.039	0.863	0.556	0.082	0.688
English Language Learners	0.036	0.165	0.56	0.83	0.293	0.237	0.458
Students with disabilities	0.148	0.404	0.334	0.868	0.37	0.203	0.525
Urban	0.267	0.564	0.258	0.843	0.358	0.194	0.476
Asian	0.05	0.038	0.093	0.871	0.408	0.163	0.563
Black	0.083	0.191	0.282	0.863	0.384	0.181	0.538
Hispanic	0.123	0.346	0.344	0.84	0.344	0.201	0.46
Multiple race/ethnicity	0.015	0.016	0.13	0.831	0.395	0.152	0.519
Native American	0.003	0.004	0.154	0.844	0.411	0.144	0.500
White	0.727	0.405	0.068	0.849	0.499	0.119	0.579
All students	1.00	1.00	0.122	0.849	0.418	0.162	0.528

Note. The low-income group includes all students who had qualified for free or reduced-price lunch at some point during their enrollment in the Massachusetts public-school system.

FIGURES

Figure 1

Predicted probability of passing high-school science exam on first attempt, conditional on 8th grade science scores, by selected grades and subject tests (left panel) and by subject test (right panel)

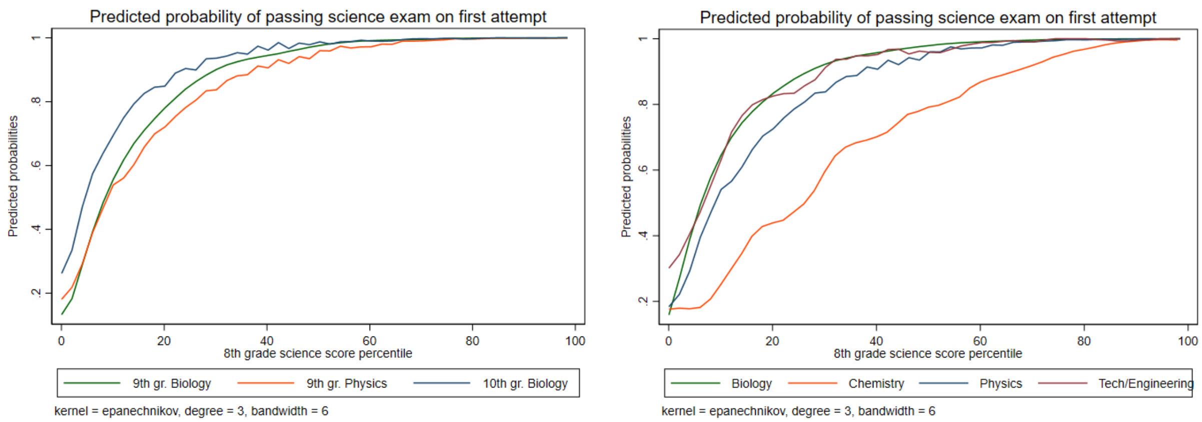
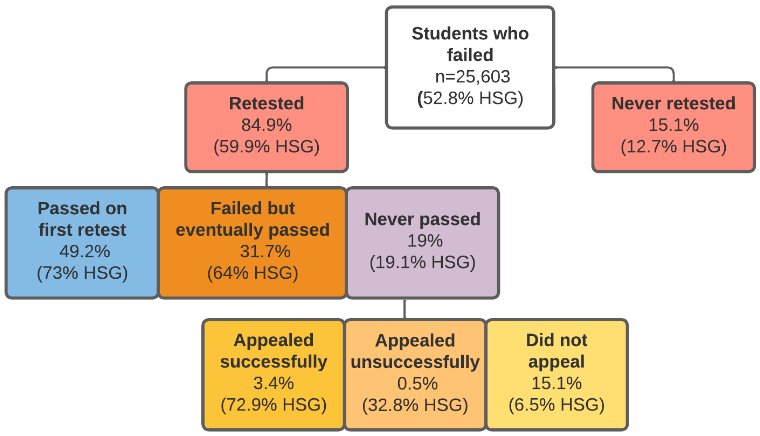


Figure 2

Retest and appeal behavior and success rates for students who initially failed the high-school science exam in the expected high-school graduation cohorts of 2010-12



Notes. HSG = high-school graduation rate within 2 years of taking the science test (10th grade test-takers) or 3 years of taking the science test (9th grade test-takers)

Figure 3

Sample mean probabilities of on-time high-school graduation by gender (top panel) and college graduation by family income (bottom panel) at score points near the passing threshold on the science high-school exit exam

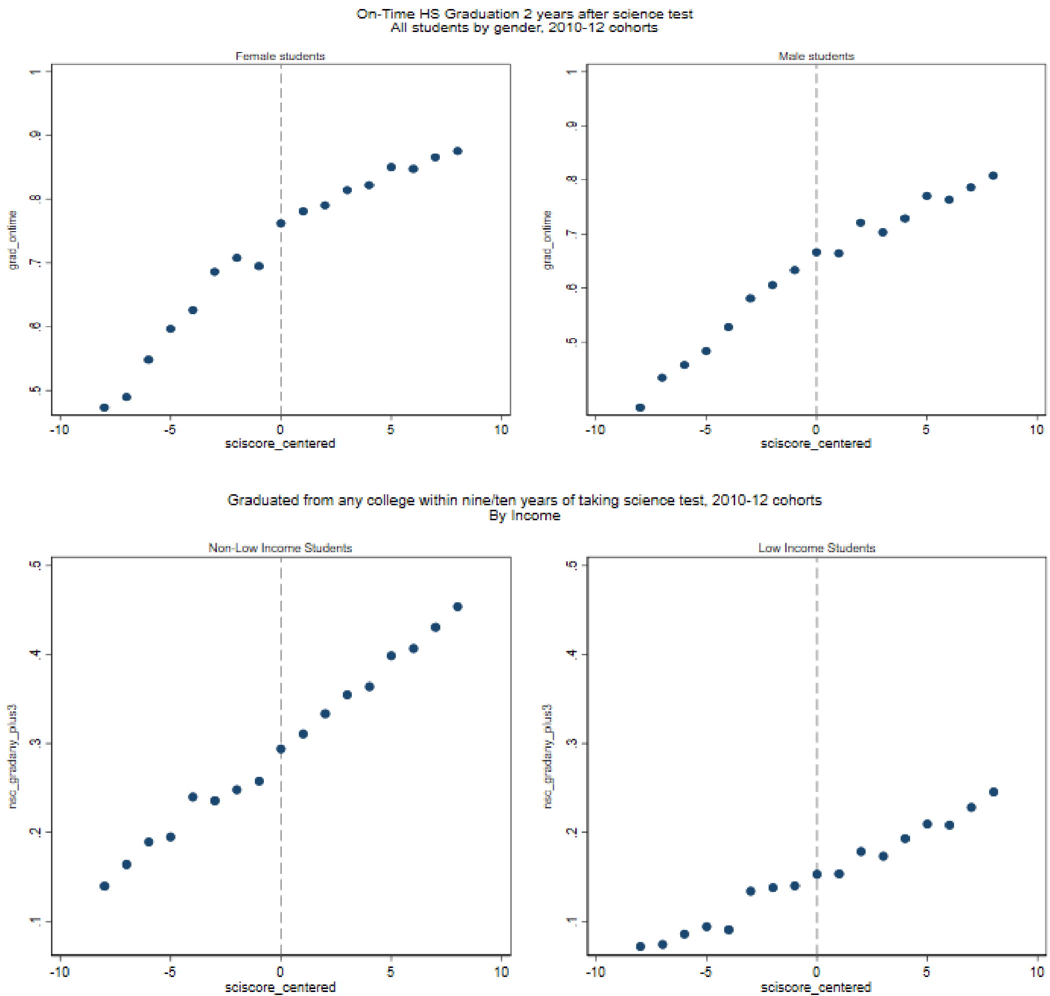
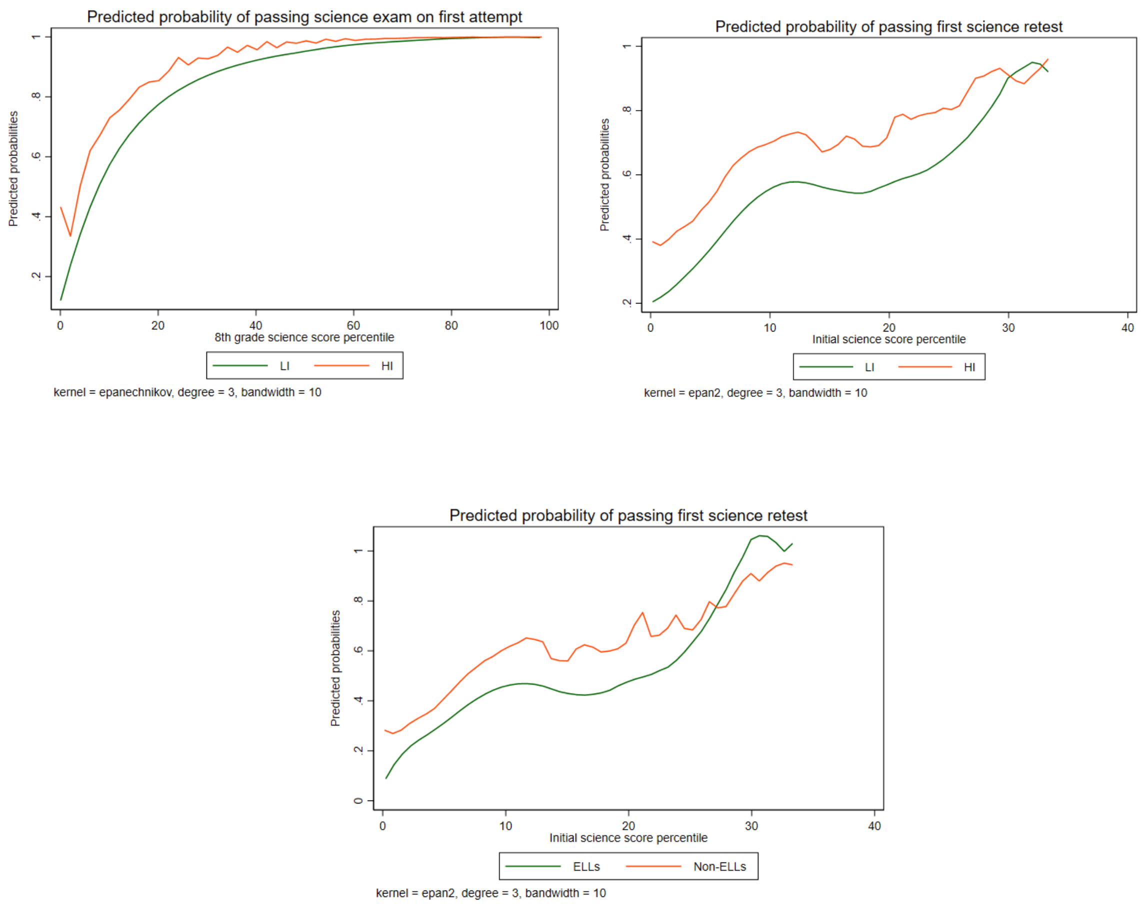


Figure 4

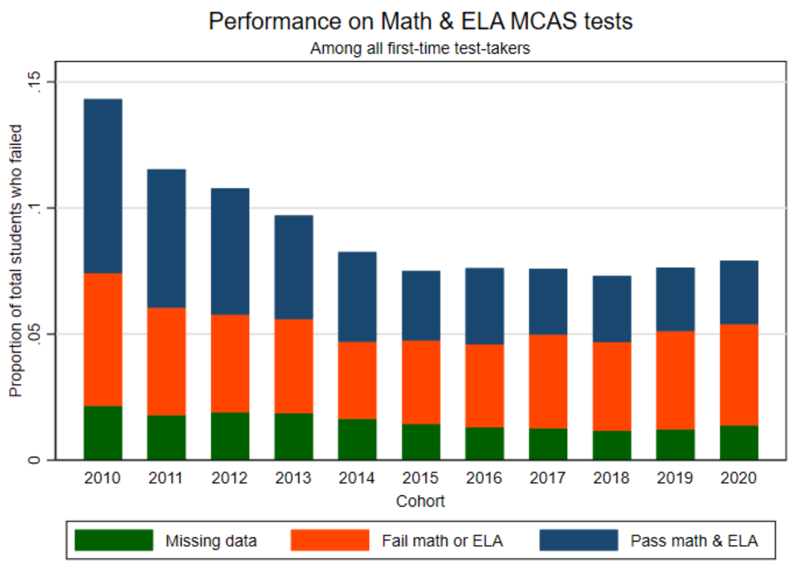
Predicted probability of passing high-school science exam on first attempt, conditional on 8th grade science scores, and on passing first retest, conditional on initial high-school score, by income (top panel) and English language proficiency (bottom panel)



Note: Plot of the predicted probability of passing the science exam on the first attempt is omitted for English Learners because 42% of these students are missing on 8th grade science scores.

Figure 5

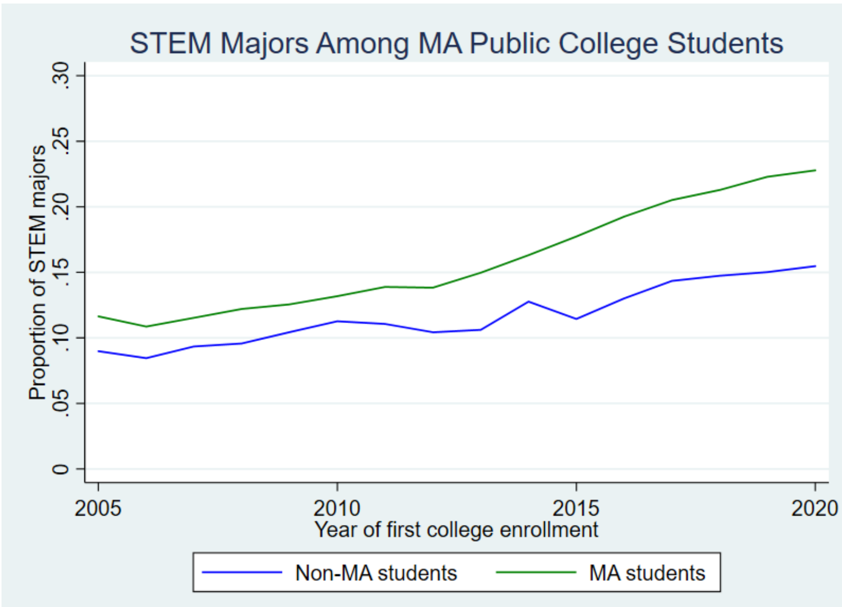
Failure rates for first-time Massachusetts high-school science takers over time, by passing status on the other two MCAS tests



Note: Students with missing data are those who took the science exam but not math and/or ELA.

Figure 6

Proportion of science, technology, engineering, and mathematics (STEM) majors among students enrolled in Massachusetts public colleges and universities over time, by where they attended high school



Notes: The sample for the figure includes students who first enrolled in a Massachusetts public institution of higher education in the fall term of each academic year (2005 denotes the 2004-05 academic year). "MA students" are those who took the Massachusetts high-school exit examination as 10th graders; "Non-MA students" are those who did not.