**The Consequences of MCAS Exit Examinations for Struggling Low-Income Urban Students**

John P. Papay

Richard J. Murnane

John B. Willett

Harvard Graduate School of Education[[1]](#footnote-1)

July 2008

**Abstract**

The growing prominence of high-stakes exit examinations has made questions about their effects on student outcomes increasingly important. We capitalize on a natural experiment to evaluate the causal effects of failing a high-stakes test on high school completion for students scheduled to graduate from Massachusetts high schools in 2006. With these exit examinations, states divide a continuous test score into two categories – pass and fail – so students with essentially identical performance may have different test outcomes. We find that, for low-income urban students on the margin of passing, failing the 10th grade mathematics examination reduces the probability of on-time graduation by eight percentage points. The large majority (89%) of students who fail the 10th grade mathematics examination retake it. However, although low-income urban students are just as likely to retake the test as apparently equally skilled suburban students, they are much less likely to pass this retest. Furthermore, failing the 8th grade mathematics examination reduces by three percentage points the probability that low-income urban students stay in school through 10th grade. Importantly, we find no effects for suburban students or wealthier urban students.

**The Consequences of MCAS Exit Examinations for Struggling Low-Income Urban Students**

***I. Introduction***

 As part of standards-based educational reforms introduced over the past two decades, many states have implemented exit examinations that students must pass in order to earn high school diplomas. Advocates argue that such examinations create incentives for students to work at learning important cognitive skills. By certifying that high school graduates have mastered the state-defined academic content standards, the examinations may also increase the economic value of a high school diploma (Evers & Walberg, 2002). Opponents of these tests suggest that they put unnecessary stress on students and encourage them to drop out of high school. They also argue that such tests place the greatest burden on the very groups who are already struggling in the educational system, such as low-income and special needs students (Thomas, 2005; Jones, Jones, & Hargrove, 2003). Because high school graduation is associated with many positive life outcomes, the question of how high-stakes testing affects high school completion rates is important to educational policymakers.

While past research has focused primarily on the overall effects of imposing high-stakes accountability on aggregate student outcomes, we look at the consequences of exit examinations on individual students within a high-stakes testing regime. Using data from Massachusetts, a state that has earned a national reputation for rigorous content standards and for English language arts (ELA) and mathematics assessments that are well aligned to the standards, we examine whether failing the 10th grade mathematics examination on the first try affects the probability of on-time high school graduation.

The greatest challenge in determining the effect of failing the MCAS on the probability of on-time high school graduation lies in identifying a reasonable comparison group. Because students who pass the test have different abilities, on average, than those who fail, we cannot simply compare the outcomes of these two groups. However, recognizing that the state determines who passes and who fails using an arbitrarily defined cut score, we compare students who barely passed with those who just failed. These students are quite similar, with scores on the test differing in some cases by just one point. We capitalize on the “natural experiment” provided by the minimum passing score to examine the causal impact of failing the statewide 10th grade mathematics examination on the probability of on-time high school graduation.

An exit examination can prevent students from graduating from high school in three ways: fear of failing may cause them to drop out before taking the test; failing the examination may cause them to drop out before re-taking it; and failure to pass even after multiple attempts may prevent graduation. We refer to these mechanisms as *Fear*, *Discouragement*, and *Repeated Failure*. We use causal and descriptive analyses to explore which of these mechanisms – if any – affects students.

We find that, for equally able low-income, urban students near the cutoff, failing the 10th grade mathematics exit examination reduces the probability of on-time graduation by eight percentage points. In contrast, failing the test does not affect the probability of on-time graduation for wealthier urban students or suburban students on the margin of passing. Thus, the combination of low family income and urban schooling makes students particularly susceptible to the effects of failing. Importantly, we cannot distinguish here between a negative effect of failing the examination and a positive effect of passing it – students who fail the test may be disappointed with their performance and drop out of school, while students who pass may be encouraged to persist in school. Regardless, the practice of dividing students with essentially the same ability into two categories by this cut score has an impact on student outcomes; this effect poses an important challenge for urban districts.

Furthermore, for a typical low-income urban student on the margin of passing the 8th grade mathematics examination, failing that test also reduces the probability of persisting through 10th grade by three percentage points. This provides some evidence that *Fear of Failure* may play a role in students’ decisions. Interestingly, we find that failing the 10th grade ELA examination does not affect the probability of graduation for low-income urban students on the margin of passing.

We supplement these causal conclusions with descriptive analyses that explore possible sources of these effects for urban students with low family income. Here, we focus on students who fail the mathematics exit examination when they first take it at the end of grade 10, exploring their persistence and success on retests. Students who fail have many opportunities to retake the examination before graduation, with some students completing six retests. At each retest opportunity, more than 80% of students who fail continue to retake the test. Massachusetts students show remarkable persistence, but relatively few students exhaust all of their retest opportunities. Instead, of the students who fail and never pass the examination, nearly two-thirds stop taking retests and drop out of school, presumably because their poor test performance has discouraged them. Here, we find important differences for urban, low-income students compared to their suburban peers. Among students with the same initial test scores, low-income urban students who fail the statewide mathematics examination at the end of the 10th grade are just as likely as suburban students to retake the test, but they are much less likely to pass on retest. Differences in academic support could explain this pattern.

In Section II, we provide a brief discussion of standards-based reforms, their development in Massachusetts, and past research on the effects of high-stakes testing. In Section III, we explain our data sources, measures, and analytic strategy. Here, we justify our ability to make causal claims from these data. In Section IV, we detail our main findings. In Section V, we perform several sensitivity analyses to verify the robustness of our results. We conclude with a discussion of our findings and the questions they raise for policy-makers.

***II. Background and Context***

Standards-Based Educational Reforms and High-Stakes Testing

In the years since the 1983 publication of *A Nation at Risk*, the standards-based reform movement has gained momentum and exerted substantial influence on state and federal education policy. While the details of these reform efforts vary greatly from state to state, common components include content standards in core academic subjects and regular testing to monitor student progress toward mastering these standards. In addition to developing accountability structures for schools, many states have begun attaching consequences for students to their performance on the state-wide examinations. Currently, 25 states have or are phasing in examinations, typically in English language arts (ELA) and mathematics, that high school students must pass in order to graduate (Center on Education Policy, 2007). In most states, including Massachusetts, students first take these exit examinations as 10th graders. Students who fail typically have multiple opportunities to retake the examination before graduation.

Critics of high-stakes examinations argue that they may lead some students to drop out of high school (Thomas, 2005; Jones, Jones, & Hargrove, 2003). A 1999 *National Research Council* report cites qualitative work suggesting that “graduation tests pose no threat to most students, but, among those who fail them, they increase a sense of discouragement and contribute to the likelihood of dropping out” (Heubert & Hauser, p. 175).

Any policy that causes students to drop out of school has substantial consequences because high school graduation remains a gateway into better paying jobs and post-secondary education. Employers recognize and reward the skills that college graduates possess, especially the ability to engage in non-routine problem-solving and to communicate effectively (Levy & Murnane, 2004). Because it causes students to complete less education, dropping out also reduces students’ quality of life in a variety of dimensions, including reduced health, wealth, and happiness (Oreopoulos, 2007), increased criminality (Lochner & Moretti, 2004), and increased mortality rates (Lleras-Muney, 2005).

Given the importance of high school completion and the possible negative consequences of high-stakes testing, many scholars have explored whether exit examinations reduce graduation rates. This work has taken two main forms: some researchers have examined the effect of statewide policy changes on aggregate student outcomes, while others have focused on the relationship between an individual student’s performance on the test and that student’s probability of graduating in states with high-stakes testing regimes.

Much early work examining aggregate outcomes used correlational evidence; Clarke, Haney, & Madaus (2000) review this literature and conclude that “high stakes testing programs are linked to decreased rates of high school completion.” Exploiting variation in exit examination policies across states and/or over time, some recent work provides at least tentative support for these correlational conclusions (Reardon & Galindo, 2002; Warren, Jenkins, & Kulick, 2006; Nichols, Glass & Berliner, 2006). In contrast, Carnoy & Loeb (2002), Greene & Winters (2004), and Carnoy (2005) find no relationship between state accountability policies, including high school exit examinations, and high school completion rates. Some recent work suggests that exploring aggregate patterns may obscure heterogeneity in effects for different groups of students. Dee and Jacob (2006) find increased dropout rates only for urban and minority students, while Jacob (2001) finds similar patterns only for the lowest achieving students.

Research that examines the relationship between individual student performance on exit examinations and high school completion remains much less common. Using data from the *Florida Minimum Competency Test* from 1987-91, Griffin & Heidorn (1996) find a relationship between student performance and drop-out rate only for students with high GPAs. While the authors control for selected student characteristics, the results cannot be interpreted causally because students who fail the examination likely differ from those who pass in critical unobserved dimensions. Griffin & Heidorn also focus on the impact of a minimum competency test, which differs substantially from the current incarnation of state-mandated high school exit examinations. Cornell, Krosnik & Chang (2006) examine a group of students who were wrongly informed that they had failed the Minnesota high-stakes examination. Most of these students reported some negative academic impact of “failing” this test.

Martorell (2005) provides causal estimates of the effect of failing a high school exit examination on high school graduation, using a regression discontinuity analysis similar to the one we employ in this paper. He finds no effect of failing the Texas exit examination on high school graduation for students who barely failed. This finding holds for every examination until the very last administration of a student’s senior year. As students run out of testing opportunities, failing the examination does prevent them from graduating because they cannot satisfy state requirements.

We do not address the overall consequences of standards-based accountability in Massachusetts. Instead, like Martorell (2005), we look at how this policy plays out in a state committed to standards-based reform. In other words, we look at the effects of dividing a continuous measure of student proficiency into two categories – pass and fail – at an arbitrary cut score. We extend Martorell’s research in several respects. Most importantly, we look for (and find) heterogeneous causal effects. In Massachusetts, examining only aggregate impacts masks a substantial effect for low-income urban students; as a result, we focus our analyses on this group. Second, we examine additional mechanisms by which exit examinations may decrease high school graduation, including the possibility that students drop out even before taking the 10th grade test. Third, we conduct descriptive analyses that shed light on the sources of the heterogeneous causal impacts. Finally, we make use of data from a state quite different from Texas.

The Massachusetts Context

 In the 15 years since the Massachusetts legislature passed the *Massachusetts Education Reform Act* of 1993, the state has invested more than one billion dollars per year in additional funding for K-12 public education. The investment has borne considerable fruit. For example, a 2006 study by the Fordham Foundation praised the Massachusetts academic standards as the most rigorous in the country (Finn, Julian, & Petrilli, 2006).[[2]](#footnote-2) A 2006 report by *Education Week* concluded that the state-wide tests used to assess the English language arts and mathematical skills of Massachusetts students (part of the Massachusetts Comprehensive Assessment System (MCAS)) were well aligned with the state’s demanding academic standards. While this report gave an average grade of B- to the standards and accountability systems developed by states, it gave a grade of A to the Massachusetts system (*Quality Counts*, 2006).

Most importantly, Massachusetts students are doing well and have improved performances markedly on the National Assessment of Educational Progress (NAEP) examinations in recent years. In 2007 Massachusetts’ 4th graders ranked first nationwide on the NAEP reading and mathematics tests and second nationwide on the writing test. The state's 8th graders ranked first in mathematics, tied for first in reading, and third in writing on the NAEP tests (NCES, 2008). Furthermore, since the introduction of state testing under standards-based reform, the state’s NAEP performance has improved rapidly. As Figure 1 shows, Massachusetts 8th graders not only far exceed the national average, but their performance has increased much more rapidly than the national average. Thus, it is in the context of a system that has brought about significant accomplishments that we examine the consequences for students of failing the MCAS examination.

FIGURE 1 ABOUT HERE

 Massachusetts began administering the MCAS mathematics and ELA examinations in 1998. For the class of 2003, the 10th grade tests became high-stakes exit examinations. Students must pass both tests in order to receive a high school diploma. The state allows students to take the tests without time constraints and to retake them repeatedly if they fail, attempting explicitly to make the MCAS as minimal a barrier to graduation as possible.[[3]](#footnote-3) Critics, however, claim that even with these safeguards, the examinations do indeed prevent students from graduating.

Of the nearly 70,000 students who took the mathematics examination in 8th grade in 2002, 76% went on to graduate on time in Massachusetts in 2006. We can partition those students who did not graduate on time into two groups – those who did not persist to take the 10th grade examination (9%) and those who took the 10th grade test but did not graduate two years later (15%). Thus, most students who did not graduate left the system after taking the 10th grade examination; we focus first on this population and return to the group who dropped out before 10th grade later in the paper.

That students who passed the 10th grade MCAS examination on their first attempt graduate at greater rates than students who fail is not surprising – all students must pass the test to graduate. Of the 66,347 students in the 2006 graduating cohort who took the 10th grade MCAS mathematics examination for the first time in 2004, 87% passed on their first try. However, students who failed faced substantial risk of dropping out: only 50% of them went on to graduate on time, compared to 90% of the students who passed.

While striking, this descriptive pattern does not confirm that the exit examinations pose a barrier to graduation. A student’s MCAS scores are associated with a variety of other characteristics, such as academic proficiency, motivation, and access to educational resources, that also affect their probability of graduation. As a result, we would expect students who fail the examination to drop out at greater rates, even in the absence of any testing requirement. The direct relationship between MCAS score and the graduation rate among students who did pass the 2004 test provides evidence for this conclusion. Among these students, 73% who just passed graduated on time, compared to 98% of students with a perfect score. Thus, a challenge in our current study involves disentangling the effects of failing the examination from the effects of student ability and other background characteristics related to test performance.

Conceptually, we would like to take students who scored identically, right at the pass/fail cut score, and randomly assign them to either “pass” or “fail” the test. By randomly assigning students, the two groups would be “equal on average” on all characteristics, both observable and unobservable, prior to treatment. We could then identify any differences in the ultimate outcome (high school graduation) as a causal effect of failing the examination. Such an experiment is, of course, both impossible and unethical. However, we can take advantage of the state’s imposition of a minimum passing score to provide a natural experiment from which we can draw the same causal conclusions. By examining students with nearly identical MCAS performance, but just on either side of this cutoff, we can interpret any differences in their graduation outcomes as the causal effect of failing the examination for these students “on the margins” of passing (Shadish, Cook, & Campbell, 2002). We discuss this regression discontinuity approach in more detail in Section III.

Research Questions

We first examine the effect of failing the 10th grade examination, paying particular attention to impacts on low-income, urban students. We then attempt to explore the *Fear of Failing* effect that may arise as students predict they will not pass the 10th grade test and drop out before even taking it. We cannot identify this effect cleanly, but we can get some sense of its magnitude by examining student performance on the 8th grade test. Finally, we look at students who fail their 10th grade test, examining their persistence and success on retests. We explore whether students who fail and drop out do so because of *Discouragement* (students give up and drop out after failing one or more of the examinations) or through *Repeated Failure* (after exhausting the available retest opportunities, students still have not satisfied the testing requirements). Specifically, we address three primary research questions:

RQ1. Does failing the high school exit examination as a 10th grader make students on the margin of passing less likely to graduate from high school on time?

RQ2. Among students who fail the 10th grade exit examination and do not graduate, is the primary mechanism one of *Discouragement* or *Repeated Failure*?

RQ3. Does failing the 8th grade test cause students on the margin of passing to drop out before taking the 10th grade examination?

***III. Research Design***

Data Sources

The Massachusetts Department of Elementary and Secondary Education has compiled a comprehensive database that tracks students longitudinally throughout high school, allowing for clear description of student graduation outcomes. For the 2006 graduating cohort, the records contain each student’s MCAS mathematics and ELA test results, demographic characteristics, and status at cohort graduation, including whether the student graduated, dropped out, is still enrolled, transferred out, was expelled, or any of eleven possible outcomes. This dataset allows for much more precise estimation of the probability of high school completion than do previous studies, and it permits investigation of the direct link between student performance on high-stakes tests and graduation outcomes at the individual level.

Our dataset includes 83,892 student records from across the state of Massachusetts. To analyze the effect of failing the 10th grade examination (our first research question), we focus on members of the 2006 graduating cohort who first took the 10th grade mathematics MCAS examination as sophomores in 2004 and for whom the examination was a high-stakes test. This sample thus includes students who entered the state between 8th and 10th grade and those who missed the 8th grade examination, but it does not include students who entered the state after 10th grade. Our final sample for addressing the first research question includes 66,347 students.[[4]](#footnote-4) To address our third research question, we use the 69,127 students in the same cohort who took the 8th grade mathematics examination. This sample includes students who dropped out of school before 10th grade.

Measures

To address our first research question, we created a dichotomous outcome variable, named *GRAD*, that indicates whether the student graduated from high school in Massachusetts on time in 2006 (1=graduated on-time in Massachusetts; 0 otherwise). Districts report the values of individual student graduation outcomes to the Department of Elementary and Secondary Education using the state’s *Student Information Management System* (SIMS). Note that students can be coded as zero either for dropping out of school, for moving out of state before graduation, or for continuing in high school without graduating. In Section V, we explore the sensitivity of our results to this outcome definition. We created several additional outcome variables for the descriptive analyses that we used to address our second research question. For students who failed the examination, we created dichotomous outcomes that indicate whether the student retook the test (*RETAKE*) and whether they passed this retest (*RETAKE\_PASS*). Finally, to address our third research question, we created another dichotomous outcome measure, named *TAKE10th*, that indicates whether a student who took the 8th grade mathematics examination persisted in school to take the 10th grade test (1=persisted to take the test; 0 otherwise).

The dataset contains a record of scores from every MCAS mathematics and ELA examination that each student took from 8th grade[[5]](#footnote-5) through high school graduation. The state reports raw scores, scaled scores, and performance level for each test. A scaled score of 220 qualifies as passing, with a different performance rating each 20 points, as follows: (a) *200 to 218*: Failing, (b) *220 to 238*: Needs Improvement, (c) *240 to 258*: Proficient, and (d) *260 to 280*: Advanced. Since multiple raw scores translate to a single scaled score, we use raw scores in our analyses in order to preserve fine-grained performance differences on the test.[[6]](#footnote-6) For the 10th grade mathematics examination, raw scores ranged from 0 to 60; students who earned more than 20 points passed the test.[[7]](#footnote-7) To implement our regression discontinuity approach, we centered students’ raw scores by subtracting out the value of the corresponding minimum passing score. On these re-centered continuous predictors, *MATH* and *ELA*, a student with a score of zero had achieved the minimum passing score. We also created a dichotomous predictor, *PASS*, to indicate whether the student passed the examination (1=student passed; 0 otherwise).

The dataset also includes the values of several key control predictors, such as student race and gender as well as dichotomous variables indicating whether the student was classified as limited English proficient (*LEP*), special education (*SPED*), low-income (*LOWINC*), attending a high school in one of Massachusetts’s 22 urban school districts (*URBAN*), or appearing in the 10th grade sample without an 8th grade test score (*NEWSTUDENT*).[[8]](#footnote-8) Each of these indicators is coded 1 for those who belong to the category, and 0 otherwise. Overall, 26% of the students attended urban schools and 28% of students were identified as low income. Low-income students tended to cluster in urban schools: 63% of urban students lived in poverty, compared to just 16% of suburban students.

Data Analyses

We address our first and third research questions by conducting identical analyses with the relevant outcome variables. We describe below the analyses that we use to address our first research question, which concerns the impact of just failing the 10th grade mathematics examination on the probability of on-time high school graduation. To explore whether just failing the 8th grade mathematics test reduces persistence to 10th grade (our third research question), we replace outcome *GRAD* by outcome *TAKE10th*.

Because students who score better on the MCAS have a higher probability of graduating from high school on time, we cannot make causal inferences about the impact of failing the test simply by comparing the graduation rates of students who pass and fail the examination. However, under conditions that we discuss below, we can analyze data from our natural experiment – using a regression discontinuity strategy – to make such causal inferences *for students at the margin of passing*.[[9]](#footnote-9) Because the probability that a student will pass the examination goes from zero to one at a single cut score, our discontinuity is sharp.

The internal validity of our regression discontinuity analyses – and consequently our ability to make unbiased causal inferences about the impact of exit examinations – relies on several critical assumptions about the relationship between student MCAS score and graduation. First, we assume that the true relationship between MCAS score and graduation is smooth and continuous around the cut score, except for the impact of passing/failing the examination itself. Second, we assume that we can model this smooth relationship accurately in the region around the cut score. Finally, we assume that students cannot adjust their effort levels to just achieve the minimum passing score and they can only influence their passing status through their test score. If these assumptions hold, then we can use the data from our natural experiment to test whether the smooth relationship between graduation and MCAS performance is disrupted at the cut score. If so, the magnitude of this disruption provides an unbiased estimate of the causal impact of failing the examination for students at the cut score.

Our key challenge is to estimate accurately the probability of on-time graduation for students immediately on either side of the cut score. The difference in the probability of on-time graduation between students scoring at the cutoff who just passed and those who just failed then becomes our estimate of the effect of failing the examination. Essentially, we model the relationship between probability of on-time graduation and MCAS score, on either side of the cut score, as seen in Figure 2. We predict the probability that students who failed the MCAS would have graduated on time if they had just achieved a passing score but still failed; this is illustrated by point A in Figure 2. In the figure, the vertical difference between A and B is the effect of failing the MCAS.

FIGURE 2 ABOUT HERE

In our analyses, we use students who failed the test to estimate point A and students who passed to estimate point B. We model this continuous relationship using a non-parametric smoothing process with local linear regression (Imbens & Lemieux, 2007; Hahn, Todd, & Van der Klaauw, 2001; Fan, 1992). We describe this method in detail in Papay, Murnane, and Willett, 2008. Here, we focus attention on the single linear regression analysis that centers on the cut score and estimates points A and B. In this regression, we use only observations within a narrow bandwidth, *h*, on either side of the cut score, as follows:[[10]](#footnote-10)

 (1)

for the *ith* individual. In this model, parameter  represents the causal effect of passing the 10th grade MCAS mathematics examination. If its estimated value is statistically significant and positive, then we know that classifying a student as *passing* the high-stakes test at the cut score, as opposed to *failing* it, causes the student’s probability of graduating from high school on-time to *increase* discontinuously.

 We extend this simple model in several ways. First, we include selected student background covariates (**Xi**). Second, because our primary outcome is a binary predictor that indicates whether the student graduates from high school on time, we replicate our analysis using logistic regression. Here, we limit our analysis to those observations that fall within a narrow window around the cut score. Finally, we examine the impact of test failure on high school graduation for particular subgroups of students, including urban students from low-income families. To conduct these latter analyses, we fit separate regressions for each subgroup.

For two reasons, we focus on the statistically significant impact of just passing/failing the MCAS mathematics test on high school graduation for low-income urban youth. First, the educational challenges facing these students have received national attention. Consequently, understanding the impact of high-stakes testing on the academic prospects for struggling low-income urban students is especially relevant to educational policy formulation. Second, the data currently available to us are insufficient to support exploration of other interesting questions, such as the effect of just failing the 10th grade MCAS test on urban special education students. We plan to examine additional subgroup effects in future research after we have increased our sub-sample sizes by pooling data across multiple graduation cohorts.

To address our second research question, we conduct analyses in which we explore *why* failing the 10th grade MCAS mathematics test reduces the probability of high school graduation for low-income urban students, but not for their wealthier or suburban peers. However, we interpret these results only descriptively because the additional analyses cannot support unbiased causal inference. In these descriptive analyses, we explore patterns of test-taking persistence and success for students who fail, in order to see whether low-income urban students are less likely than wealthier or suburban students to retake the examination or to pass their first retest. Here, we fit probit models of the following form on the sample of students who failed the 10th grade mathematics examination:

 (2)

for the *ith* student. In our model, our principal research interest focuses on the parameter sum, *β*1+*β2*, which represents the difference between low-income urban students and suburban students in the probability of retaking the test. By including mathematics and ELA test scores and whether the student passed the ELA examination in the model, we explicitly compare students with the same proficiency on both the mathematics and ELA examinations.

We conduct similar analyses to examine retesting success, replacing the outcome *RETAKEi* with *RETAKE\_PASSi*. This analysis proves slightly more complicated because we restrict our sample to students who fail the test.[[11]](#footnote-11) As a result, we use an instrumental variables approach, recommended by Hanushek & Rivkin (2006). We discuss the issues raised and our approach to resolve them more fully in Papay, Murnane, & Willett (2008).

***IV. Findings***

(1) Effect of failing the high-stakes exit examination on high school graduation

 Failing the 10th grade MCAS mathematics examination reduces the probability that a low-income, urban student on the margins of passing will graduate from high school on-time by eight percentage points (p=0.015). Given that 26% of low-income, urban students who just pass the exam do not graduate on time, this effect is quite substantial. We find no such effects for wealthier urban students or for suburban students, regardless of family income. Thus, it is the interaction of low family income and an urban environment that appears to render students, on average, more susceptible to the effects of failing.

We present the fitted non-parametrically smoothed relationship between graduation and MCAS mathematics score for typical low-income urban students from our preferred specification in Figure 3. For these low-income urban students at the margin, failing the examination substantially reduces their subsequent probability of graduation. Visually, this effect appears as an interruption in the underlying smooth relationship between the probability of graduation and the MCAS mathematics score at the cut score. For perspective, we have included the sample mean probabilities of on-time graduation at each MCAS score level.

FIGURE 3 ABOUT HERE

The effects for wealthier urban and suburban students are not statistically significant. However, the point estimates indicate that for these subgroups, students on the margins of passing who just fail have a slightly *greater* probability of on-time graduation than students who just pass. This seemingly counterintuitive pattern could stem from efforts by schools with ample resources to focus attention on the relatively few students with failing MCAS scores. Recent research by Neal and Schanzenbach (2007) lends some support for this claim; the authors find that, in the Chicago Public Schools, teachers face and respond to incentives to focus instruction on students who seem likely to improve their performance on the high-stakes examination.

(2) Persistence and success in retesting among students who fail

 Overall, the 8,269 students who failed the mathematics MCAS on their first try in 2004 showed remarkable persistence in retaking the examination. Nearly 89% took the examination at least one more time and, of these students, 68% went on to pass the test at some point in high school. On average, students who never passed the examination retook it twice before giving up. As the sample histogram in Figure 4 illustrates, on each retest, approximately 35% of the students passed. Among those who failed each retest, most students (85 to 90 percent) decided to retake it yet another time. Although not shown, the numbers of students pursuing retests declines precipitously after the fourth retest: only 113 students retook the examination a fifth time, and only 7 took a sixth retest. Thus, very few students took advantage of all their retest opportunities.

FIGURE 4 ABOUT HERE

Among students who failed their first test, we find evidence for both *Discouragement* and *Repeated Failure*. Here, we examine students whose initial test scores placed them close to the cut score but who never pass a retest. More than two-thirds of these students stop taking retests at some point and do not attempt the March 2006 examination, the last retest before the cohort’s graduation. Presumably, these students became discouraged and dropped out of school. However, one-third of these students persist to the March 2006 retest. Over 85% of these students have taken at least four retests, showing remarkable persistence. For these students, *Repeated Failure* appears to be the mechanism at play as they exhaust all of their retest opportunities but cannot satisfy the graduation requirement. These patterns support Massachusetts’s claim that most students have ample opportunities to retake the examination.

 In Figure 5, we present the fitted probability of retaking the examination (top panel) and passing the first retest (bottom panel) as a function of initial mathematics test score.[[12]](#footnote-12) It shows that, among students with the same MCAS scores on the initial tests, low-income urban students are no less likely than suburban students to retake the mathematics examination. However, low-income urban students are nearly ten percentage points less likely to pass this retest than suburban students with the same initial scores (p<0.001).

FIGURE 5 ABOUT HERE

(3) Effect of failing the 8th grade examination on persistence to 10th grade

 Although the 8th grade examination does not carry high stakes for students, performance on the test is clearly related to the probability that students remain in school through 10th grade. For low-income urban students on the margin of passing the 8th grade mathematics test, failing the examination reduces the probability of continuing in school and taking the 10th grade MCAS examination by three percentage points (p=0.16). While this effect is not statistically significant, we arrive at nearly identical, but more precise and statistically significant results using a slightly larger bandwidth around the cut score. Because only eleven percent of low-income urban students who just pass the examination leave the system before 10th grade, this three percentage point decline is noteworthy. In Figure 6, we illustrate this pattern by plotting the fitted non-parametrically smoothed relationship between persistence to 10th grade and mathematics score for low-income urban students, indicating that the probability of persisting jumps at the cut score between Passing and Failing.

FIGURE 6 ABOUT HERE

 (4) Effect of failing the English language arts examination on high school graduation[[13]](#footnote-13)

 Inspecting raw data in Massachusetts suggests that the mathematics examination is a larger hurdle to on-time graduation than the ELA examination. Most students who failed the 10th grade ELA examination also failed the mathematics test, while among students who only failed one of the tests, three times as many failed mathematics as ELA. Examining the ELA examination proves interesting, however, because detected patterns differ from the mathematics results. Failing the 10th grade ELA examination does not reduce the probability of graduation for any students, including low-income urban students, on the margin of passing. We illustrate the relationship between ELA score and probability of graduation for typical low-income urban students in Figure 7. Here, the figure displays no discontinuous jump in the probability of graduating at the cut score, suggesting that failing the ELA examination does not affect students’ likelihood of on-time graduation.

FIGURE 7 ABOUT HERE

***V. Sensitivity Analysis***

 As discussed above, for regression discontinuity analyses to identify a causal effect of failing the MCAS examinations on student graduation, several assumptions must hold. First, the rule that determines whether a student has passed or failed the examination must be externally determined, and rigidly applied across all students, while all other observed and unobserved characteristics of the student must vary smoothly and continuously around the cut score. Second, the relationship linking the probability of graduation and test score must be estimated accurately in the immediate vicinity of the cut score. In this section, we address these two primary concerns and describe other sensitivity analyses that we conduct to assess the robustness of our results.

External Establishment of Cut Scores

 The state-established cut scores do indeed produce a sharp discontinuity in treatment and thus create an extremely plausible natural experiment. Because the raw score needed to pass the examination differs from year to year and is only calculated after students take examination, it seems highly unlikely that students could decide knowingly to fall just above, or just below, the cut score. Furthermore, the state imposes these performance labels strictly, so that any student with a score of 20 points on the 2004 administration of the 10th grade mathematics examination failed, while any student with a score of 21 points passed. We perform several tests to verify these assumptions and find no reason to doubt them. We provide a more detailed discussion in Papay, Murnane, & Willett (2008).

Accurate estimation of the relationship between graduation and MCAS mathematics score

For estimates of the treatment effect to be unbiased, we must predict credibly and precisely what the probability of graduation would have been for students who failed the MCAS mathematics examination if they had scored 21 points on the test and still failed. We address this issue by modeling the smooth relationship between the probability of graduation and test score non-parametrically, making few assumptions about the relationship between MCAS score and graduation. Here, our primary decision involves the choice of bandwidth, *h*, or the number of observations around the cut score that we use in estimation. To explore the sensitivity of our results to this decision, we vary the bandwidth systematically. In the top panel of Table 1, we present the fitted effects of failing the 10th grade mathematics examination on on-time graduation for each different bandwidth. In the middle and bottom panels of Table 1, we present parallel results for the effects of failing the 8th grade examination on persistence to 10th grade and for the effects of failing the 10th grade ELA examination on on-time high school graduation. Regardless of the bandwidth that we choose, our main results are unchanged – for urban, low income students, failing the 8th grade mathematics examination reduces the probability of persisting to 10th grade and failing the 10th grade mathematics examination reduces the probability of on-time graduation. However, we find no effects for other groups of students or for any group failing the 10th grade ELA examination. Our estimates for the effect of failing the 8th grade examination for marginal urban students range from 2.7 to 3.7 percentage points, and are quite insensitive to bandwidth choice. Our estimates of the effect of failing the 10th grade examination range from 5.8 to 13.1 percentage points. In all cases, this result is statistically significant. Finally, we fit our main models using logistic, rather than local linear, regression to account for the binary outcome. We find nearly identical results.

TABLE 1 ABOUT HERE

Definition of outcome variable

We choose to present our main analyses using on-time graduation as our primary outcome measure. However, one concern is that students who fail the MCAS may remain in school and graduate in subsequent years, or that they may drop out and earn a General Equivalency Diploma instead of graduating from high school. We find that our results are quite robust to the definition of our outcome. Here, we use three different outcome measures: graduated on-time or still enrolled in school; dropped out; and graduated on-time or obtained a GED. As seen in Table 2, in all cases we find statistically significant effects of passing the examination ranging from 7.2 to 9.1 percentage points.

TABLE 2 ABOUT HERE

***VI. Discussion***

 This paper addresses several important questions about the effects of the state accountability system on Massachusetts high school students. To put these effects in context, it is important to recall the evidence cited earlier. Under standards-based educational reforms, the average reading and mathematics performances of Massachusetts elementary school students have improved markedly. In 2007, the state’s reading and mathematics performances on the NAEP ranked first in the nation. Thus, we do not see the evidence that we present as an attack on the demonstrably successful educational reform effort in Massachusetts. Instead, we document unanticipated consequences of efforts to prepare all students to meet the demands of 21st century life. These consequences are important and need to be at the center of efforts to make standards-based reforms work for all Massachusetts students in the years ahead.

To recap, we find that, for low-income urban students on the margin of passing, failing the 8th grade mathematics examination reduces the probability of persisting to 10th grade by three percentage points, while failing the 10th grade examination reduces the probability of on-time graduation by eight percentage points. We find no effects of failing for wealthier urban students or suburban students. Again, these estimates are only valid for students at the margins of passing the examination, under the high-stakes testing regime in Massachusetts.

Importantly, we know nothing about whether these students are better (or worse) off than they would have been in the absence of standards-based reform. However, low-income, urban students with essentially the same proficiency on the state test have substantially different graduation outcomes simply because they are categorized as “passing” or “failing” the examination. This effect raises an important challenge for urban school districts. We also have no information about the extent to which the requirement to pass the MCAS affects the probability of on-time graduation for students well below the passing score. As a result, we cannot estimate how much of the state dropout rate for low-income urban youth is due to the imposition of the exit examination. However, because 60% of students who do not graduate on time actually pass the MCAS, failing the test is clearly only one of many factors that contribute to the dropout decision.

We see several complementary explanations for the finding that failing the 10th grade mathematics examination reduces the likelihood of graduation for urban students from low-income families, but not for more affluent or suburban students. First, we cannot distinguish whether just failing the examination causes these students to drop out or whether just passing it causes them to remain in school. Low-income urban students who pass may feel encouraged that they are doing well in school and may decide to persist to graduation. Similarly, students who pass may get more teacher attention or may be promoted more readily through school, leading to improved graduation outcomes.

On the other hand, low-income urban students who fail the examination may become discouraged or subject to institutional responses that reduce their likelihood of graduating on time. Families of low-income urban students may lack the resources to help them overcome the hurdle posed by failing the examination. Low-income urban students typically attend high schools in which many students have failed the 10th grade MCAS examinations. These schools are struggling to figure out how, with very limited resources, to respond to this problem. Finally, the interaction between school and home contexts may produce these effects. Interestingly, the different consequences for failing the ELA examination than for failing the mathematics examination suggest that urban schools may devote more resources to or be more successful at remediation in reading and writing than in mathematics.

That suburban students, including those from low-income families, appear to face no barrier from failing the 10th grade MCAS mathematics test suggests that their schools have found ways to support both low-income and wealthier students who have failed. These suburban schools typically have many fewer students who fail the examination, so they can afford to provide more personalized attention and remediation. In some Massachusetts districts, schools match teachers with students who failed the exit examination in order to provide one-on-one tutoring. In such an environment, it is not surprising that these students may in fact have more in-school adult contact and encouragement than students who just passed, and may in fact graduate at greater rates.

 That most students who fail the 10th grade mathematics examination retake it and that low-income urban students retake the test at similar rates as their wealthier urban or suburban peers are also encouraging. These findings suggest that these students are receiving the message that they should persist and retake the test. As a result, schools have time to work with these students and prepare them to meet the graduation requirements. However, low-income urban students are much less likely to pass this retest, even when comparing students with the same initial examination performance. Finding the explanation for this pattern is an important topic for research, with critical implications for improving equality of educational opportunity.

 Our findings raise several questions for researchers, educators, and policymakers in Massachusetts and other states. First, the absence of effects of high-stakes testing on high school completion for suburban students (including those from low-income families) suggests that it is possible to overcome the initial disappointment associated with failing a high-stakes examination. Learning more about the initiatives that improve student retention could be helpful for districts struggling to support many failing students. A related question that we intend to pursue in future work is whether some urban districts are more successful than others in supporting students who failed the 10th grade mathematics examination. If that is the case, then understanding the successful efforts of some urban districts might help others to improve their support to struggling mathematics learners.

Especially intriguing is the finding that marginally failing the high-stakes ELA examination does not reduce the probability that low-income urban 10th graders graduated on time, while marginally failing the mathematics examination does reduce the probability of on-time graduation. Why the difference? Do urban districts concentrate resources on programs to improve their low-income students’ ELA skills? Does the structure of the examinations make remediation easier in ELA than in mathematics for students on the border of passing?

Our finding that the *Fear of Failing* the 10th grade examination induces some low-income urban students to drop out before even taking it raises additional questions. Failing the 8th grade examination gives students some sense of their probable performance on the 10th grade test, but discerning students should recognize that scores on either side of the cutoff are not substantively different. Nonetheless, we find a moderate effect of failing on persistence to 10th grade for these very students. What is the mechanism at play here? Does the “failing” label affect a student’s self-concept? Do students pay attention only to the performance level that their score puts them in, not on how close they are to passing? Or, does this effect reflect school or parental responses, such as retaining students or removing them to private schools?

Another question concerns the extent to which the consequences of exit examinations depend on their content and format. The 10th grade MCAS mathematics test is relatively demanding compared to the exit examinations used by other states. Not only does it assess students’ skills in a range of topic areas, it does so with questions that contain relatively complex language. Also, some test items call for open-ended responses while others require students to explain their answers. Supporters of the Massachusetts examinations argue that good instruction in mathematics is the only way to prepare students to do well on the test, and that simply drilling students on released test items is not an effective way to improve MCAS scores. The payoff to drill, as opposed to good mathematics instruction, may vary among the examinations used by different states. This difference may influence the success of various remediation programs.

This research argues for the importance of examining heterogeneous effects. In future work, we hope to explore more fully the effects of failing on different groups of students, including those with limited English proficiency. It also raises the question of whether the types of differential impacts we observe in Massachusetts may also be present in other states, especially those that use relatively demanding exit examinations. A corollary is the importance of finding the explanations for any observed differential effects of exit examinations. Finding differences in the probability of retaking the examination between groups suggests one policy problem. Finding differences in success rates among those who do retake the examination, as we do, suggests a different problem. We need to understand more carefully what messages and remediation efforts low-income urban students are receiving that encourage them to retake the examination but do not prepare them for success. Finally, we wonder why the effect for urban students varies by income. Do wealthier students attend different schools, or do they receive additional support outside of school?

In summary, the requirement that high school students achieve passing scores on relatively rigorous state-administered examinations in order to obtain a high school diploma is a relatively new phenomenon in the United States. The content, format, and difficulty of such tests vary widely across states, as do opportunities for re-taking the examinations and support for those who fail. Future research needs to go beyond the question of whether failing a particular exit examination affects the probability of high school graduation. It needs to examine the extent to which the consequences of failing an exit examination depend on the attributes of the examination, the testing system, the student, and the quality of support available to struggling students.

*Figure 1*. Comparison of recent Massachusetts and nationwide National Assessment of Educational Progress scaled scores, for 8th grade mathematics from 1992 to 2008.



Figure 2. Diagrams illustrating the regression discontinuity design, with both a linear relationship between the assignment variable and treatment (top panel) and a curvilinear relationship (bottom panel). The difference between A and B represents the causal effect of treatment (failing MCAS). Here, X is the MCAS math score; Y is the probability of graduation.





 *Figure 3*. Fitted smoothed non-parametric relationship (bandwidth=6) between the probability of on-time graduation and 10th grade mathematics score for low-income urban students, with the sample mean probabilities of graduation overlaid.



We plot the non-parametric regression fit without student-level covariates. For data tables, see Papay, Murnane, & Willett, 2008.

*Figure 4*. Sample histogram presenting the frequencies of students who failed the 10th grade mathematics examination and who subsequently retook the examination, along with their performance on retest.



*Figure 5*. Fitted relationship (from Table 2) between the probability of retaking the examination (top panel) or passing the first retest (bottom panel) and initial 10th grade mathematics score for low-income urban students and suburban students who failed their first examination (plotted in the immediate region of the pass/fail cut-score for white female students not classified as special education or limited English proficient who just passed the ELA test) (n=8,225).





For data tables, see Papay, Murnane, & Willett, 2008.*Figure 6*. Fitted smoothed non-parametric relationship (bandwidth=6) between the probability of persisting to 10th grade and 8th grade mathematics score for low-income urban students, with the sample mean probabilities of graduation overlaid.



We plot the non-parametric regression fit without student-level covariates. For data tables, see Papay, Murnane, & Willett, 2008.

*Figure 7*. Fitted smoothed non-parametric relationship (bandwidth=8) between the probability of on-time high school graduation and 10th grade ELA score for low-income urban students, with the sample mean probabilities of graduation overlaid.



We plot the non-parametric regression fit without student-level covariates. For data tables, see Papay, Murnane, & Willett, 2008.

*Table 1*. Estimated causal impacts of failing the 10th grade mathematics, 8th grade mathematics, and 10th grade ELA examinations, for different bandwidths by subgroup, with standard errors in parentheses. Results for the optimal bandwidth, h\*, appear in bold.

|  |
| --- |
| **Panel I: 10th Grade Mathematics** |
|  | **Bandwidth (h)** |
| **Group** | **4** | **5** | **6** | **7** | **8** |
| Urban, Low Income | 0.131 | \*\* | 0.103 | \*\* | **0.080** | \* | 0.065 | \* | 0.058 | \* |
|  | (0.041) |  | (0.036) |  | **(0.033)** |  | (0.031) |  | (0.029) |  |
| Urban, Not Low Income | -0.025 |  | -0.024 |  | **-0.052** |  | -0.007 |  | -0.007 |  |
|  | (0.067) |  | (0.059) |  | **(0.054)** |  | (0.050) |  | (0.046) |  |
| Suburban, Low Income | -0.050 |  | -0.022 |  | **0.023** |  | 0.009 |  | 0.003 |  |
|  | (0.052) |  | (0.045) |  | **(0.042)** |  | (0.038) |  | (0.036) |  |
| Suburban, Not Low Income | 0.000 |  | -0.018 |  | **-0.015** |  | -0.016 |  | -0.027 |  |
|  | (0.034) |  | (0.030) |  | **(0.027)** |  | (0.025) |  | (0.024) |  |
| **Panel II: 8th Grade Mathematics** |
|  | **Bandwidth (h)** |
| **Group** | **4** | **5** | **6** | **7** | **8** |
| Urban, Low Income | 0.032 |  | 0.029 |  | **0.027** |  | 0.037 | \* | 0.034 | \* |
|  | (0.024) |  | (0.021) |  | **(0.019)** |  | (0.018) |  | (0.017) |  |
| Urban, Not Low Income | 0.047 |  | 0.027 |  | **0.048** |  | 0.029 |  | 0.013 |  |
|  | (0.036) |  | (0.031) |  | **(0.028)** |  | (0.026) |  | (0.025) |  |
| Suburban, Low Income | 0.011 |  | 0.012 |  | **-0.002** |  | 0.008 |  | 0.014 |  |
|  | (0.024) |  | (0.022) |  | **(0.020)** |  | (0.018) |  | (0.017) |  |
| Suburban, Not Low Income | -0.008 |  | -0.003 |  | **-0.008** |  | -0.010 |  | -0.013 |  |
|  | (0.011) |  | (0.010) |  | **(0.009)** |  | (0.008) |  | (0.008) |  |
| **Panel III: 10th Grade ELA** |
|  | **Bandwidth (h)** |
| **Group** | **6** | **7** | **8** | **9** | **10** |
| Urban, Low Income | -0.002 |  | 0.006 |  | **0.011** |  | 0.019 |  | 0.010 |  |
|  | (0.039) |  | (0.036) |  | **(0.034)** |  | (0.032) |  | (0.030) |  |
| Urban, Not Low Income | -0.121 |  | -0.090 |  | **-0.052** |  | -0.023 |  | 0.023 |  |
|  | (0.076) |  | (0.072) |  | **(0.067)** |  | (0.063) |  | (0.060) |  |
| Suburban, Low Income | -0.046 |  | -0.031 |  | **-0.016** |  | 0.003 |  | 0.006 |  |
|  | (0.052) |  | (0.048) |  | **(0.044)** |  | (0.042) |  | (0.007) |  |
| Suburban, Not Low Income | -0.017 |  | -0.001 |  | **0.022** |  | 0.027 |  | 0.032 |  |
|  | (0.038) |  | (0.035) |  | **(0.033)** |  | (0.031) |  | (0.030) |  |

Notes: \* p<0.05, \*\* p<0.01, \*\*\* p<0.001.

*Table 2*. Parameter estimates, standard errors, and approximate *p*-values at the cut score from the non-parametric regression analysis of the effect of failing the 10th grade mathematics examination on three different graduation outcomes, for urban, low-income students (from the single regression centered at the cut score with bandwidth *h\**).

|  |  |  |  |
| --- | --- | --- | --- |
| **Predictor** | **Graduated or still enrolled** | **Dropped Out** | **Graduated or earned GED** |
| Intercept | 0.660 | \*\*\* | 0.273 | \*\*\* | 0.607 | \*\*\* |
|  | (0.031) |  | (0.027) |  | (0.033) |  |
| MATH | 0.015 | \* | 0.002 |  | 0.025 | \*\*\* |
|  | (0.007) |  | (0.006) |  | (0.007) |  |
| PASS | 0.091 | \*\* | -0.072 | \*\* | 0.081 | \* |
|  | (0.030) |  | (0.026) |  | (0.033) |  |
| PASSxMATH | -0.024 | \*\* | 0.001 |  | -0.031 | \*\*\* |
|  | (0.008) |  | (0.007) |  | (0.009) |  |
| African-American | 0.117 | \*\*\* | -0.089 | \*\*\* | 0.059 | \*\* |
|  | (0.020) |  | (0.017) |  | (0.022) |  |
| Asian-American | 0.081 | \* | -0.067 | \* | 0.061 |  |
|  | (0.035) |  | (0.030) |  | (0.038) |  |
| Hispanic | 0.039 |  | -0.054 | \*\* | -0.011 |  |
|  | (0.021) |  | (0.018) |  | (0.021) |  |
| Mixed/Other Race | 0.241 | \*\*\* | -0.152 | \*\*\* | 0.204 | \*\* |
|  | (0.054) |  | (0.044) |  | (0.068) |  |
| Native American | -0.081 |  | -0.139 |  | -0.086 |  |
|  | (0.135) |  | (0.072) |  | (0.138) |  |
| Pacific Islander | 0.145 |  | -0.015 |  | -0.208 |  |
|  | (0.201) |  | (0.189) |  | (0.261) |  |
| Limited English Proficient | 0.027 |  | -0.020 |  | 0.007 |  |
|  | (0.022) |  | (0.018) |  | (0.024) |  |
| Special Education | -0.003 |  | -0.012 |  | -0.028 |  |
|  | (0.019) |  | (0.016) |  | (0.020) |  |
| Female | 0.031 | \* | -0.039 | \*\* | 0.073 | \*\*\* |
|  | (0.015) |  | (0.012) |  | (0.016) |  |
| New Student | -0.071 | \*\* | 0.003 |  | -0.050 | \* |
|  | (0.023) |  | (0.018) |  | (0.024) |  |
| R2 | 0.036 | 0.018 | 0.045 |
| Bandwidth (h\*) | 6 | 6 | 6 |
| N | 3469 | 3469 | 3469 |

Notes: \* p<0.05, \*\* p<0.01, \*\*\* p<0.001.

**References**

Carnoy, M. (2005). Have state accountability and high-stakes tests influenced student progression rates in high school? *Educational Measurement: Issues and Practice*. Winter, 19-31.

Carnoy, M. & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, *24*(4), 305-331.

Center on Education Policy. (2007). *“It’s different now”: How exit examinations are affecting teaching and learning in Jackson and Austin.* Retrieved April 30, 2007, from [http://www.cep-dc.org/highschoolexit/JacksonAustin/Jackson&Austin.pdf](http://www.cep-dc.org/highschoolexit/JacksonAustin/Jackson%26Austin.pdf)

Clarke, M., Haney, W. & Madaus, G. (2000). High stakes testing and high school completion. *National Board on Educational Testing and Public Policy Statement, 1*(3). Retrieved April 30, 2007, from <http://www.bc.edu/research/nbetpp/publications/v1n3.html>.

Cornell, D.G., Krosnik, J.A. & Chang, L. (2006). Student reactions to being wrongly informed of failing a high-stakes test: The case of the Minnesota Basic Standards test. *Educational Policy, 20*(5), 718-751.

Dee, T.S. & Jacob, B.A. (2006). Do high school exit exams influence educational attainment or labor market performance? Cambridge, MA: NBER Working Paper 12199. Retrieved October 28, 2007, from <http://www.nber.org/papers/w12199>.

Evers, W.M., & Walberg, H.J., eds. (2002). *School accountability*. Stanford, CA: Hoover Institution Press.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association, 87*(420): 998-1004.

Finn, C.E., Julian, L., & Petrilli, M.J. (2006). *The state of state standards.* Washington, D.C.: The Fordham Foundation. Retrieved March 26, 2008 from <http://www.edexcellence.net/foundation/publication/publication.cfm?id=358>.

Greene, J.P. & Winters, M.A. (2004). *Education working paper: Pushed out or pulled up? Exit examinations and dropout rates in public high schools.* New York: Center for Civic Innovation at the Manhattan Institute for Policy Research.

Griffin, B.W. & Heidorn, M.H. (1996). An examination of the relationship between minimum competency test performance and dropping out of high school. *Educational Evaluation and Policy Analysis*, 18(3), 243-252.

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica, 69*(1): 201-209.

Hanushek, E.A., & Rivkin, S.G.. (2006). School quality and the black-white achievement gap. Cambridge, MA: NBER Working Paper 12651. Retrieved June 26, 2008, from <http://www.nber.org/papers/w12651>.

Heubert, J.P. & Hauser, R.M., eds. (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press, 1999.

Imbens, G. & Lemieux, T. (2007). Regression discontinuity designs: A guide to practice. Cambridge, MA: NBER Working Paper 13039. Retrieved October 28, 2007, from <http://www.nber.org/papers/w13039>.

Jacob, B.A. (2001). Getting tough? The impact of high school graduation exams. *Educational Evaluation and Policy Analysis, 23*(2), 99-121.

Jones, M.G., Jones, B.D., & Hargrove, T.Y. (2003). *The unintended consequences of high-stakes testing*. Lanham, MD: Rowman & Littlefield Publishers, Inc.

Levy, F. & Murnane, R. (2004). *The new division of labor: How computers are creating the next job market.*  Princeton, N.J.: Princeton University Press; New York: Russell Sage Foundation.

Lleras-Muney, A. (2004). The relationship between education and adult mortality in the United States. *Review of Economic Studies,* 72(1).

Lochner, L., & Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American Economic Review,* 94(1), 155-189.

Martorell, F. (2005). Does failing a high school graduation exam matter? Unpublished working paper: Author.

Massachusetts Department of Education. (2002). *2001 MCAS technical report.* Retrieved April 30, 2007, from <http://www.doe.mass.edu/mcas/2002/news/01techrpt.pdf>.

Massachusetts Department of Education. (2005). *2004 MCAS technical report.* Retrieved April 30, 2007, from <http://www.doe.mass.edu/mcas/2005/news/04techrpt.pdf>.

National Center for Education Statistics. (2008). State comparisons: National Assessment of Educational Progress (NAEP). Washington, DC: U.S. Department of Education. Retrieved April 5, 2008 from <http://nces.ed.gov/nationsreportcard/nde/statecomp/>

Neal, D. & Schanzenbach, D.W. (2007). Left behind by design: Proficiency counts and test-based accountability. Unpublished working paper: Author.

Nichols, S.L., Glass, G.V, & Berliner, D.C. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives, 14*(1). Retrieved April 30, 2007, from <http://epaa.asu.edu/epaa/v14n1/>.

Oreopoulos, P. (2007). Do dropouts drop out too soon? Wealth, health, and happiness from compulsory schooling. *Journal of Political Economy,* 91: 2213-2229.

Papay, J.P., Murnane, R.J., & Willett, J.B. (2008). The consequences of high school exit examinations for struggling low-income urban students: Evidence from Massachusetts. Cambridge, MA: NBER Working Paper 14186. Retrieved July 28, 2008, from <http://www.nber.org/papers/w14186>.

Quality Counts. (2006). Quality counts at 10: A decade of standards-based education. *Education Week, 25*(17): 74.

Reardon, S.F. & Galindo, C. (2002). *Do high-stakes tests affect students’ decisions to drop out of school? Evidence from NELS.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin Company.

Thomas, R.M. (2005). *High stakes testing: Coping with collateral damage.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Warren, J.W., Jenkins, K.N., & Kulick, R.B. (2006). High school exit examinations and state-level completion and GED rates, 1975 through 2002. *Educational Evaluation and Policy Analysis, 28*:2, 131-152.

1. The authors thank Carrie Conaway, the Director of Planning, Research, and Evaluation of the Massachusetts Department of Elementary and Secondary Education, for providing the data and for answering many questions about data collection procedures. Participants in the May 1, 2008 NBER economics of education workshop provided helpful comments. Financial support was provided by the U.S. Department of Education Institute for Education Sciences (Grant Number R305A080127) and the Harvard Graduate School of Education Dean's Summer Fellowship. The views expressed herein are those of the authors and do not necessarily reflect the views of the Massachusetts Department of Elementary and Secondary Education. [↑](#footnote-ref-1)
2. This same Fordham Foundation report, *The State of State Standards 2006*, pointed out that the Massachusetts standards were exceptional. In contrast, “two-thirds of schoolchildren in America attend class in states with mediocre (or worse) expectations for what their students should learn” (Finn, Julian, & Petrilli, 2006). [↑](#footnote-ref-2)
3. The state has a performance appeals process in place that allows students to demonstrate their proficiency in alternate ways. It also offers alternative assessments to certain students. Only 314 of the state’s 57,000 graduates in 2006 satisfied the requirement using either of these two alternative routes. [↑](#footnote-ref-3)
4. The state identifies slightly fewer than 3,000 students (less than 5% of the total sample) who are not in the “final 2006 cohort,” meaning that they moved out of the state before high school graduation. Using only the 63,361 individuals in the “final cohort” does not alter our results. We include the full sample to account for any effects the high-stakes examination has on student mobility. [↑](#footnote-ref-4)
5. Technically, students took the middle school ELA examination in 7th grade and the mathematics examination in 8th grade. For simplicity, we refer to these examinations as the “8th grade” tests. [↑](#footnote-ref-5)
6. For more information on MCAS scoring and scaling, see the MCAS Technical Reports (MA DOE, 2002, 2005). [↑](#footnote-ref-6)
7. For the 8th grade mathematics test, students had to score 22 points to pass, and for the 10th grade ELA examination the minimum passing score was 39. [↑](#footnote-ref-7)
8. Some of these students moved into the state after 8th grade, while others simply had missing 8th grade test scores. Because we cannot distinguish between these two groups, we cannot interpret this variable as a pure indicator of new students to the state. [↑](#footnote-ref-8)
9. For a more detailed description of the regression discontinuity approach see Shadish, Cook, & Campbell (2002). [↑](#footnote-ref-9)
10. We estimate robust (Huber-White) standard errors to account for both the clustering of students within schools and heteroscedasticity in the dichotomous outcome. [↑](#footnote-ref-10)
11. The authors thank Steven Rivkin for pointing out this issue and for his helpful suggestions for addressing it. [↑](#footnote-ref-11)
12. In the bottom panel, we use predicted math score because of the IV approach used for this analysis. [↑](#footnote-ref-12)
13. Because the middle school ELA test for the 2006 cohort occurred in 7th grade, one year earlier than the mathematics test, the state data system, which began in 2001, cannot match students as accurately for this test. As a result, we cannot examine the effects that *Fear of Failing* the ELA examination may have on persistence to 10th grade. [↑](#footnote-ref-13)