PRELIMINARY DRAFT: DO NOT QUOTE OR CITE

The Consequences of High School Exit Examinations for Struggling Urban Students: Evidence from Massachusetts

John P. Papay
Richard J. Murnane
John B. Willett

Harvard Graduate School of Education*

April, 2008

* The authors thank Carrie Conaway, the Director of Planning, Research, and Evaluation of the Massachusetts Department of Education, for providing the data and for answering many questions about data collection procedures. 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.

PRELIMINARY DRAFT: DO NOT QUOTE OR CITE

Abstract

The growing prominence of high-stakes exit examinations has made questions about their effects on student outcomes increasingly important. Exit examinations can cause students to drop out of school for several reasons: because they fear taking the test, because they fail it and become discouraged, or because they repeatedly retake the examination and cannot pass it. We use a natural experiment, with a regression discontinuity design, to evaluate the causal effects of high-stakes testing on high school completion for the cohort scheduled to graduate from Massachusetts high schools in 2006. We find that, for low-income urban students on the margin of passing, failing the 10^{th} grade mathematics examination reduces the probability of on-time graduation by approximately eight percentage points. Among students who fail the 10^{th} grade mathematics examination, we find that low-income urban students are just as likely to retake the test as apparently equally skilled suburban students, but they are much less likely to pass this retest. Furthermore, failing the 8^{th} grade mathematics examination reduces by three percentage points the probability that low-income urban students stay in school through 10^{th} grade. We find no such effects for wealthier urban or suburban students, regardless of family income.

PRELIMINARY DRAFT: DO NOT QUOTE OR CITE

The Consequences of High School Exit Examinations for Struggling Urban Students: Evidence from Massachusetts

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, including greater probability of employment, higher lifetime earnings and better health (Bureau of Labor Statistics, 2007; Haveman & Wolfe, 1984; Card, 1999), the question of how high-stakes testing affects high school completion rates is important to educational policymakers.

We capitalize on a natural experiment, using a regression discontinuity design, to examine the causal impact of failing the statewide 10th grade mathematics examination on the probability of on-time high school graduation. Our data come from Massachusetts, a state that has earned a national reputation for rigorous content standards and English Language Arts (ELA) and mathematics assessments that are well aligned to the standards.

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 affects students.

We find that failing the Massachusetts mathematics examination at the end of the 10th grade reduces the probability that low-income, urban students on the margin of passing will graduate on-time by eight percentage points. In contrast, failing the test does not reduce 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. 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. 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 findings that explore possible sources of these effects for urban students with low family income. Students who fail have many opportunities to retake the examination before graduation, with some students completing six retests. However, very few students exhaust all of these retest opportunities. Instead, discouragement, as opposed to repeated failure, is the primary reason why failing the 10th grade examination prevents some low-income urban students from graduating. Furthermore, we find that, 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 explore several threats to our study's validity. 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 10^{th} 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. In April 2007, the *Bureau of Labor Statistics* reported an unemployment rate of 7.2% for individuals with less than a high school diploma, compared with 4.1% for high school graduates. Of those individuals who do work, high school graduates earn substantially more: in 2005, the median high school graduate, age 25-34, earned 27% more than the median high school dropout.[1]

High school graduation also matters because a diploma represents almost a necessary prerequisite for college enrollment, and a college education contributes markedly to subsequent labor market outcomes. College completion (both at 2-year and 4-year institutions) is associated with a substantial earnings premium. Workers aged 25-34 with an associate's degree earn, on average, 29% more than high school graduates; bachelor's degree recipients earn 58% more.[2] This evidence suggests that 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).

---

[1] We define high school dropouts as students who have at least some high school education. Statistics for high school and college graduate earnings come from the U.S. Census Bureau *Current Population Survey*, 2006 Annual Social and Economic Supplement.

[2] While these statistics could simply reflect "ability bias" if more able students, who would have earned more regardless of schooling, persist in school longer, a great deal of economic research has found substantial returns to education after accounting for this "ability bias" (see Card, 1999).

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. Neither line of inquiry has yielded definitive answers about the effects of exit examinations.

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 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 millions of dollars in standards-based educational reforms. These investments have 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).[3] 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, the percentages of Massachusetts 4th grade students that have scored at the Proficient or higher level on the National Assessment of Educational Progress (NAEP) examinations have increased 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). 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.

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 for all students in that they 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

---

[3] 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).

fail, attempting explicitly to make the MCAS as minimal a barrier to graduation as possible.[4] 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

[4] 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.

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.[5] 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 a "pass" or a "fail" condition. This assignment process would render them equivalent in expectation on all observable and unobservable characteristics prior to treatment, allowing us to identify any differences in the ultimate outcome (high school graduation) as a causal effect of simply failing the examination, rather than of earning lower scores. Such an experiment is, of course, both impossible and unethical. However, we can take advantage of the state's exogenous imposition of a minimum passing score to provide a natural experiment from which we can draw equivalent causal conclusions. By examining students with nearly identical MCAS performance, but just on either side of this exogenously-assigned 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.

<u>Research Questions</u>

As mentioned above, exit examinations can increase dropout rates in three ways: through *Fear* (students predict that they will not pass and drop out before taking the test); through *Discouragement* (students give up and drop out after failing one or more of the examinations); and through *Repeated Failure* (after exhausting the available retest opportunities, students still

[5] For scores above the passing standard, the estimated correlation between the raw MCAS mathematics score and the proportion of students who graduate on time is 0.965, suggesting a very strong positive linear relationship between MCAS performance and probability of on-time high school completion.

have not satisfied the testing requirements). In our research, we explore the relative importance of these mechanisms, paying particular attention to impacts on low-income urban students. 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 for whom failing the 10th grade exit examination does pose a barrier to high school graduation, 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 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 10^{th} grade examination (our first research question), we focus on members of the 2006 graduating cohort who first took the 10^{th} 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 8^{th} and 10^{th} grade and those who missed the 8^{th} grade examination, but it does not include students who entered the state after 10^{th} grade. Our final sample for addressing the first research question includes 66,347 students.[6] To address our third research question, we use the 69,127 students in the same cohort who took the 8^{th} grade mathematics examination. This sample includes students who dropped out of school before 10^{th} 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 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. To address our third research question, we created another dichotomous outcome measure, named *TAKE10th*, that indicates whether a student who took the 8^{th} grade mathematics examination persisted in school to take the 10^{th} grade test (1=persisted to take the test; 0 otherwise).

The dataset contains a record of scores from every MCAS mathematics and ELA

[6] 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.

examination that each student took from 8th grade[7] 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.[8] For the 10th grade mathematics examination, raw scores ranged from 0 to 60; students who earned more than 20 points passed the test.[9] To implement our regression discontinuity approach, we centered students' raw scores by subtracting out the value of the corresponding minimum passing score. The values of the re-centered continuous predictors, *MATH* and *ELA*, were such that 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*).[10] 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

---

[7] 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.
[8] For more information on MCAS scoring and scaling, see the MCAS Technical Reports (MA DOE, 2002, 2005).
[9] For the 8th grade mathematics test, students had to score 22 points to pass, and for the 10th grade ELA examinations the minimum passing score was 39.
[10] 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.

tended to cluster in urban schools: 63% of urban students lived in poverty, compared to just 16% of suburban students.

We created several additional variables for use in the descriptive analyses that we used to address our second research question. For students who failed the examination, we created a vector of mutually exclusive dichotomous variables that indicated whether the student retook the test and passed it on the second try (*RETAKE_PASS*), retook the test and failed (*RETAKE_FAIL*), or never retook the test (*QUIT*).

Data Analyses

We address our first and third research questions by conducting identical regression discontinuity 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 the regression discontinuity strategy first proposed by Thistlethwaite & Campbell (1960) – to make such causal inferences *for students at the margins of passing*.[11] Because the probability that a student will pass the examination goes unequivocally from zero to one at a single cut score, our discontinuity is sharp.

[11] For a more detailed description of the regression discontinuity approach see Shadish, Cook, & Campbell (2002).

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 population relationship between MCAS score and graduation is continuous around the cut score, except for the impact of passing/failing the examination itself. Second, we assume that we can model this continuous relationship credibly and precisely 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 score is disrupted at the cut score. If so, the magnitude of the discontinuity in the outcome provides an unbiased estimate of the causal impact of failing the examination for students at the cut score. Thus, we obtain an estimate of the average treatment effect for students on the margin of passing.

Our key challenge thus involves estimating accurately the probability of graduation for students immediately on either side of the cut score. We estimate the effect of failing the examination as a difference in the probability of on-time graduation between students scoring at the cutoff who just passed (γ_{pass}) and just failed (γ_{fail}).[12] In our analyses, we use observations above the cut score to estimate γ_{pass} and observations below the cut score to estimate γ_{fail}. Because we do not know the precise functional form of the relationship between MCAS score and the probability of graduation, we model this continuous relationship using a nonparametric smoothing process to estimate γ_{pass} and γ_{fail}. A further complication arises as our parameters of

[12] Technically, $\gamma_{pass} = \lim_{MATH_i \to 0^+} \left[P(GRAD_i = 1) \mid MATH_i\right]$ and $\gamma_{fail} = \lim_{MATH_i \to 0^-} \left[P(GRAD_i = 1) \mid MATH_i\right]$

interest – γ_{pass} and γ_{fail} – are estimated at boundary points. As standard nonparametric smoothing strategies have poor boundary properties, Hahn, Todd, & Van der Klaauw (2001) recommend estimating these limits with local linear regression.[13]

Our implementation of nonparametric smoothing using local linear regression follows closely the recommendations of Imbens and Lemieux (2007).[14] We conduct our nonparametric smoothing within a linear probability specification of the standard regression discontinuity design. Specifically, at each MCAS score point, we estimate a linear regression function using only observations within a narrow bandwidth, *h*, around the point to predict the probability of graduation for each observation. As we move this bandwidth through our data range, we therefore generate locally predicted values at each MCAS score point; linking these estimates together creates the requisite smoothed nonparametric regression line. Here, the extent of the smoothing depends on the choice of bandwidth, *h*. Because we can only make causal claims about the effect of failing for students at the cut score, in our later analyses we focus attention on the single locally-linear regression analysis that centers on the cut score and estimates γ_{pass} and γ_{fail}. In this regression, then, we use only observations within bandwidth *h* on either side of the cut score, as follows:[15]

$$p(GRAD_i = 1) = \beta_0 + \beta_1 MATH_i + \beta_2 PASS_i + \beta_3 (PASS_i \times MATH_i) + \varepsilon_i \qquad (1)$$

for the i^{th} individual. While our nonparametric smoothing approach does not, by definition,

[13] Fan (1992) shows that, unlike most nonparametric smoothing techniques, local linear regression does not require boundary modifications.
[14] Ludwig and Miller (2007) use a similar strategy. Our approach differs in our choice of a rectangular rather than a triangular kernel for the non-parametric smoothing; however, Imbens & Lemieux (2007) argue that "more sophisticated kernels rarely make much difference" (p. 16) and instead recommend assessing robustness to different bandwidth choices, as we do in Section V.
[15] We estimate robust (Huber-White) standard errors to account for both the clustering of students within schools and heteroscedasticity in the dichotomous outcome.

return parameter estimates,[16] we can interpret the estimates from this single locally-linear fit in (1); these estimates represent the instantaneous slopes and intercepts for students at the cut score. In this model, parameter $\beta_2 = \gamma_{pass} - \gamma_{fail}$ represents the causal effect of passing the 10th grade MCAS mathematics examination on the population probability of on-time high school graduation for students at the cut score. 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 to *increase* discontinuously.

Our nonparametric procedure requires that we choose a suitable bandwidth, *h*, for the smoothing procedure and consequently for defining the region around the discontinuity in which we fit and interpret the model in (1). In our analyses, we select an optimal bandwidth, *h**, using all of our data by applying the cross-validation procedure described by Imbens & Lemieux (2007). Essentially, this procedure determines the bandwidth that minimizes the mean squared error in the predicted boundary points, leading to an optimal tradeoff of bias and precision for the estimation of γ_{pass} and γ_{fail}.[17] In our analyses, we obtain an optimal bandwidth of between four and six raw score points depending on the model specification, as indicated below. However, in sensitivity analyses in Section V, we re-present our main results using several different bandwidths to show that our main conclusions are robust to the choice of bandwidth.

[16] For example, the overall relationship between MCAS score and probability of graduation cannot be represented by a single slope throughout the data range.

[17] In other words, we determine a predicted probability of graduation $(\hat{GRAD}_i(h))$ for each observation *i* using only observations within *h* points to the left of $MCAS_i$ for students who failed and to the right of $MCAS_i$ for students who passed the examination. We determine the mean squared error of these predictions across the entire sample. We then systematically vary the bandwidth, *h*, choosing as *h** the value of *h* that minimizes this mean squared error. More formally, *h** = arg $\min_h \frac{1}{N}\sum_{i=1}^{N}(\hat{GRAD}_i(h) - GRAD_i)^2$. Because our ultimate objects of interest are the parameter estimates at the cut score, Imbens & Lemieux recommend excluding observations in the tails from the cross-validation determination. As data are less dense in the tails, including these observations may lead to over-smoothing. As a result, we eliminate the 10% of the observations on either side of, and most remote from, the cutoff.

We extend this simple model in several ways. First, we include a vector of selected student background covariates ($\mathbf{X_i}$) to improve precision and to eliminate small sample biases that result from including observations not immediately at the cut score (Imbens & Lemieux, 2007). Second, because our primary outcome is a dichotomous predictor that indicates whether the student graduates from high school on time, we replicate our analysis by specifying the probability of on-time high-school graduation as a logistic function of predictors. Here, we limit our analysis to those observations that fall within a narrow window around the cut score. For consistency with our earlier nonparametric smoothing, we choose a window whose width extends the optimal bandwidth of h^* on either side of the cut score. Again, we systematically vary this window width in Section V in order to test the robustness of our findings.[18]

Finally, we also examine the impact of test failure on high school graduation for particular groups of students, including urban students from low-income families. To conduct these latter analyses, we add all possible interactions between predictors $PASS_i$, $MATH_i$, $LOWINC_i$, and $URBAN_i$, up to and including the four-way interaction among the predictors, to our regression equation in (1). Again, our main results here derive from a single local linear regression analysis that incorporates only observations within an optimal bandwidth, h^*, on either side of the cut-off. The presence of the additional interaction terms permits the hypothesized relationship between graduation and mathematics score to differ for urban and suburban students from both low-income and wealthier families, above and below the cut score. While we find nearly identical results fitting separate regressions for each subgroup, our preferred approach pools data on the subgroups into a single analysis and provides slightly more

[18] In preliminary analyses, we investigated whether higher-order non-linear polynomial specifications of *MATH* score were required within the logistic model, including quadratic and cubic polynomial specifications. These specifications did not lead to improvements in model fit, within the narrow regression discontinuity window that we have selected for the analysis, and so we present results from the more parsimonious linear specification here.

precise estimates.

We focus on the statistically significant impact of just passing/failing the MCAS mathematics test on high school graduation for low-income urban youth for two reasons. 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 10^{th} 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 10^{th} 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 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. For these latter analyses, we fit logistic regression models of the following form on the sample of students who failed the 10^{th} grade mathematics examination:

$$\begin{aligned} &Log-Odds(RETAKE_i = 1) = \beta_0 + \beta_1(URBAN_i \times LOWINC_i) + \beta_2 URBAN_i + \beta_3 MATH_i + \\ &\beta_4 ELA_i + \beta_5 PASS\_ELA_i + \tilde{\boldsymbol{\alpha}}' \mathbf{X}_i \end{aligned} \tag{2}$$

for the i^{th} student. In preliminary analyses, we found that low-income and wealthier suburban students were indistinguishable from one another in terms of their probability of retaking the examination or of passing their first retest. By omitting the main effect of dichotomous predictor

$LOWINC_i$ from the hypothesized model, we implicitly treat all suburban students, regardless of family income, as the reference group. In our model, our principal research interest focuses on the parameter sum, $\beta_1+\beta_2$, which represents the difference between low-income urban students and suburban students in the log-odds 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.

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=0.018). 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. In Table 1, we present parameter estimates and approximate *p*-values from our local regression analyses using observations that fall within our "optimal" window of *h** on either side of the cut score. Models 1a and 1b present our findings for all students from equation (1), both with and without time-invariant student demographic controls; Models 2a and 2b represent our preferred specification that distinguishes effects for different subgroups, again with and without time-invariant student demographic controls.

TABLE 1 ABOUT HERE

To interpret the estimates presented in Table 1 more easily, we present the fitted nonparametrically smoothed relationship between graduation and MCAS mathematics score for typical low-income urban students from our preferred specification in Figure 1.[19] 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 1 ABOUT HERE

In the full model, 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.

<u>(2) Persistence and success in retesting among students who fail</u>

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.

---

[19] We can also recover the fitted relationship between graduation and MCAS mathematics score for the three other categories of students (wealthier urban, low-income suburban, and wealthier suburban). However, as our analyses show no effects on these groups, we decide to focus on the relationship for low-income urban students.

As the sample histogram in Figure 2 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. Furthermore, relatively few students in our regression discontinuity sample failed the March 2006 examination, the last retest before the cohort's graduation. Approximately 55% of both low-income urban and other students in this sample who took this test passed it. As a result, we find that *Repeated Failure* does not account for the different effects for low-income urban students. Instead, these students appear to become discouraged after failing and drop out.

FIGURE 2 ABOUT HERE

These retesting patterns support the Massachusetts Department of Education's claim that most students have ample opportunities to retake the examination. They also provide evidence that allows us to distinguish between two mechanisms that may reduce the probability of graduation for low-income urban students who fail the MCAS mathematics examination as 10th graders. Table 2 includes parameter estimates and approximate *p*-values from fitting the models specified in equation (2) to predict the probability that students who failed the 10th grade mathematics examination retake and pass the first retest. In Figure 3, 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. 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

($p<0.001$).

TABLE 2 ABOUT HERE

FIGURE 3 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. We present results in Table 3 from a regression discontinuity analysis of this outcome. 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 ($p=0.12$). While this effect is not statistically significant in the model estimated with optimal bandwidth, we arrive at nearly identical, but more precise and statistically significant results using a slightly larger bandwidth. 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 4, we illustrate this pattern by plotting the fitted nonparametrically 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.

TABLE 3 ABOUT HERE

FIGURE 4 ABOUT HERE

(4) Effect of failing the English language arts examination on high school graduation[20]

Inspecting raw data in Massachusetts suggests that the mathematics examination is a

[20] 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.

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. In Table 4, we present parameter estimates and approximate *p*-values from our local linear regression analyses, again using only observations that fall within our "optimal" window, centered on the cut score. We illustrate the relationship between ELA score and probability of graduation for typical low-income urban students in Figure 5. 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.

TABLE 4 ABOUT HERE

FIGURE 5 ABOUT HERE

V. Threats to Validity

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 exogenous 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 other threats to the internal validity of our study.

Exogenous Establishment of Cut Scores

The cut scores established by the Massachusetts Department of Education serve as an extremely plausible source of exogenous variation and do indeed produce a sharp discontinuity in treatment. Because the state's scoring and scaling of the MCAS examinations use complicated psychometric procedures, the raw score needed to pass the examination differs from year to year, based on the student population being assessed and the selection of the test items specifically included on the test form. Thus, it seems highly unlikely that students could decide knowingly to fall just above, or just below, the cut score. Furthermore, the state DOE imposes these performance labels strictly, so that any student with a score of 20 points *fails* the examination, while any student with a score of 21 points *passes*. Thus, the discontinuity is both exogenous and sharp.[21]

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. We address this issue by modeling the smooth relationship between the probability of graduation and test score nonparametrically, using a local linear regression approach. Here, our primary specification decision then involves the choice of bandwidth, *h*. Our preferred models use optimal bandwidths chosen through the cross-validation procedures described above.

[21] We performed several additional tests to verify the exogeneity of the MCAS cut score, as recommended by Imbens & Lemieux (2007). We examined a histogram of the 10th grade mathematics scores to explore continuity around the cut score. We find that 899 students just failed the exam, while 900 just passed it. We also examined histograms of other covariates not affected by the examination to identify any apparent discontinuities around the cut score and found none. Finally, we split our sample into students who passed and students who failed in order to estimate effects at "pseudo-discontinuities" declared at the median mathematics scores of these subsamples. In all cases, we find no reasons to doubt the robustness of our results.

To explore the sensitivity of our results to differences in bandwidth selection, we vary the bandwidth systematically, refitting our principal smoothed nonparametric models in each case. In the top panel of Table 5, we present the fitted effects of failing the 10th grade mathematics examination on on-time graduation for each subgroup as a function of different bandwidths. In the middle and bottom panels of Table 5, 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 bandwidth, 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 2.9 to 3.9 percentage points, and are quite insensitive to bandwidth. Our estimates of the effect of failing the 10th grade examination range from 5.7 to 12.6 percentage points. In all cases, we reject the null hypothesis that the parameter value is zero.

TABLE 5 ABOUT HERE

Finally, we explore the sensitivity of the results to the choice of functional form of the relationship between probability of on-time graduation and MCAS mathematics score. As an alternative to our smoothed nonparametric specification, we fit logistic regression models that incorporate only observations in selected narrow "windows" around the MCAS cutoff. The top panel of Table 6 contains the critical predicted logistic regression coefficients and standard errors from models in which we estimate the impact of failing the 10th grade mathematics examination on the probability of on-time graduation. To facilitate interpretation, the bottom panel contains estimates in probability units of the causal impact of failing on the fitted probability of on-time graduation for a typical student. The results from the logistic regression analysis mirror almost identically those provided by our nonparametric approach.

TABLE 6 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 performance of Massachusetts elementary school students has improved markedly. In 2007, the state's reading and mathematics performance on the NAEP ranks first in the nation. Thus, in our view, the evidence we present should not be seen as an attack on the educational reform effort in Massachusetts. Instead, it is more fruitful to view our results as evidence of unanticipated consequences of efforts to prepare all students to meet the demands of 21st century life. The 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. We 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.

We see several complementary explanations for the finding that failing the 10th grade mathematics examination presents a barrier to graduation for urban students from low-income families, but not for more affluent or suburban students. One is that families of low-income urban students may lack the resources to help them overcome the hurdle posed by failing the examination. A second is that more suburban and affluent urban students may be concentrated in the strongest public schools, with the greatest capacity to help students who fail. A third is that the interaction between school and home contexts may produce these effects. Finally, the different consequences for failing the ELA examination than for failing the mathematics examination suggest that urban schools may devote more resources to remediation in reading and writing rather than to remediation in mathematics or may be less successful in developing effective remediation programs 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 low-income urban students who fail the 10th grade mathematics examination retake it at similar rates as their wealthier urban or suburban peers is encouraging. It suggests that these students are receiving the message that they should persist and retake the test. 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? Did 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?

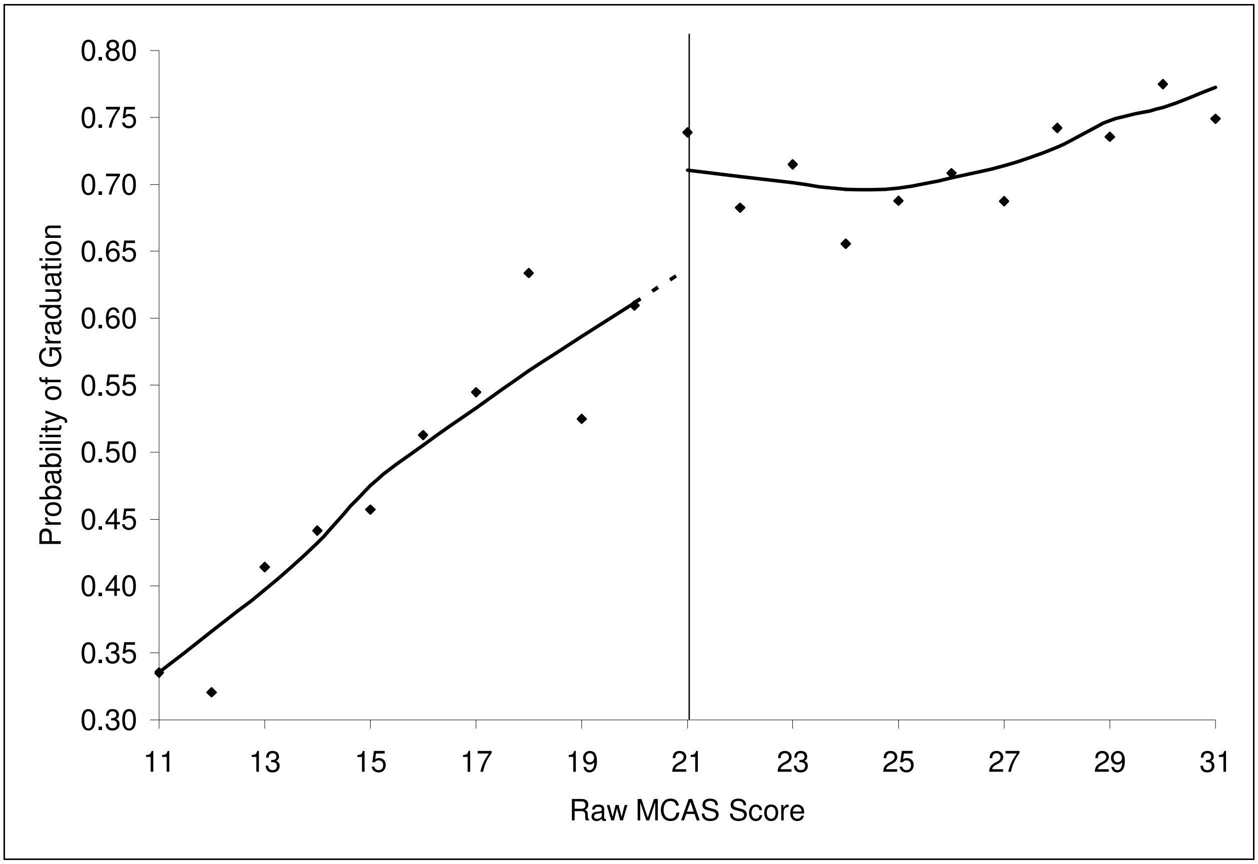
Our findings that the *Fear of Failing* the 10th grade examination induces some low-income urban students to drop out before even taking it raise additional questions. Failing the 8th grade examination gives students some sense of their probable performance on the 10th grade test, but discerning students should recognize that scores on either side of the cutoff are not substantively different. Nonetheless, we found 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?

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 and 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. In future work, we hope to explore more fully the effects of failing on students with limited English proficiency. 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. Fitted smoothed nonparametric 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 nonparametric regression fit without student-level covariates.

Figure 2. 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.

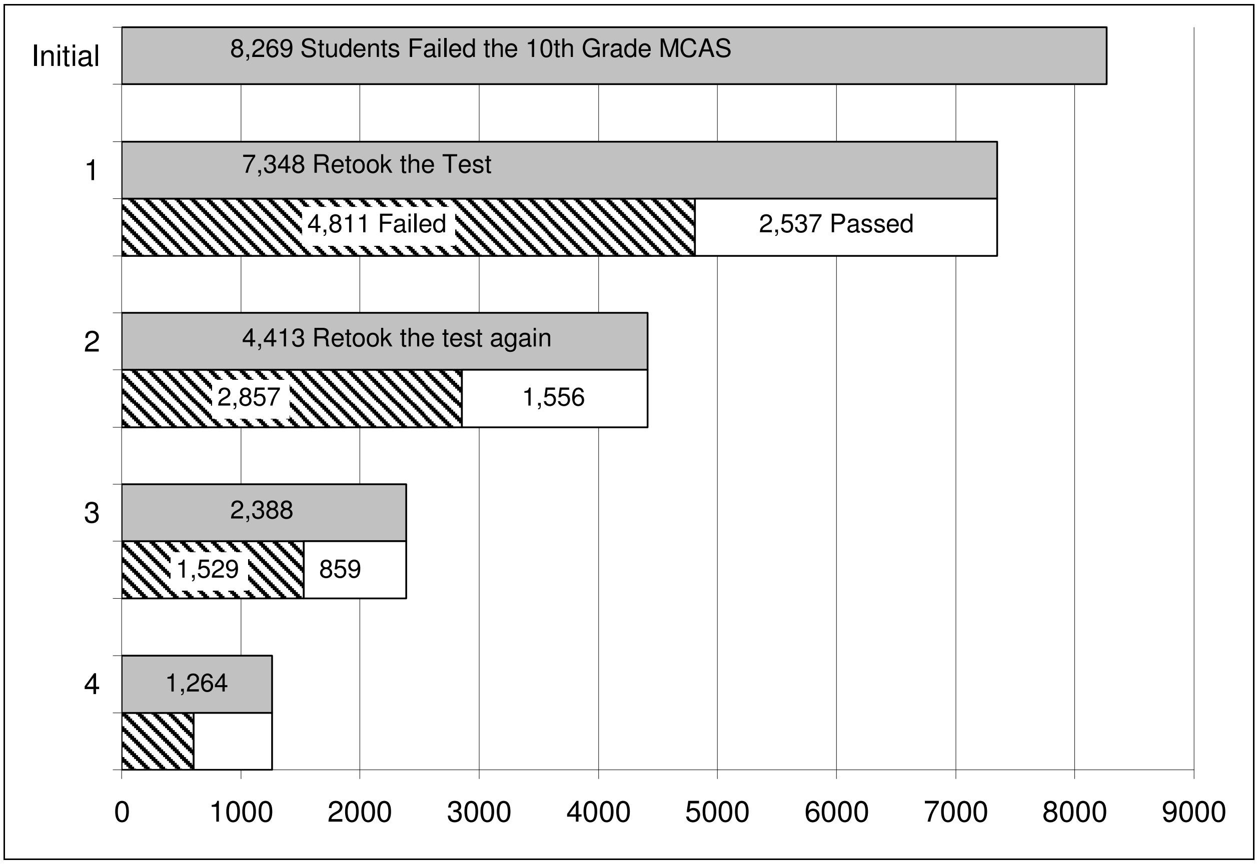
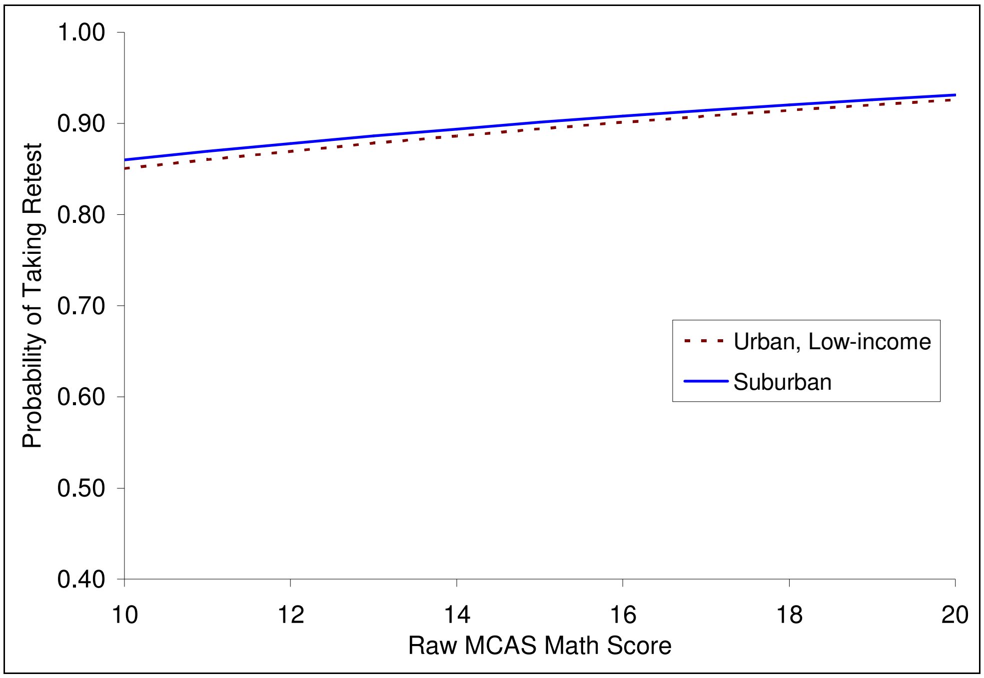


Figure 3. Fitted relationship (from Table 2, Final Models) between the probability of retaking the examination (top panel) or passing the first retest (bottom panel) and first 10th grade mathematics score for low-income urban students and suburban students who failed their first examination (plotted 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).



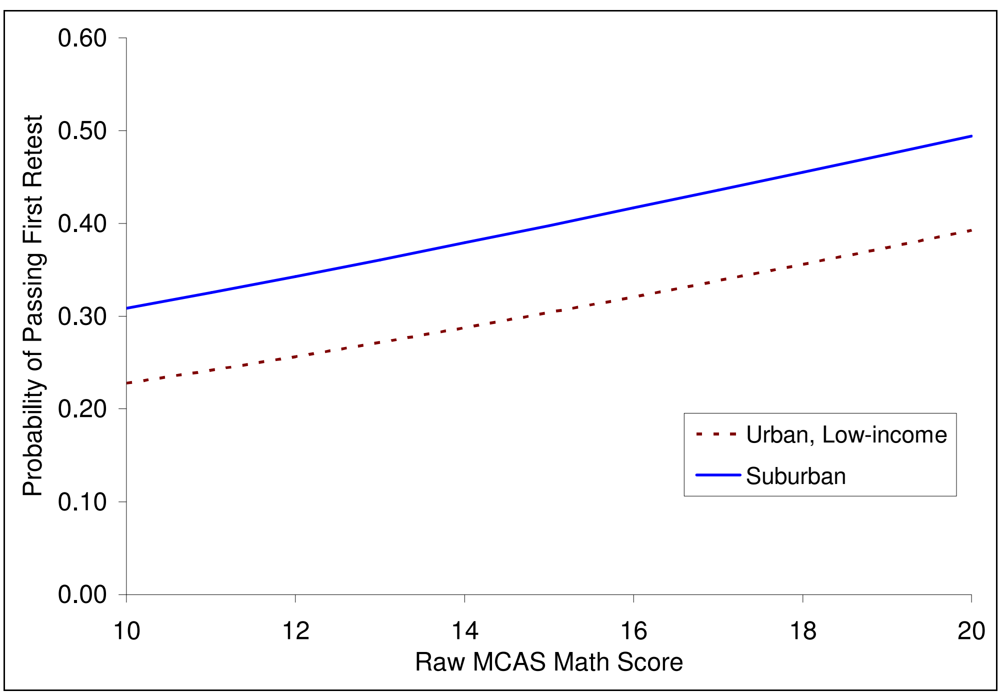
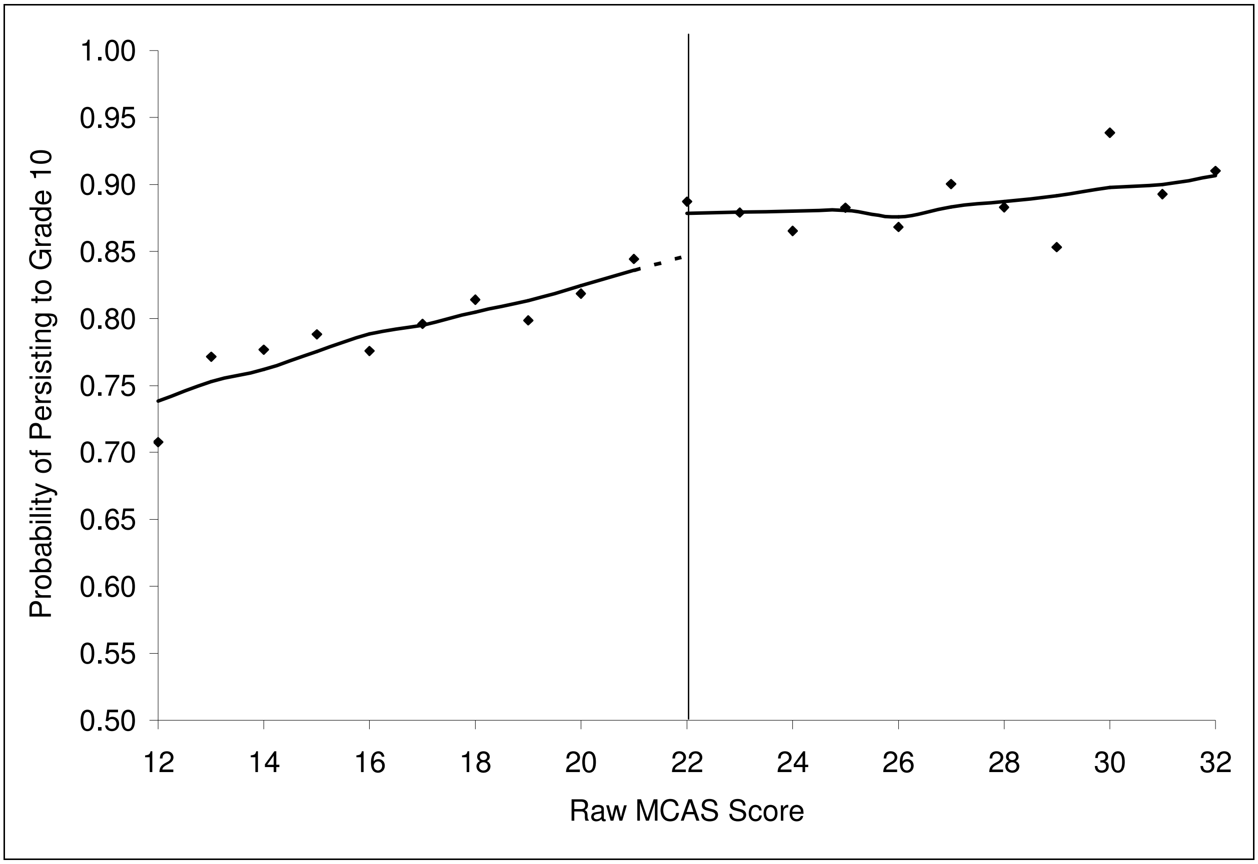
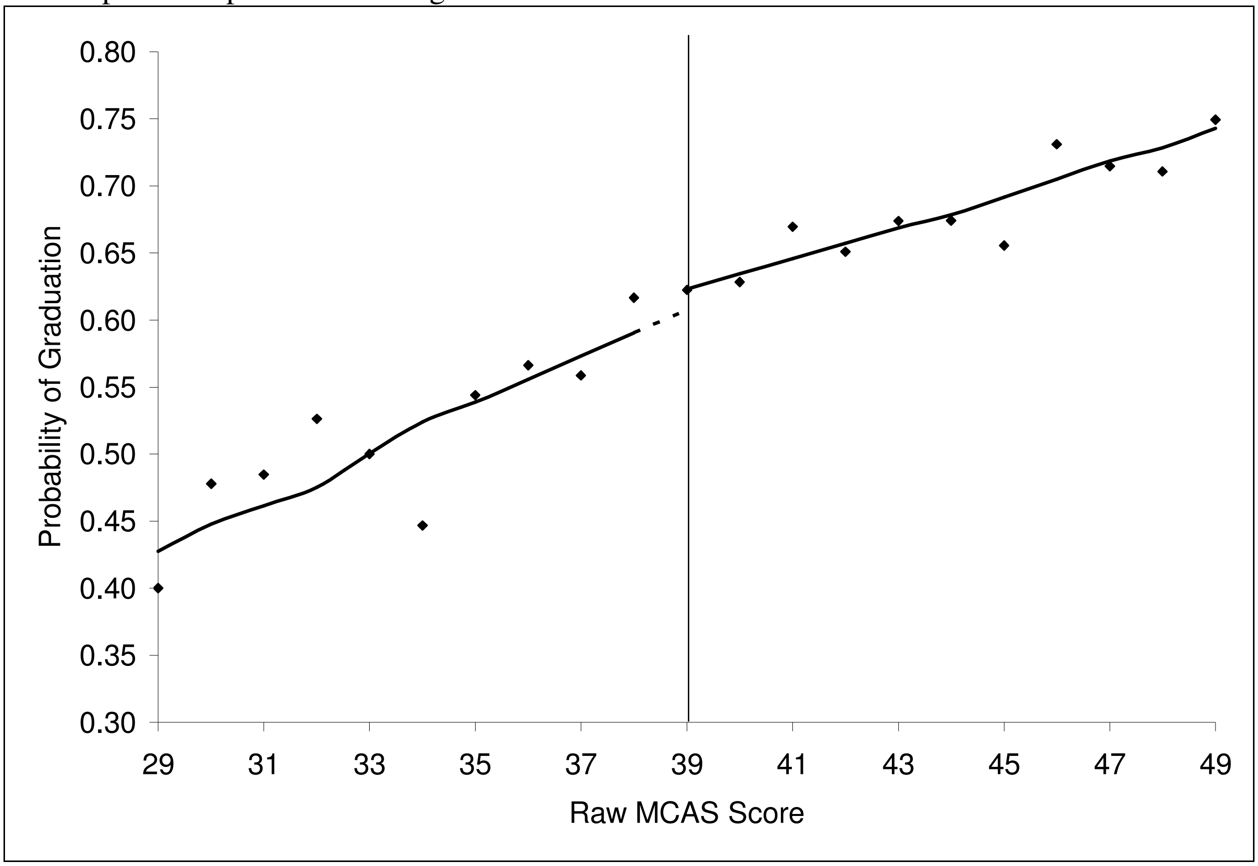


Figure 4. Fitted smoothed nonparametric relationship (bandwidth=6) between the probability of persisting to 10th grade and 8th grade mathematics score for low-income urban students, with the sample mean probabilities of graduation overlaid.



We plot the nonparametric regression fit without student-level covariates.

Figure 5. Fitted smoothed nonparametric 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 nonparametric regression fit without student-level covariates.

Table 1. Parameter estimates and approximate *p*-values at the cut score from the nonparametric regression analysis of the effect of failing the 10th grade mathematics examination on on-time graduation (from the single regression centered at the cut score with bandwidth *h**).

Predictor	Model 1a	Model 1b	Model 2a	Model 2b
Intercept	0.707 ***	0.718 ***	0.787 ***	0.741 ***
MATH	0.023 **	0.022 **	0.025 ***	0.024 ***
PASS	0.028	0.025	-0.011	-0.014
PASSxMATH	-0.016	-0.017 *	-0.016 *	-0.014
MATHxURBAN			0.02	0.023
MATHxPASSxURBAN			-0.009	-0.013
MATHxLOWINC			-0.004	-0.003
MATHxPASSxLOWINC			-0.007	-0.01
MATHxURBANxLOWINC			-0.016	-0.019
MATHxPASSxURBANxLOWINC			0.002	0.006
PASSxURBAN			-0.034	-0.037
PASSxLOWINC			0.031	0.038
PASSxURBANxLOWINC			0.088	0.092
Low-income		-0.057 ***	-0.073	-0.077
Urban		-0.078 ***	-0.065	-0.058
URBANxLOWINC			-0.011	-0.026
African-American		0.064 ***		0.056 ***
Asian-American		0.074 *		0.068 **
Hispanic		-0.005		-0.011
Mixed/Other Race		0.093		0.113 **
Native American		0.036		0.02
Pacific Islander		-0.44 ***		-0.285 *
Limited English Proficient		0.019		0.005
Special Education		0.027 *		0.029 **
Female		0.082 ***		0.081 ***
New Student		-0.076 ***		-0.07 ***
R^2	0.012	0.04	0.04	0.055
Bandwidth (h*)	4	4	6	6
N	8289	8289	11869	11869

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

Table 2. Parameter estimates and approximate *p*-values from the logistic regression analysis of the probability of retaking the examination and passing the first retest, among all students who originally failed (n=8,225).

Predictors	**Probability of Retaking Examination**		**Probability of Passing the First Retest**	
	Full Model	Final Model	Full Model	Final Model
Intercept	2.819 ***	2.817 ***	0.037	0.024
URBANxLOWINC	0.247	0.243 *	0.145	0.112
Urban	-0.322 *	-0.320 **	-0.540 ***	-0.526 ***
Low-income	-0.005		-0.035	
MATH	0.079 ***	0.079 ***	0.078 ***	0.078 ***
ELA	0.045 ***	0.045 ***	0.023 ***	0.023 ***
PASSxELA	-0.295 *	-0.294 *	0.170	0.170 *
African-American	0.673 ***	0.672 ***	-0.158 *	-0.164 *
Asian-American	0.082	0.082	0.107	0.103
Hispanic	0.130	0.129	-0.409 ***	-0.416 ***
Mixed/Other Race	1.570 *	1.569 *	-0.360	-0.367
Native American	0.899	0.898	-0.032	-0.036
Pacific Islander	2.048 *	2.047 *	-0.293	-0.295
Limited English Proficient	0.350 **	0.350 **	-0.246 *	-0.247 *
Special Education	0.517 ***	0.517 ***	-0.074	-0.074
Female	0.160 *	0.160 *	-0.138 **	-0.139 **
New Student	-0.671 ***	-0.671 ***	0.037	0.037
-2*Log Likelihood	4945.80	4945.80	9273.66	9273.90
Pseudo-R^2	0.13	0.13	0.09	0.09

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

Table 3. Parameter estimates and approximate *p*-values at the cut score from the nonparametric regression analysis of the effect of failing the 8th grade mathematics examination on persistence to 10th grade (from the single regression centered at the cut score with bandwidth *h**).

Predictors	Model 3a	Model 3b	Model 4a	Model 4b
Intercept	0.904 ***	0.924 ***	0.945 ***	0.934 ***
MATH (8th grade)	0.009 ***	0.008 ***	0.006 **	0.006 **
PASS (8th grade)	0.007	0.006	-0.009	-0.008
PASSxMATH	-0.005 *	-0.005 *	-0.002	-0.002
MATHxURBAN			-0.004	-0.004
MATHxPASSxURBAN			0.002	0.002
MATHxLOWINC			0.002	0.002
MATHxPASSxLOWINC			0.000	-0.001
MATHxURBANxLOWINC			0.008	0.008
MATHxPASSxURBANxLOWINC			-0.010	-0.009
PASSxURBAN			0.057	0.056
PASSxLOWINC			0.008	0.007
PASSxURBANxLOWINC			-0.025	-0.026
Low-income		-0.020 ***	-0.042 *	-0.040 *
Urban		-0.076 ***	-0.130 ***	-0.131 ***
URBANxLOWINC			0.074 *	0.076 *
African-American		0.018 *		0.016
Asian-American		0.007		0.002
Hispanic		-0.010		-0.012
Mixed/Other Race		0.103 ***		0.103 ***
Native American		-0.032		-0.030
Pacific Islander		0.009		0.010
Limited English Proficient		-0.019		-0.020
Special Education		0.006		0.004
Female		0.020 ***		0.020 ***
R^2	0.009	0.03	0.029	0.032
Bandwidth (h*)	6	6	6	6
N	25456	25456	25456	25456

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

Table 4. Parameter estimates and approximate *p*-values at the cut score from the nonparametric regression analysis of the effect of failing the 10th grade ELA examination on on-time graduation (from the single regression centered at the cut score with bandwidth *h**).

	Model 5a	**Model 5b**	**Model 6a**	**Model 6b**
Intercept	0.676 ***	0.702 ***	0.743 ***	0.709 ***
ELA Score	0.023 ***	0.022 ***	0.021 ***	0.022 ***
PASS	0.004	0.004	0.007	0.004
PASSxELA	-0.010 *	-0.010 *	-0.010 *	-0.010 *
ELAxURBAN			0.000	0.001
ELAxPASSxURBAN			0.016 *	0.014
ELAxLOWINC			-0.007	-0.008
ELAxPASSxLOWINC			0.005	0.006
ELAxURBANxLOWINC			0.010	0.010
ELAxPASSxURBANxLOWINC			-0.025 *	-0.024 *
PASSxURBAN			-0.093 *	-0.074 *
PASSxLOWINC			-0.053 *	-0.043
PASSxURBANxLOWINC			0.132 **	0.118 *
Low-income		-0.057 ***	-0.016	-0.033
Urban		-0.094 ***	-0.082 *	-0.098 **
URBANxLOWINC			-0.012	-0.015
African-American		0.050 ***		0.049 ***
Asian-American		0.110 ***		0.108 ***
Hispanic		-0.003		-0.004
Mixed/Other Race		0.133 ***		0.137 ***
Native American		0.154 **		0.151 **
Pacific Islander		-0.272 *		-0.279 *
Limited English Proficient		0.101 ***		0.097 ***
Special Education		0.030 **		0.027 **
Female		0.056 ***		0.056 ***
New Student		-0.056 ***		-0.051 ***
R^2	0.030	0.057	0.046	0.060
Bandwidth (h*)	8	8	8	8
N	11730	11730	11730	11730

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

Table 5. 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	0.126**	0.100**	**0.078***	0.063*	0.057*
	(0.041)	(0.036)	**(0.033)**	(0.031)	(0.029)
Urban, Not Low Income	-0.017	-0.023	**-0.052**	-0.008	-0.006
	(0.067)	(0.059)	**(0.054)**	(0.050)	(0.046)
Suburban, Low Income	-0.051	-0.026	**0.024**	0.009	0.003
	(0.052)	(0.045)	**(0.042)**	(0.038)	(0.036)
Suburban, Not Low Income	0.000	-0.017	**-0.014**	-0.016	-0.028
	(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	0.035	0.031	**0.029**	0.039*	0.034*
	(0.024)	(0.021)	**(0.019)**	(0.018)	(0.017)
Urban, Not Low Income	0.047	0.032	**0.048**	0.028	0.014
	(0.036)	(0.032)	**(0.028)**	(0.026)	(0.025)
Suburban, Low Income	0.015	0.014	**-0.001**	0.009	0.015
	(0.024)	(0.022)	**(0.020)**	(0.018)	(0.017)
Suburban, Not Low Income	-0.009	-0.004	**-0.008**	-0.011	-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	-0.030	-0.014	**0.005**	0.028	0.033
	(0.031)	(0.029)	**(0.027)**	(0.026)	(0.025)
Urban, Not Low Income	-0.105*	-0.077	**-0.070**	-0.039	-0.025
	(0.048)	(0.045)	**(0.043)**	(0.041)	(0.039)
Suburban, Low Income	-0.095**	-0.077*	**-0.039**	-0.029	-0.035
	(0.035)	(0.033)	**(0.031)**	(0.030)	(0.028)
Suburban, Not Low Income	-0.028	-0.012	**0.004**	0.016	0.018
	(0.023)	(0.021)	**(0.020)**	(0.019)	(0.018)

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

Table 6. Estimated causal impact of failing the 10th grade mathematics examination on on-time high school graduation from a logistic regression model, for samples within windows of different widths around the cut score. Panel I presents the estimated logistic regression coefficients, with standard errors in parentheses; Panel II presents the fitted differences in the probability of graduation for a typical student. Results for the optimal bandwidth, h*, appear in bold

Panel I: Logistic regression coefficients (standard errors in parentheses)					
	Width of window around discontinuity				
Group	+/- 4	+/- 5	+/- 6	+/- 7	+/- 8
Urban, Low Income	0.580**	0.467**	**0.370***	0.300*	0.274*
	(0.182)	(0.161)	**(0.147)**	(0.136)	(0.127)
Urban, Not Low Income	-0.073	-0.108	**-0.224**	-0.029	-0.019
	(0.304)	(0.266)	**(0.246)**	(0.227)	(0.211)
Suburban, Low Income	-0.226	-0.101	**0.133**	0.060	0.029
	(0.251)	(0.218)	**(0.198)**	(0.181)	(0.169)
Suburban, Not Low Income	0.002	-0.089	**-0.056**	-0.061	-0.118
	(0.184)	(0.163)	**(0.149)**	(0.138)	(0.131)

Panel II: Probability of graduation					
	Width of window around discontinuity				
Group	+/- 4	+/- 5	+/- 6	+/- 7	+/- 8
Urban, Low Income	0.123**	0.099**	**0.077***	0.062*	0.057*
Urban, Not Low Income	-0.014	-0.022	**-0.044**	-0.006	-0.004
Suburban, Low Income	-0.039	-0.018	**0.025**	0.011	0.006
Suburban, Not Low Income	0.000	-0.014	**-0.009**	-0.009	-0.018

The "typical" student in the regression discontinuity sample is a white female, not classified as either LEP or as special education.

References

Bureau of Labor Statistics, U.S. Department of Labor. (2007). *Table A-4: Employment status of the civilian population 25 years and over by educational attainment*. Retrieved May 10, 2007 from http://www.bls.gov/news.release/empsit.t04.htm.

Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 3A, pp. 1801-1863). Amsterdam: Elsevier Science.

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

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.

Haveman, R.H., & Wolfe, B.L. (1984). Schooling and economic well-being: The role of nonmarket effects. *The Journal of Human Resources*, *19*(3), 377-407.

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.

Ludwig, J. and Miller D. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *Quarterly Journal of Economics*, *122*(1), 159-208.

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.

Massachusetts Department of Education. (2007). *2007 NAEP tests: Summary of results for Massachusetts*. Retrieved March 26, 2008, from http://www.doe.mass.edu/news/news.asp?id=3692.

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/.

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

Reardon, S.F. & Galindo, C. (April, 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.

Thistlethwaite, D.L. & Campbell, D.T. (1960). Regression-discontinuity analysis: An alternative to the *ex post facto* experiment. *The Journal of Education Psychology, 51*(6), 309-317.

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

U.S. Census Bureau. (2006). *Current population survey: 2006 annual social and economic supplement*. Retrieved April 30, 2007, from http://pubdb3.census.gov/macro/032006/perinc/new03_037.htm.

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.