EdWorkingPaper No. 24-645

Introducing a High-School Exit Exam in Science: Consequences in Massachusetts

Ann Mantil
Brown University

John P. Papay
Brown University

Preeya Pandya Mbekeani
American Institutes for Research

Richard J. Murnane
Harvard University

Preparing students for science, technology, and engineering careers is an urgent state policy challenge. We examine the design and roll-out of a science testing requirement for high-school graduation in Massachusetts. While science test performance has improved over time for all demographic subgroups, we observe rising inequality in failure rates and retest success. English learners, almost 8% of all test-takers, account for 53% of students who never pass. We find large differences by family income, even conditional on previous test scores, that raise equity implications. Using a regression-discontinuity design, we show that barely passing the exam increases high-school graduation and college outcomes of students near the score threshold, particularly for females and students from higher-income families.

VERSION: July 2025

Suggested citation: Mantil, Ann, John Papay, Preeya Pandya Mbekeani, and Richard J. Murnane. (2025). Introducing a High-School Exit Exam in Science: Consequences in Massachusetts. (EdWorkingPaper: 22-645). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/9e56-3074

Abstract

Preparing students for science, technology, and engineering careers is an urgent state policy challenge. We examine the design and roll-out of a science testing requirement for high-school graduation in Massachusetts. While science test performance has improved over time for all demographic subgroups, we observe rising inequality in failure rates and retest success. English learners, almost 8% of all test-takers, account for 53% of students who never pass. We find large differences by family income, even conditional on previous test scores, that raise equity implications. Using a regression-discontinuity design, we show that barely passing the exam increases high-school graduation and college outcomes of students near the score threshold, particularly for females and students from higher-income families.

Keywords: exit exams, high school, high-stakes testing, K-12 educational policy, science

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 evidence on the science skills and knowledge of American middle- and high-school students was prompting serious concern. For example, 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). While the most common response of state policymakers to this problem has been to increase science high-school course requirements, several states instituted a science exit examination that students must pass in order to receive a diploma.

Our study of the consequences of introducing such a policy sits at the intersection of the established literature on exit exams and the less developed research base on state initiatives to support science skill development. In the years after the exit exam was introduced, science test scores in Massachusetts increased substantially, and districts shifted their curricular offerings to align with the test. While we cannot estimate the overall impact of introducing the science exam on student outcomes, due to the missing counterfactual of what science preparation would have looked like in the absence of the test, we analyze effects on the margins. We show that barely passing the test increases educational attainments compared to barely failing. However, this pattern does not preclude the possibility that the exit exam's impact is positive across the board, just more so for some students than for others.

Consistent with the extensive literature documenting that female students report lower levels of interest in science and in pursuing careers in STEM fields than males of similar achievement (Desy et al., 2011; Fredricks & Eccles, 2002; Riegle-Crumb et al., 2011; Sadler et al., 2012; Wang & Degol, 2017), we find a differential impact by gender. Females who barely passed the science test were more likely to graduate from high school on time than females who barely failed. We see no corresponding effect for males. This finding has important implications for our understanding of science skill development among female students. Also, our study is one of very few to look at impacts on college graduation 9-10 years after students take the exit exam. Similar to Papay et al. (2022), we find positive effects of barely passing the exam of about three percentage points for students from higher-income families, but not for students from low-income families, on this crucial outcome.

The present study is also the first that we know of to describe the many design issues inherent in instituting a state-wide science exam and the responses it elicited from district and school leaders. We show that the introduction of the science exam had heterogenous impacts across groups of students, posing especially great challenges for English learners (ELs). In the graduating cohorts of 2018 to 2020, ELs comprise just under 8% of all test-takers but 41% of the students who fail on their first attempt and 53% of those who never pass.

After reviewing previous research on high-school exit exams and presenting the policy context in Massachusetts, we address three main research questions:

(1) How did the initial cohorts of test-takers fare on the science exit exam?

(2) For the subset of students who scored near the passing threshold in the early years of the test, how did barely passing (as opposed to barely failing) affect students' high-school graduation and college outcomes?

(3) How have students' testing outcomes and schools' responses to the exit exam changed over time?

Background and Context

Exit Exams

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; Caves & Ballestra, 2018; Dee & Jacob, 2006; Hemelt & Marcotte, 2013; Holme et al., 2010; Warren et al., 2008). These studies seek to explore the overall impact of introducing the policy on educational outcomes. Most research has focused on high-school graduation, but the few studies 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 ELA on students who score near the cutoff

(Clark & Martorell, 2014; Ou, 2010; Papay et al., 2010; Papay et al., 2014; Papay et al., 2022; Polson, 2018; Reardon et al., 2010); we adopt this approach to examine impacts in science. This prior work on math and ELA exams largely finds that barely passing (as opposed to failing) increases the probability of high-school graduation, particularly for students from low-income families. Papay et al. (2022) was the first study to estimate impacts on college graduation and found positive effects of passing for students from higher-income families, but not those from low-income families.

How schools respond to exit-exam requirements has received less attention and study. In studies of high-stakes testing in elementary grades, 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 (Booher-Jennings, 2005; Neal & Schanzenbach, 2010). Vasquez Heilig & Darling-Hammond (2008) found evidence of selective retention in grade, skipping over the tested grade, and encouraging dropout in urban Texas high schools experiencing exit-exam accountability pressures. The advent of test-based accountability and the Common Core curriculum, both focused heavily on mathematics and English/language arts (ELA), shifted resources and attention away from untested subjects like science (Arold & Shakeel, 2001; Au, 2007; Dee et al., 2013; Holme, 2008; Murnane & Papay, 2010). Vogler (2008) showed that a high-stakes exit exam in history elicited more instructional time devoted to test preparation and practice with the test format than a similar statewide exam with lower stakes.

State Policies to Promote Science Mastery

The wide-ranging literature on science education at the school and classroom levels (see, for example, Lederman & Abell, 2014; Potvin & Hasni, 2014; Xie et al., 2015) includes limited evidence about the effectiveness of state-level policies targeting high-school students' mastery of science. While an exit exam is one policy option, the more common strategy is to institute or strengthen state-level mandatory course requirements in science. Massachusetts is one of only four states that does not currently require a minimum number of science courses or credits in order to graduate (Erwin et al., 2023). However, the available evidence suggests that mandatory coursework has very limited impacts on students' mastery of STEM content. For example, using data from the National Education Longitudinal Study of 1988, Teitelbaum (2003) found three responses to policies increasing the number of science courses high school students must pass. First, compliance was incomplete, with some students graduating without fulfilling the science course requirement. Second, on average, students did earn more credits in science. Third, however, there was no increase in scores on science exams.

In a study based on data from the Chicago Public Schools, Montgomery et al. (2010) examined the effects of a policy change increasing from one to three the number of science courses high-school students must pass in order to graduate. Similar to Teitelbaum (2003), they found that the new requirement did increase the number of science courses students completed but did not increase science learning or college outcomes. More recent evidence from California showed that more rigorous science course requirements did not have any effect on students' proficiency on the state's science assessment (Gao, 2021).

Mandatory coursework in science, then, appears to lack potency as a catalyst for improved student preparation for postsecondary study and careers in STEM fields. An exit exam, as adopted by Massachusetts, represents a different approach. The state sets standards for what students are expected to learn and requires a demonstration of proficiency, in the form of passing an exit exam, in order to receive a high school diploma. The expectation is that this will induce improvements in curricular and instructional rigor and in students' motivation to learn the material. In the 2018-19 school year, ten states had a high-school exit exam in science (Gewertz, n.d.). This paper describes how Massachusetts addressed the many questions involved in the design of an exit exam in science and explains what happened afterwards, as educators and students responded to the new requirement.

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 passing the 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 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. The state's strong performance and highly rated science standards, which were the first in the nation to incorporate technology and engineering, seemed to position its 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 (Conaway, 2007). Concern over the interest level in STEM among students graduating from the state's high schools was accompanied by rising alarm over the adequacy of their preparation for college-level work.

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 (P. Reville, personal communication, March 16, 2022; Sacchetti, 2005b). 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 piece 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).

Design of the Science Exit Exam

The Massachusetts science exit-exam policy is more nuanced and complex than those in most other states with similar requirements. In Massachusetts, the June science exams are offered in four content areas: introductory physics, biology, chemistry, and technology/engineering. Each exam has different forms, all of which include a set of common items (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.[2] Students receive a score report with their scaled score, which they can easily compare to the state's passing threshold. 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; 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 U.S. Department of Education allowed the state to administer the test in 9th grade provided that it assessed 10th-grade standards (J. Nellhaus, personal communication, March 25, 2022). This flexibility was offered for multiple reasons: (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 same 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 have their score "banked" as their official 10th grade score if they passed. 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 of 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.[3] 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 never pass the science exam; 18% of students who initially fail fall into this category. Districts have the option to submit a cohort or portfolio appeal for such students (Massachusetts DESE, 2020).[4] 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 to 5% of students who fail the exam on their first attempt successfully fulfill the requirement via an appeal.

Data and 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 include information on students' high-school course-taking. College enrollment and graduation data come from the National Student Clearinghouse (NSC), which matches to state records using names and dates of birth. These NSC data include student information from 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 ($52,380 in 2023 dollars). Importantly, students whom we identify as "higher income" are quite heterogeneous: they come from families whose incomes fell just above this threshold as well as those who were quite wealthy. Other demographic indicators, including student grade level, 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 a 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.[5] Essentially, we compare outcomes of students who barely passed the exit exam to those of students whose scores were just below the passing threshold.

We focus on several main outcomes. 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 state's test requirements, as well as other local graduation requirements. College outcomes include indicators for whether the student enrolled within four years of the test (or five years if the test was first taken in 9th grade) and whether the student graduated from college within nine (or 10) years. Finally, we estimate impacts on students' science course-taking behavior in high school, namely the number of full-year equivalent (FTE) science courses students passed after taking their first exam.

We conceptualize the treatment as barely passing the science examination on the first attempt, although we could equivalently define this as barely failing the test. The forcing variable is a student's raw score from their first attempt, 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, other than the difference explained by their test score. 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 effect of barely passing, as compared to barely failing, the science exam. This estimate only applies to students who score quite near the cutoff on their first attempt.

We use estimating equations of the following form, using $COLLGRAD_i$ as the outcome variable in this example:

$$p(COLLGRAD_i = 1) = \beta_0 + \beta_1 SCI_i^C + \beta_2 PASS\_SCI_i + \beta_3 SCI_i^C \times PASS\_SCI_i + \boldsymbol{\gamma}' \, \boldsymbol{X}_i + \varepsilon_i$$

for student i, where SCI^C is the centered raw science score and *PASS_SCI* is an indicator for passing status. The causal effect of interest is β_2, which represents the difference in the probabilities of graduating from a four-year college for students who just passed the science exam and otherwise similar students who just failed. If its estimated value is statistically significant and positive, we can conclude that barely passing causes the student's probability of graduating from college to *increase* discontinuously, on average, in the population. We discuss the substantive implications of this interpretation below.

We calculate optimal bandwidths (h*) using the cross-validation procedure described by Imbens & Lemieux (2008)[6] for each outcome of interest and test the robustness of the results to bandwidth selection (see Online Technical Appendix). 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, (X_i), including the demographic variables listed above 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 use the full sample and interact the passing and score variables with the relevant demographic indicator of interest (e.g., gender, family income).

Initial Test Performance and Impacts

To analyze the roll-out of the science requirement and its initial consequences for students, 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 roughly 70,000 per year.

In the following sections, we describe overall patterns in test performance and success on retest, the impacts of passing on students scoring near the cutoff, and evidence of heterogeneity in impacts by demographic subgroup.

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, as shown in Table 1. 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.[7] 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 (math and ELA).

<Insert Table 1 about here>

We see differences in performance by the timing and choice of test in the initial three cohorts of science test-takers, as shown in Table 2. 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 2 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.[8] These three groups included over 94% of test-takers in the initial years of the exam. Focusing on the lower half of the distribution, 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. Most students tested in 9th grade, probably with the goal of spreading the required exit exams out over two years. However, 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 the 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. However, the decision meant that students confronted different challenges in achieving the proficiency standard.

The subject test students initially took also influenced their subsequent retest behavior, as shown in Table 2. Among biology test-takers who failed and then retested, 92% took the biology test again. But students who failed 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 only in biology. 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 had already left the Massachusetts public-school system.

Figure 2 depicts the different pathways 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 initially taking the science test (for first-time 10th grade test-takers) or three years (for 9th grade test-takers).[9] The overall on-time graduation rate for these cohorts was 85%, but for students who initially failed a science test, the rate was only 53%.

<Insert Figure 2 about here>

Most students (85%) who initially failed science did retest. Almost half passed on their first retest attempt, while another third needed multiple retest attempts but eventually did achieve a passing score. Among students who initially 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 students; districts filed an appeal for just over 3% of all students who initially failed and retested, as shown in Figure 2. Most appeal attempts met with success, and students who were granted an appeal graduated from high school at very similar rates as their peers who passed on their first retest.

Causal Impacts for Students Near the Cutoff

Using a regression-discontinuity approach, we estimate how test performance affected later educational attainments for students scoring near the passing threshold. As shown in the first row of Table 3, barely passing increased both on-time high-school graduation and graduation in five years for the first three cohorts affected by the science requirement. 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.[10]

<Insert Table 3 about here>

There is also evidence of an overall impact on college outcomes for students near the cutoff. The point estimates for enrolling and graduating from any college are statistically significant in Table 3 at both the optimal bandwidth of h*=2 and at h=3. We conclude that for students who scored near the passing threshold on their first attempt, passing the science exam increased the probability of college graduation by one to two percentage points, on average.

Heterogeneity by Subgroup

A key consideration in assessing the consequences of introducing the science exam is whether there are differences in failure rates, retest participation and success, and impacts 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 also investigated differences by race/ethnicity, urbanicity, and disability status, but did not find clear-cut patterns using those indicators, particularly compared to the three detailed below.

Gender

In the first three cohorts of students subject to the science test requirement, first-time test-takers were evenly divided by gender. As seen in Table 1, there was no difference between male and female students in the probability of passing on the first attempt. Among the 12% who failed, 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, with a difference of 10 percentage points favoring females.

We find clear evidence of heterogeneous effects by gender in our regression-discontinuity analyses of the impacts of barely passing vs. barely failing the exam. As shown in Table 3, 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 ($p < 0.001$).

In the top panel of Figure 3, we plot the sample mean probabilities of graduation for female and male students at each test score point, centered on the cutoff. 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 3 to bandwidth selection strongly support the inference of an impact on the probability of high-school graduation for female students, but not for 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 (see Figure A-4, for example) contain no clear evidence of discontinuities at the cutoff.

<Insert Figure 3 about here>

Family Income

About 43% of science test-takers in the first three cohorts were from low-income families, as shown in Table 1. 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 8^{th} 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 8^{th} 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 8^{th} grade passed the high-school science test the first time that they took it. But among those scoring below the state median as 8^{th} graders, students from low-income families were less likely to pass than those from more advantaged backgrounds with the same 8^{th} 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 may be contributing factors.

Among those who failed a science exam, retest participation was quite similar among low-income and higher-income students (Table 1), 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 are pronounced differences by family income in the predicted probability of passing the first retest even 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, on average, students from low-income families received less preparation and support for retesting than did students from higher-income families.

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 rates 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 from *both* low-income and higher-income families. The coefficients for the two groups reported in Table 3 do not differ significantly (p=0.113), and visual discontinuities are apparent in both plots of the sample mean probabilities of graduation for students around 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, 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.

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 1.

English learners also experienced markedly less success on retest than other groups. While their rate of retest participation was quite similar to that of 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 1. In results not shown, we find that the intersection of limited English language skills 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 suggests no impacts of barely passing the exam on the later educational attainments of ELs (Table 3). 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.

Differences Across Cohorts

Since the introduction of the science exit-exam requirement, Massachusetts has experienced substantial demographic shifts in its public-school student population (Mantil et al., 2022; 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.[11] The new "economic disadvantage" measure captures students up to a lower household income threshold than did NSLP eligibility and is therefore not directly comparable. 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 students from historically disadvantaged groups, the failure rate on 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 passed the science exam 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 that 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, despite the fact that these students fall further in the tail of the performance distribution, the on-time high-school graduation rate of these students has remained quite stable across cohorts 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 make up 59% of those who fail in recent cohorts, whereas initially, students who failed were evenly split by gender. Nearly 90% of students who fail fall into the "economically disadvantaged" category for family income, compared to 82% of NSLP-eligible students in the earlier years. Additionally, most students who fail are ELs and/or students with disabilities; only 21% of those who fail on their first attempt carry neither designation.

Within each of these demographic subgroups, though, 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 failed the science test on their first attempt. In the earliest cohorts, 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). Economically disadvantaged students in recent years fail the test at the same rate as that of all students in the 2010-12 cohorts.

As in the initial cohorts, about 85% of students who fail their first science exam attempt to pass on retest, and just under half 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. 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).

As the passing rate has increased, the students who now score below the cutoff are increasingly 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 focus on supporting students in multiple subject areas.

Schools' Responses to the Exit-Exam Policy

Since the introduction of the science exam, the timing and subject-test choices made by students and schools have changed dramatically. In the initial three cohorts, more than half of students (58%) took advantage of the option to test early in 9th grade (Table 4).[12] 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 the biology exam, and 21% the physics exam. Because very few students were taking the chemistry and technology/engineering tests, the state discontinued them following the spring 2023 administration.

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 +0.07 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 February or June 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 about 90% of students in the first cohort affected by the policy, had nearly all students test in only 1-2 grade/subject groups (e.g., 9th grade biology only; 9th grade physics & 10th grade biology). For example, the share of high schools in which nearly all students took the biology exam as 9th 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 9th grade science exam attended schools in which fewer than 5% of 9th 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 in the cohorts of 2018-20) did not take a science course at all. For those who did, most took 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 took another biology course the next year. In fact, half of all students who initially failed the science test ended up taking biology the next year.

By focusing on test-takers near the passing threshold in the graduating cohorts of 2015-17, we can examine how barely passing or failing the test affects students' course-taking behavior. Here, we use our regression-discontinuity model to estimate impacts on the number of full-year science course equivalents (FYEs) that students passed, from the year after their first attempt at the science exam through the year following their expected high-school graduation. Students who barely failed the test went on to pass more science FYEs than those who barely passed; however, this impact was present only for those who tested in 10th grade, not 9th (see Figure 6; full results available on request). Among these 10th graders, barely failing caused students to take and pass 0.21 more science FYEs in the three years following the test, compared to those who barely passed. This same pattern by grade is present in the endogenously selected sample of students who remained enrolled in Massachusetts high schools through their expected year of graduation.

<Insert Figure 6 about here>

Because nearly half of 10th grade test-takers attended high schools where virtually all students tested in 10th grade, these high schools – including CTE schools and those with a three-semester biology course sequence – may have mandated additional science coursework for students who failed the MCAS. In other schools with both 9th and 10th grade test-takers, failing scores in the latter group may be met with more urgency, given that those students have less time before their expected graduation date to pass a science exam.

Discussion

Massachusetts policymakers adopted a high-school exit exam in science with a clear motivating goal: increasing students' preparation for and interest in postsecondary coursework and careers in STEM fields. They targeted for improvement the rigor and accessibility of the high-school science curriculum, as well as students' willingness to take advantage of science course offerings. But the best policies to accomplish this were not obvious. As discussed in the beginning of this paper, the evidence on the consequences of mandating more high-school science coursework was not encouraging. Instead, Massachusetts built on its existing policy by adding a science exit exam as a high-school graduation requirement.

While extensive prior research has analyzed exit exams in math and ELA, science is different in at least three key respects. First, a key motivation behind the science exam was to focus more attention, on the part of both educators and students, on science – a discipline that had been deemphasized in favor of math and ELA and that was critical for the development of the state's future workforce. Second, science exams are potentially more vulnerable to construct-irrelevant variance for English learners, due to the use of complex vocabulary and extended reading passages in test items. Finally, a large body of research on gender and science instruction and achievement makes the question of differential impacts for females and males uniquely salient for a science exam.

We look at the effects of this decision, both overall and with reference to subgroups of students. First, the exit exam appears to have contributed to the standardization of high-school science course sequences, particularly in 9th and 10th grades. In recent years, more than three-quarters of students have tested for the first time as 9th graders, and the proportion of first-time test-takers in 9th grade biology has steadily grown over time.

Second, it is clearly the case that average scores on the science exit exam rose dramatically in the first few years after its introduction, overall and for all student groups. The interpretation of this trend, however, is complicated by the literature on score inflation, particularly in the case of high-stakes tests. While it could be that science score gains derived from improved teaching and reflect a deeper grasp of the entire science curriculum, they could also result from increased familiarity with the test, inappropriate test preparation, and/or narrowing of instruction to tested content (Center for Education Policy, 2007; Hamilton et al, 2007; Koretz, 2008; Koretz & Hamilton, 2006; Murnane & Papay, 2010).

Because of these considerations, we cannot draw firm conclusions about the system-wide consequences of the introduction of the science requirement on students' science knowledge. While a much lower percentage of first-time science test-takers scored below the passing threshold over time, this pattern may or may not reflect increased student mastery of the state's curricular standards in science. Even if it did, however, there would be additional difficulties in attributing those improvements solely to the exit-exam policy because other changes took place concurrently, including increased state spending on public education.

While it is not possible to determine the general equilibrium effects of the introduction of the science exit exam, we can provide more convincing evidence about effects on the margins. First, for the initial cohorts of science test-takers, we have enough longitudinal data to shed light on the impact of exit-exam performance on the later outcomes of students scoring near the passing threshold. Barely passing the exam had the largest effects on high-school graduation for female test-takers. Previous studies using regression-discontinuity designs have not reported differential impacts of barely passing an exit exam by gender[13]; our own analysis of the introduction of the Massachusetts math exam found high-school graduation effects that were quite similar for males and females on the threshold of passing. The novel finding of a differential impact by gender in science is not explained by differences in the grade and the subject in which males and females first tested, nor by other demographics like race/ethnicity, English proficiency, and family income. The passing rates of male and female students on the 8th grade science test were similar.

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, perseverance, and perception of their own abilities 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).

In reality, both dimensions are likely at play for individual students, and the broader question is whether one or the other is larger. To shed light on these alternative explanations, we use female students' 8th grade science test performance to differentiate between those who had passed or failed as middle-schoolers (Papay et al., 2016). 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 8th grade science assessment. Among these students, barely passing the high-school examination 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 girls had failed the science test in middle school and may have expected to fail again. Passing the science exit exam may have increased their confidence and effort level. By contrast, we find no evidence of impacts for females who had passed in 8th grade, nor for males regardless of their 8th grade test performance. This implies that educators should focus particular attention on females who are struggling in science as early as middle school in order to counter cultural stereotypes and shore up female students' perceptions of their own skills and abilities.

It is important to note that our RD estimates of these marginal effects apply only to students near the passing cutoff in the initial years of the exit exam. The student population in Massachusetts public high schools has changed in key respects since then. For example, English learners made up four percent of all test-takers when the policy was introduced, compared to nine percent in 2023. In the first cohort affected by the policy, about 900 test-takers were immigrant newcomers not proficient in English, compared to nearly 3,500 in the 2025 graduating cohort.

Over time, as the pool of test-takers who fail has shrunk dramatically, English learners and those with a diagnosed disability comprise a steadily larger share of those falling below the cutoff. From our conversations with educators around the state, it seems clear that the exit-exam policy puts enormous pressure on schools with large numbers of such students. The intention behind the graduation requirement is to elicit constructive responses. However, we have heard anecdotal evidence of science teachers, especially those with large shares of immigrant newcomer students, pruning the curriculum to exclude all but the most easily taught topics on the test. This type of response could lead English learners and students with disabilities to be exposed to less of the state's high-school science standards than other students, on average.

More than 15 years after the rollout of the science exit exam, the policy's consequences for English learners and students with disabilities continue to be of concern. The imperative of attending to test-takers who fail, a group disproportionately comprised of the state's most vulnerable high-school students, is the most critical implication of our results.

Notes

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

[2] In 2022, the Next-Generation MCAS science tests were administered for the first time; the scale of the new assessments ranges from 440 to 560.

[3] The exception was an additional retest opportunity in April 2010, offered for students in the first cohort impacted by the new requirement who had yet to pass a science exam. 1,616 students participated in this administration.

[4] 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.

[5] For more on the intuition behind regression-discontinuity designs, see Bloom (2012).

[6] We follow Imbens and Lemieux's (2008) recommendation to exclude observations that fall far from the cut score (thereby "winsorizing" the data) to avoid over-smoothing. Consequently, we eliminate 10% of the observations on either side of, and most remote from, the cutoff.

[7] 84% of students who took the science test passed both the science and math tests on the first attempt (nearly all students in the state passed the ELA test). About 6% passed math but not science, 4% failed both, and 2% passed science but not math. (Another 4% of science test-takers never took the math test.) Thus, most students who failed science passed the math test, suggesting that the results we present are not driven by the overlap in failures on the two tests. Moreover, student performance on the math exam is potentially endogenous for students who take science in 9th grade.

[8] 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.

[9] 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.

[10] 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 3 are from models using these different bandwidths, estimates from the same bandwidth (either h=2 or h=3) follow the same pattern described here.

[11] In 2015, Massachusetts shifted from the NSLP 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).

[12] 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.

[13] The exception is Reardon et al (2010), who find 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.

References

Arold, B. & Shakeel, M. D. (2021). The unintended effects of the Common Core State Standards on non-targeted subjects. Available at SSRN: https://ssrn.com/abstract=3868271

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5): 258-267. https://doi.org/10.3102/0013189X07306523

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. https://www.nber.org/papers/w19182

Bloom, H.S. (2012). Modern Regression Discontinuity Analysis. *Journal of Research on Educational Effectiveness*, *5*(1), 43–82. https://doi.org/10.1080/19345747.2011.578707

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. https://doi.org/10.1080/00220671.2016.1226158

Center on Education Policy. (2007). *Choices, changes and challenges: Curriculum and instruction in the NCLB era.* Washington, DC: Jennifer McMurrer.

Clark, D., & Martorell, P. (2014). The signaling value of a high school diploma. *Journal of Political Economy, 122*(2), 282–318. https://doi.org/10.1086/675238

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. https://doi.org/10.3102/01623737124670

Desy, E.A., Peterson, S.A., & Brockman, V. (2011). Gender Differences in Science-Related Attitudes and Interests among Middle School and High School Students. *Science Educator 20*(2), 23-30.

Erwin, B., Brown, D., and Mann, S. (2023). 50-state comparison: High-school graduation requirements 2023. Education Commission of the States. https://reports.ecs.org/comparisons/high-school-graduation-requirements-2023-04

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.

Frandsen, B. (2017). Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete. In M. D. Cattaneo and J. C. Escanciano, Eds. *Advances in Econometrics: Vol. 38—Regression Discontinuity Designs: Theory and Applications* (pp. 29–72). Bingley, UK: Emerald.

Fredricks, J.A. & Eccles, J.S. (2002). Children's competence and value beliefs from childhood through adolescence: growth trajectories in two male-sex-typed domains. *Developmental Psychology 38*, 519–33.

Gao, N. (2021). *Does raising high school graduation requirements improve student outcomes?* Public Policy Institute of California.

Gewertz, C. (n.d.). Which states require an exam to graduate? *Education Week*. 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. https://nces.ed.gov/nationsreportcard/pdf/main2005/2006466.pdf

Hamilton, L.S., Stecher, B.M., Marsh, J.A., McCombs, J.S., Robyn, A., Russell, J.L., Naftel, S., & Barney, H. (2007). *Standards-based accountability under No Child Left Behind: Experiences of teachers and administrators in three states.* RAND Corporation.

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). https://www.jstor.org/stable/42001531

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. https://doi.org/10.1016/S1479-3539(08)16006-2

Imbens, G. W., and T. Lemieux. (2008). "Regression Discontinuity Designs: A Guide to Practice." *J. Econometrics* 142 (2): 615–35

Institute of Medicine. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. https://doi.org/10.17226/11463.

Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Harvard University Press.

Koretz, D. M., & Hamilton, L. S. (2006). Testing for accountability in K-12. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 531-578). Westport, CT: American Council on Education/Praeger.

Lederman, N.G. & Abell, S.K. (Eds.). (2014). *Handbook of research on science education, volume II*. Routledge. https://doi.org/10.4324/9780203097267

Mantil, A., Papay, J.P., Ferguson, I.M., Mbekeani, P.P., & Murnane, R.J. (2023). *Rising numbers, unmet needs: Immigrant newcomers in Massachusetts high schools.* Providence, RI: Educational Opportunity in MA, Annenberg Institute at Brown University.

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/.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics, 142*: 698–714.

Montgomery, N., Allensworth, E., & Correa, M. (2010). *Passing through science: The effects of raising graduation requirements in science on course-taking and academic achievement in Chicago*. Research Report. Consortium on Chicago School Research.

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. https://www.jstor.org/stable/20799160

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. https://doi.org/10.1162/rest.2010.12318

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. https://doi.org/10.1016/j.econedurev.2009.06.002

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.* Annenberg Institute of School Reform, Brown University. https://annenberg.brown.edu/sites/default/files/LiftingAllBoats_FINAL.pdf

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, 44*(4), 717-733. 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. DOI: 10.3368/jhr.51.2.0713-5837R

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. https://doi.org/10.1080/19345747.2013.819398

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. https://doi.org/10.3102/0162373709352530

Polson, C. (2018). TAKS-ing students? Texas exit exam effects on human capital formation. *Economics Of Education Review, 62*, 129-150. https://doi.org/10.1016/j.econedurev.2017.09.009

Potvin, P., & Hasni, A. (2014). Interest, motivation and attitude towards science and technology at K-12 levels: A systematic review of 12 years of educational research. *Studies in Science Education*, *50*(1), 85–129.

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*(4), 498–520. https://doi.org/10.3102/0162373710382655

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.

Riegle-Crumb, C., Moore, C., & Ramos-Wada, A. (2011). Who wants to have a career in science or math? Exploring adolescents' future aspirations by gender and race/ethnicity. *Science Education 95*, 458–76.

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.

Sadler, P.M., Sonnert, G., Hazari, Z., & Tai, R. (2012). Stability and volatility of STEM career interest in high school: A gender study. *Science Education 96*, 411–27.

Teitelbaum, P. (2003). The influence of high school graduation requirement policies in mathematics and science on student course-taking patterns and achievement. *Educational Evaluation & Policy Analysis* 25(1), 31-57.

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.

Vasquez Heilig, J., & Darling-Hammond, L. (2008). Accountability Texas-style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis, 30*, 75-110.

Vogler, K. E. (2008). Comparing the impact of accountability examinations on Mississippi and Tennessee social studies teachers' instructional practices. *Educational Assessment*, *13*, 1-32.

Wang, M.T. & Degol, J.L. (2017). Gender gap in Science, Technology, Engineering, and Mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Education Psychology Review 29*, 119–140.

Warren, J. R., Grodsky, E. S., & Lee, J. C. (2008). State high school exit examinations and post-secondary labor market outcomes. *Sociology of Education*, *81*(1), 77–107. https://doi.org/10.1177/003804070808100104

Xie, Y., Fang, M., & Shauman, K. (2015). STEM education. *Annual Review of Sociology*, *41*(1), 331-357.

Table 1

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=25551)	Proportion who failed	Ever retested	Passed on first retest	Never passed	HS Graduation Rate
Female	0.495	0.495	0.122	0.864	0.434	0.139	0.581
Male	0.505	0.505	0.122	0.840	0.406	0.160	0.475
Low-income	0.427	0.816	0.233	0.849	0.388	0.167	0.492
Higher-income	0.573	0.184	0.039	0.866	0.558	0.073	0.688
English Language Learners	0.036	0.165	0.558	0.832	0.294	0.225	0.458
Students with disabilities	0.148	0.404	0.334	0.870	0.371	0.190	0.524
Urban	0.267	0.564	0.258	0.846	0.359	0.181	0.476
Asian	0.050	0.038	0.093	0.873	0.409	0.153	0.563
Black	0.083	0.191	0.282	0.866	0.386	0.165	0.538
Hispanic	0.123	0.346	0.343	0.842	0.345	0.188	0.459
Multiple race/ethnicity	0.015	0.016	0.129	0.840	0.402	0.141	0.519
Native American	0.003	0.003	0.152	0.865	0.427	0.135	0.506
White	0.727	0.406	0.068	0.852	0.501	0.109	0.579
All students	1.00	1.00	0.122	0.852	0.420	0.150	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.

Table 2

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	0.625	0.135	0.413	0.433	0.154
10th grade	9,589	0.375	0.105	0.762	0.099	0.139
Biology	16,279	0.637	0.109	0.789	0.067	0.143
Chemistry	864	0.034	0.205	0.111	0.765	0.124
Physics	7,813	0.306	0.156	0.109	0.73	0.160
Tech/Engineering	595	0.023	0.106	0.166	0.681	0.153

Table 3

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	0.0376**	0.0126+	0.0203**	0.0126*	0.00324
	(0.00464)	(0.00620)	(0.00273)	(0.00344)	(0.00174)
Female	0.0697**	0.0377**	0.00593	-0.00453+	-0.000669
	(0.0010)	(0.0100)	(0.0047)	(0.0017)	(0.0008)
Male	0.00147	-0.0153**	0.0359**	0.0313**	0.00712
	(0.0092)	(0.0035)	(0.0012)	(0.0057)	(0.0044)
Low-income	0.0417**	0.0135	0.00762+	0.00646	-0.00137
	(0.00839)	(0.0110)	(0.00331)	(0.0040)	(0.0026)
Higher-income	0.0241**	0.00964	0.0585**	0.0308**	0.0181**
	(0.00509)	(0.0115)	(0.00188)	(0.0013)	(0.00147)
English language learners	-0.00221	0.00437	-0.0661*	-0.0352*	-0.0190+
	(0.0120)	(0.0241)	(0.0220)	(0.0083)	(0.0069)
Proficient in English	0.0421**	0.0136	0.0300**	0.0172*	0.00511*
	(0.00609)	(0.00882)	(0.0014)	(0.0045)	(0.0013)
N	17089	23867	17089	17089	17089

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

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)

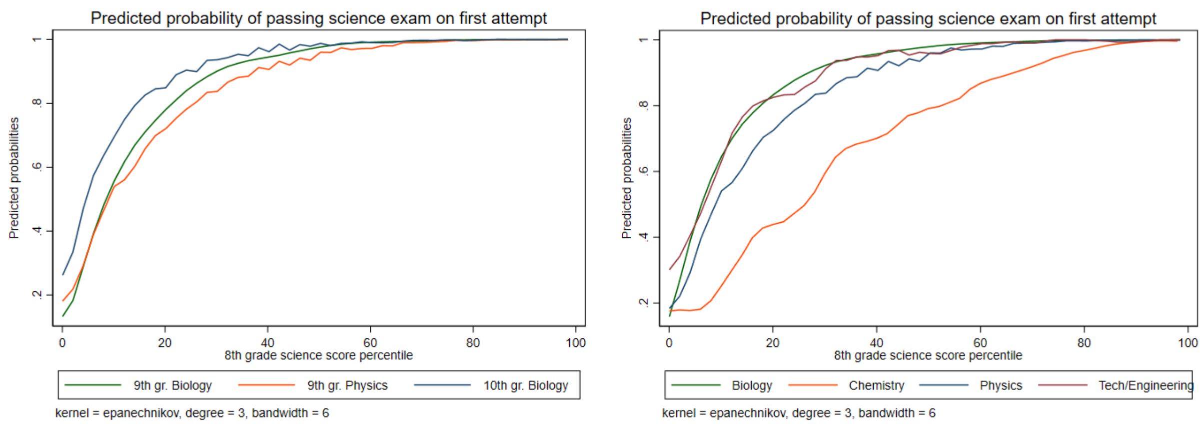
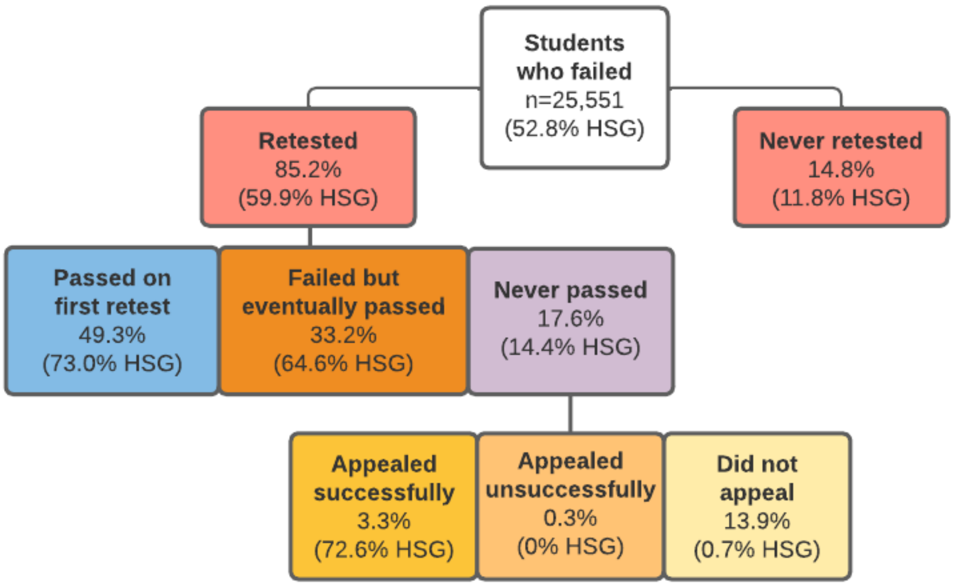


Figure 2

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). In these initial cohorts, 347 students (less than 0.2% of all test-takers) graduated from high school without passing a science test or successfully appealing, likely due to confusion over which students were in the graduating cohorts bound by the new requirement.

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

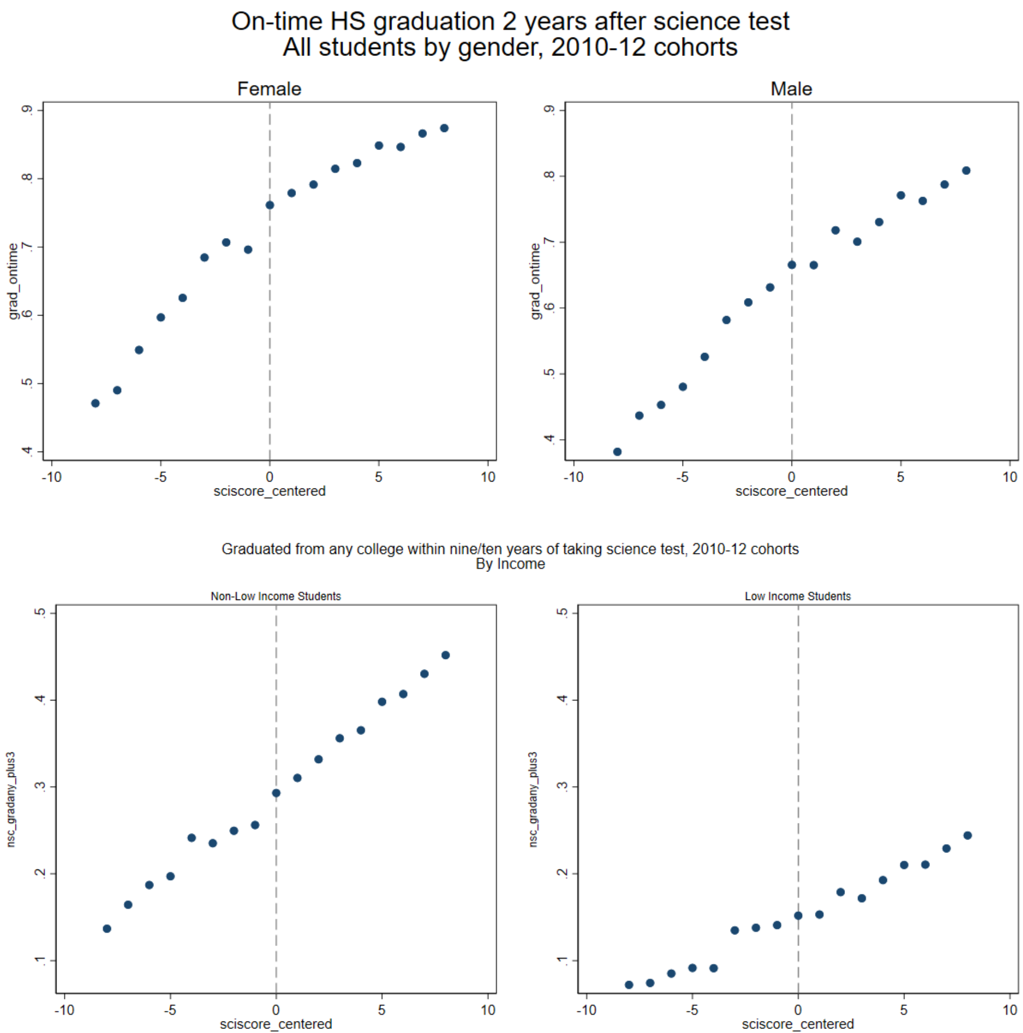
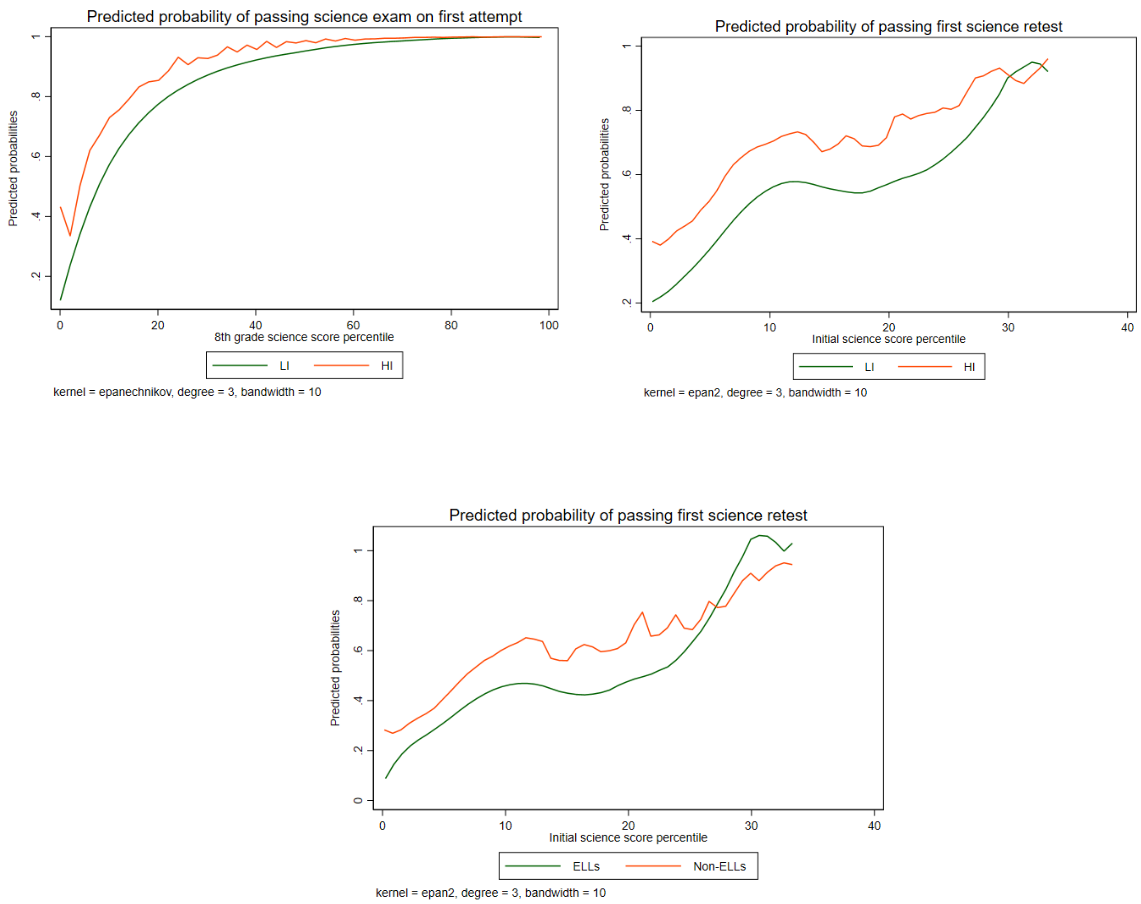


Figure 4

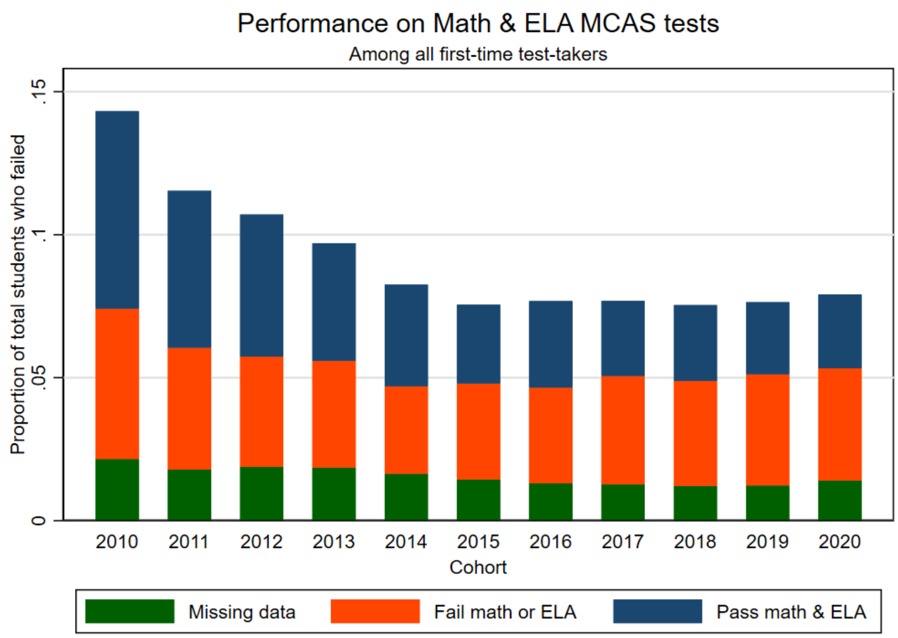
Predicted probability of passing high-school science exam on first attempt, conditional on 8th grade science scores, and on passing first retest, conditional on initial high-school score, by income (top panel) and English language proficiency (bottom panel)



Notes: While the passing cutoff was at the 15th percentile on average, it differed by subject test. In chemistry, the passing cutoff was at the 34th percentile; all test-takers above the 23rd percentile are the relatively few students who take chemistry, but the general patterns hold across subject areas. 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

Sample mean science course full-year equivalents (FYEs) at score points near the passing threshold on the science high-school exit exam for students in the graduating cohorts of 2015-17, by student grade level