HIGH-SCHOOL EXIT EXAMINATIONS AND THE SCHOOLING DECISIONS OF TEENAGERS: A MULTI-DIMENSIONAL REGRESSION-DISCONTINUITY ANALYSIS

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

High-School Exit Examinations and the Schooling Decisions of Teenagers: A Multi-Dimensional
Regression-Discontinuity Analysis
John P. Papay, John B. Willett, and Richard J. Murnane
NBER Working Paper No. 17112
June 2011
JEL No. C10,C14,I20,I21,I28,J24

ABSTRACT

We ask whether failing one or more of the state-mandated high-school exit examinations affects whether
students graduate from high school. Using a new multi-dimensional regression-discontinuity approach,
we examine simultaneously scores on mathematics and English language arts tests. Barely passing
both examinations, as opposed to failing them, increases the probability that students graduate by 7.6
percentage points. The effects are greater for students scoring near each cutoff than for students further
away from them. We explain how the multi-dimensional regression-discontinuity approach provides
insights over conventional methods for making causal inferences when multiple variables assign individuals
to a range of treatments.

John P. Papay
Brown University
Education Department
Providence, RI 02912
john_papay@mail.harvard.edu

John B. Willett
Graduate School of Education
Harvard University
6 Appian Way - Gutman 412
Cambridge, MA 02138
John_Willett@harvard.edu

Richard J. Murnane
Graduate School of Education
Harvard University
6 Appian Way - Gutman 469
Cambridge, MA 02138
and NBER
richard_murnane@harvard.edu

High-School Exit Examinations and the Schooling Decisions of Teenagers:

A Multi-Dimensional Regression-Discontinuity Analysis

After rapid growth through the first 70 years of the twentieth century, the high-school graduation rate in the United States has stagnated over the past four decades (Katz & Goldin, 2008). Although estimating the national graduation rate is difficult for many reasons, somewhere between 20 and 30 percent of U.S. high-school students drop out before graduating (Heckman & LaFontaine, 2010; Stillwell, 2010). This pattern is surprising given the growing importance of a high-school diploma for access to both post-secondary education and higher-paying jobs in the labor market. Indeed, Heckman, Lochner, and Todd (2008) show that the internal rate of return to a high-school diploma, relative to dropping out of school, has increased rapidly in recent years and now stands at greater than 50 percent.

Why do so many American teenagers drop out of high school prior to graduation? Potential explanations include poor information about labor markets, credit constraints that prevent borrowing against future income, and perceived high non-pecuniary costs of earning a high school diploma. In recent years, these non-pecuniary costs have risen for many marginal students because they must now pass exit examinations in order to graduate. Today, more than three-quarters of American public-school students face high-school exit examination requirements (Center on Education Policy, 2010). Students typically take these tests, usually in mathematics and English language arts (ELA), early in their high-school career. Although students have multiple opportunities to retake these examinations, failing an exit examination early in high school may increase the cost of graduation because students must prepare for and spend time retaking the tests. They may also pose additional psychic barriers to continuing in

1

school, as students who fail an examination may lose confidence in their academic abilities.

In this paper, we ask whether failing one or more of the state-mandated high-school exit examinations reduces the probability that students graduate from high school. Importantly, we go beyond prior research on this topic by using a multi-dimensional regression-discontinuity approach. This paper represents the first application of this generalized strategy for addressing situations in which multiple variables assign individuals to a range of different treatments. Because students must take and pass examinations in mathematics and ELA, there may be important interactions and complementarities between students' performances on these two tests. Our approach allows us to examine whether the impact of barely passing one exam on the probability of graduation depends on the student's performance on the other test.

We find that barely passing both examinations, as opposed to failing them, increases the probability that students on the margin graduate from high school by 7.6 percentage points. Passing the ELA examination increases the probability of graduating by 5.4 percentage points for students who failed the mathematics test, but only by 4.5 percentage points for students who passed in mathematics. Finally, the effects of passing seem to be much greater for students near both of the cutoffs than for students further away from them.

We begin with a brief discussion of how high-school exit examination performance can affect student outcomes. We then explain our approach, describing our dataset and the Massachusetts context as well as our extension of the regression-discontinuity design. We present our results, describe the threats to the validity of our main inferences, and conclude with a discussion of both the opportunities for further research and the limitations of this approach.

Section II: The Effects of Failing a High-School Exit Examination

In a simple model, students will continue in school through graduation if their assessment

of the marginal benefits of a high-school diploma exceeds their marginal costs. Failing an exit examination could increase the "costs" of obtaining a high school diploma in several ways. First, it presents a structural barrier for students who cannot pass the examination on retest.[1] Second, preparing for and retaking the test may simply be a burden that some students are not interested in bearing.[2] Finally, failing the examination may affect students' perceptions of themselves and their academic abilities, which may in turn affect their ideas about the returns to additional schooling. It might also produce an emotional response as students who are told they are "failing" in mathematics and/or ELA may become discouraged.

Over the past decade, researchers have begun to examine the consequences of failing high-school exit examinations, using regression-discontinuity designs to estimate the causal effect of barely failing an exit examination on subsequent student outcomes, including future academic achievement, high-school graduation, and labor-market earnings.[3] The results are mixed. Using data from Texas, Martorell (2005) found no effects of barely failing the first exit examination on either student graduation rates or labor market earnings. Reardon et al. (2010) found no statistically significant effects on graduation rates (or most other outcomes) in either

---

[1] Many states, including Massachusetts, explicitly attempt to limit this barrier by allowing students multiple retest opportunities. For example, some Massachusetts students retake the test more than six times. Nonetheless, some students still cannot meet the passing standard.

[2] They may judge that the extra time spent is not worthwhile; this may be particularly true for students who are very far away from the passing cutoff.

[3] Note that the question of how failing an exit examination affects students is substantively different from the question of how imposing an exit examination policy affects students. We focus on the first.

mathematics or ELA in California. By contrast, in earlier work in Massachusetts, we (2010) found that barely failing the first attempt on the mathematics examination reduced the probability of on-time graduation by approximately 8 percentage points for urban, low-income students, but there were no effects of barely failing the ELA examination. Ou (2010) found quite similar results in New Jersey. These differences could reflect variation in the tests or student populations across these different states.

In all of these states, students must pass multiple tests in order to graduate from high school. Thus, all researchers in this area have faced situations in which multiple tests define a range of different treatment conditions. To date, however, they have addressed this challenge in one of two ways. Martorell (2005) incorporated all of the tests into a single analysis, but he reduced the dimensionality of his forcing variables by compositing the tests into a single forcing variable that represented the minimum of a student's test scores across subjects. Because students in Texas must pass three tests in order to graduate, this approach enabled him to draw causal inferences about students on the margin of passing the test in which they performed the worst. However, this approach ignores potentially important differences in the effects of failing examinations in different subjects. Failing a mathematics examination may be substantially different for students than failing an ELA examination. For example, remediation may be much easier in one subject than the other, or differences on the tests may make failing one subject more important than the other.

The second approach involves analyzing the different subjects separately. We pursued this strategy in our earlier paper (2010), fitting separate regression models for the mathematics and ELA tests. Ou (2010) and Reardon et al. (2010) took similar approaches. However, this separate analysis assumes implicitly that treatment effects are homogeneous across levels of the

other examination; in other words, it assumes that failing the mathematics examination matters just as much for students who score "Advanced" in ELA as those who fail the ELA test. Clearly, in a regime where students must pass both tests, failing two tests may be substantially worse than failing one.

Reardon et al. (2010) take the analysis one step further, examining whether the effect of failing one test in California depended on whether students passed the other. They found no striking patterns. However, it is quite plausible that this approach still obscures substantial heterogeneity in effects across levels of the other examination. For example, failing the ELA test may matter much less for students who score very highly on the mathematics examination than for students who score near the mathematics cutoff. In other words, students who feel confident about their mathematics ability might not be affected by the additional burden of passing the ELA examination. Similarly, students who earn a very low score on the mathematics test may believe they will never be able to improve their performance enough to pass it; if so, barely failing the ELA examination would not matter.

We implement a new approach that allows us to obtain additional information about the consequences of examination performance. We estimate the effect of barely passing both of these examinations and explore how the effect of passing one test differs across levels of the other examination. In other words, we ask whether barely passing either (or both) the mathematics or ELA examination on the first try affects the probability of subsequent high-school graduation in Massachusetts for students on the margin of passing. If so, we ask whether the effect of passing the ELA examination differs by mathematics test performance, and whether the effect of passing the mathematics examination differs by ELA test performance. This set of questions requires us to model simultaneously the discontinuities in both forcing variables – the

mathematics and ELA test scores. We outline our strategy briefly in the next section and provide a more extended explanation in the appendix.

Section III: Estimating the Effect of Discontinuities in Mathematics and ELA

Data and Context

Our dataset comes from Massachusetts. During the time period that we studied, students had to pass the *Massachusetts Comprehensive Assessment System* (MCAS) tests in both mathematics and ELA in order to graduate from high school. Students take these examinations for the first time in the spring of 10th grade, and they receive their scores several months later. The state determines the passing cutoff each year, and all students with scores that fall at or above the cutoff pass it. Students have multiple opportunities to retake examinations that they fail.

We use data from the 2006, 2007, and 2008 graduating cohorts. The Massachusetts *Department of Elementary and Secondary Education* has compiled a comprehensive database that follows students longitudinally through high school and tracks students who leave the system. We focus on the 202,860 students who first took the 10th grade MCAS examinations as sophomores (in 2004, 2005, or 2006, respectively) and for whom it was a high-stakes test.[4] As we discuss later, we focus on students who score "near" the cut-scores on both tests, which we call the "joint cutoff". As a result, our analytical sample sizes are, in practice, much smaller than these 200,000 first-time test-takers.

---

[4] For less than 1% of all students, including those with serious special educational needs, the exit examination requirement can be satisfied through alternative measures and the 10th grade examination is not required. We exclude these students from our analysis.

Massachusetts is an interesting state to study because its educational system is one of the highest performing in the country and it has placed a high priority on educational reform. Since the *Massachusetts Education Reform Act* of 1993, which introduced standards-based reforms and state-based testing, Massachusetts has invested substantially in additional K-12 public education funding. This investment included a great deal of financial resources, as well as clearly defined academic standards and curriculum frameworks. Under these reforms, the state also began administering the *Massachusetts Comprehensive Assessment System* (MCAS) tests in 1998. For the class of 2003, the 10th grade tests became high-stakes exit examinations.

This focus on standards-based reform and these investments in public education appear to have paid off. The state has been praised for having the most rigorous academic standards in the country and state assessments that align closely with these standards (Finn, Julian, & Petrilli, 2006; *Quality Counts*, 2006). Furthermore, Massachusetts students are consistently among the nation's top performers on the National Assessment of Educational Progress (NAEP) examinations, and the state's NAEP performance has improved rapidly since the introduction of state testing (NCES, 2008). Beyond the K-12 education system, Massachusetts is also a state whose economic conditions are somewhat different than the rest of the country as the state has had relative economic success. With per capita income of $51,254 in 2007, the state ranked third in the country; 38% of adult residents have at least a bachelor's degree, the most in the nation (U.S. Census, 2009).

Measures

Our primary outcome variable, named *GRAD*, indicates whether the student graduated

on-time from a public high school in Massachusetts.[5] Districts report the values of individual student-graduation outcomes using the state's *Student Information Management System* (SIMS). Our dataset also contains a record of student test scores; we focus our attention on scores from the student's first attempt on each of the exit examinations. To implement our regression-discontinuity approach, we center students' raw scores by subtracting out the value of the corresponding minimum passing score, to create our forcing variables.[6] On each of these re-centered continuous predictors ($MATH^C$ and ELA^C), a student with a score of zero had achieved the minimum passing score. We also create dichotomous versions of the same predictors (*PASS_MATH* and *PASS_ELA*) to indicate whether a student's score lay above the pass/fail cutoff on the forcing variable.

The dataset also includes information on several exogenous covariates, such as student race and gender as well as indicators for whether the student was classified as limited English proficient, special education, or low-income, or attended a high school in one of the state's urban school districts. We include these covariates as well as the fixed effect of cohort in our analysis to improve the precision of our estimation. It is reassuring that our results are quite similar, albeit somewhat less precise, if we exclude them.

---

[5] We treat teenagers who leave school prior to graduation and obtain the General Educational Development credential as non-graduates.

[6] The cut scores vary by subject and year. For example, students had to earn 21 points to pass the mathematics examination in 2004, but only 19 points in 2005 and 20 points in 2006. The cut scores in ELA were 39, 38, and 35 points, respectively, across the three years. For more information on MCAS scoring and scaling, see the MCAS Technical Reports (MA DOE, 2002, 2005).

Analytic Approach

As any regression-discontinuity design, we seek to estimate the conditional mean of the outcome *at the cut score* for individuals who fall into different treatment groups (e.g., pass or fail the examination). In the case of a single exit examination, then, our parameters of interest are:

$$\mu_r(GRAD) = \lim_{x \to 0+} E[GRAD \mid MATH_i^C = x]$$

and   (1)

$$\mu_l(GRAD) = \lim_{x \to 0^-} E[GRAD \mid MATH_i^C = x]$$

Assuming that the cutoff has been assigned exogenously and we can estimate these limits credibly, the difference between these means is an unbiased estimate of causal effect of treatment (τ), for students at the cutoff:

$$\tau = \mu_r(GRAD) - \mu_l(GRAD) \qquad\qquad\qquad (2)$$

In other words, it estimates the difference in the population probability of graduating for students who pass and fail the test, on the margin of passing.

Because students must pass two exit examinations in order graduate, the two tests actually define four different "treatment" conditions:

(1) Pass Mathematics and Pass ELA;
(2) Fail Mathematics and Pass ELA;
(3) Pass Mathematics and Fail ELA; or
(4) Fail Mathematics and Fail ELA.

We are interested in all of the pairwise comparisons among these treatments. For example, comparing (1) and (2) gives us the effect of barely passing the mathematics examination for students who pass the ELA test.

In Papay, Willett, and Murnane (2011), we lay out a general multi-dimensional regression-discontinuity approach that allows us to make all of these pairwise comparisons by

incorporating discontinuities in multiple forcing variables into a single analysis. Here, we describe briefly how we implement this approach. We provide more details in an appendix.

Because the state imposes its passing criteria rigidly, the discontinuities are sharp in both the mathematics and ELA examinations. The four treatment conditions define four separate regions in the two-dimensional space spanned by the forcing variables, ($MATH^C$, ELA^C). Similar to the case with a single forcing variable, our parameters of interest are the conditional mean probabilities of graduation for individuals in each treatment condition, at the cutoff. For example, the causal effect of passing the mathematics examination instead of failing it for individuals scoring at the cutoff on the ELA test would be the difference between:

$$\mu_r(GRAD) = \lim_{x \to 0^+} E[GRAD_i \mid MATH_i^C = x, ELA_i^C = 0]$$

and   (3)

$$\mu_l(GRAD) = \lim_{x \to 0^-} E[GRAD_i \mid MATH_i^C = x, ELA_i^C = 0]$$

To estimate these limits, and the difference between them, we generalize Imbens & Lemieux's (2008) "non-parametric" approach. We first choose an "optimal" joint bandwidth around the cut-off on the forcing variables (labeled h_1^*, h_2^*) and then estimate the causal effect by conducting a local linear regression analysis using this optimal bandwidth.

Bandwidth Selection

The primary challenge in implementing our approach comes in choosing the appropriate bandwidths (h_1^*, h_2^*) for our analysis. For each observation, at each point on the ($MATH^C$, ELA^C) grid, we fit a linear regression model – within an arbitrary bandwidth (h_1, h_2) – to estimate a fitted value of the outcome at that point:

$$\hat{\mu}(MATH_i^C, ELA_i^C, h_1, h_2) = \hat{\gamma}_0 + \hat{\gamma}_1 MATH_i^C + \hat{\gamma}_2 ELA_i^C + \hat{\gamma}_3(MATH_i^C \times ELA_i^C) \quad (4)$$

In each case, we attempt to mirror the regression-discontinuity approach by only using observations that fall within the appropriate region, and estimating $\hat{\mu}(MATH_i^C, ELA_i^C, h_1, h_2)$ as if it were a boundary point.[7] For a given bandwidth (h_1, h_2), we thus estimate a fitted probability of graduation for each observation. We compare these fitted values to the observed values, across the entire sample, using a generalized version of the Imbens & Lemieux (2008) cross-validation criterion:

$$CV_{GRAD}(h_1, h_2) = \frac{1}{N} \sum_{i=1}^{N} (GRAD_i - \hat{\mu}(MATH_i^C, ELA_i^C, h_1, h_2))^2 \qquad (5)$$

Our optimal joint bandwidth, h_1^* and h_2^*, is the pair of bandwidths that minimizes the CV criterion. In Figure 1, we present a surface plot showing the estimated value of the cross-validation criterion at each of these bandwidths in $h_{MATH} \in [5,12]$ and $h_{ELA} \in [5,15]$. It reaches its minimum at (7, 10).

INSERT FIGURE 1 ABOUT HERE

Estimation

After selecting an optimal bandwidth, we estimate the causal effect of passing the examinations using local linear-regression analysis. We fit the requisite regression models in each region simultaneously, by specifying a single statistical model with 16 parameters – an intercept and slope parameters to accompany all 15 possible interactions among $MATH^C$, ELA^C, *PASS_MATH*, and *PASS_ELA*. We fit locally linear probability models of the following form,

---

[7] For example, for a student who fails both tests and for whom $MATH^C$=-3 and ELA^C=-5, we use only observations for which $MATH^C$<-3 and ELA^C<-5. By contrast, for students who pass both tests and for whom $MATH^C$=10 and ELA^C=15, we use only observations for which $MATH^C \geq 10$ and $ELA^C \geq 15$.

using observations whose mathematics and ELA scores fall within one optimal bandwidth on either side of the relevant cut-scores:

$$p[GRAD_i = 1] = \beta_0 + \beta_1 PASS\_MATH_i + \beta_2 PASS\_ELA_i + \beta_3 (PASS\_MATH_i \times PASS\_ELA_i)$$
$$+ \beta_4 MATH_i^c + \beta_5 ELA_i^c + \beta_6 (MATH_i^c \times ELA_i^c) + \beta_7 (MATH_i^c \times PASS\_MATH_i)$$
$$+ \beta_8 (ELA_i^c \times PASS\_ELA_i) + \beta_9 (MATH_i^c \times PASS\_ELA_i) + \beta_{10} (ELA_i^c \times PASS\_MATH_i) \qquad (6)$$
$$+ \beta_{11} (MATH_i^c \times ELA_i^c \times PASS\_MATH_i) + \beta_{12} (MATH_i^c \times ELA_i^c \times PASS\_ELA_i)$$
$$+ \beta_{13} (MATH_i^c \times PASS\_MATH_i \times PASS\_ELA_i) + \beta_{14} (ELA_i^c \times PASS\_MATH_i \times PASS\_ELA_i)$$
$$+ \beta_{15} (MATH_i^c \times ELA_i^c \times PASS\_MATH_i \times PASS\_ELA_i) + \gamma' Z_i + \varepsilon_i$$

Here, we include the set of student-level covariates described above (Z_i) to improve the precision of our estimation. We can interpret this single model *parametrically* for observations local to the cut score and use the standard errors to conduct appropriate statistical tests. Following Lee & Card's (2008) admonition to account for the discrete nature of assignment variables, we treat each combination of mathematics and ELA scores as a cluster in computing the standard errors. In all cases, we expect students who pass the test to have better outcomes than students who fail; as a result, we make use of one-tailed tests throughout our analysis, reporting the corresponding *p*-values.

Section IV: Results

We present the results from our analysis graphically in Figures 2 and 3. The surface plot in Figure 2 represents regression surfaces derived from our full fitted model. Each quadrant represents one of the treatment conditions: for example, students in the bottom quadrant are those who failed both examinations. To illustrate that our interpretations are only valid for students near the cut scores, we present only the part of each fitted surface that is local to one of the cut scores. The edges of these surfaces at the cut scores represent the fitted probabilities of graduation for students who just pass or just fail one of the tests; the causal effects of failing are simply the vertical distances between these edges.

To aid in interpreting these fitted effects, we present a vertical cross-section of this figure in the top panel of Figure 3. We take this cross-section at the ELA cut score; therefore, the lines represent the fitted probability of graduation for students scoring at the ELA cut score by different levels of their mathematics scores. The solid line represents students who pass the ELA test, while the dotted line represents our counterfactual – the fitted probability of on-time graduation for students scoring at the cut score who *fail* the ELA test. Thus, the estimated causal effect of passing the ELA test, as opposed to failing it, is the difference between these two lines, illustrated in the bottom panel. The results for the mathematics examination are qualitatively similar, although the implied causal effects of failing the examination are smaller.

<center>INSERT FIGURES 2 AND 3 ABOUT HERE</center>

These figures illustrate several key findings, which we summarize in Table 1. First, we estimate that, for students near the joint cut scores, barely passing both examinations increases the probability of on-time graduation by 7.6 percentage points ($p=0.011$). This effect represents an estimate of the parameter sum $\beta_1 + \beta_2 + \beta_3$ from Equation (6), and we see it in Figure 2 as the vertical distance between Region D and Region A at the cut score. This is a substantial effect, given that only 60 percent of students scoring right at the joint cutoff graduate from high school on-time.

<center>INSERT TABLE 1 ABOUT HERE</center>

Second, barely passing the ELA examination increases the probability of on-time graduation for students near the cut scores, and this effect appears to be somewhat greater for students who just fail the mathematics examination (5.4 percentage points, $p=0.063$, estimated parameter β_2 from Equation (6)) than for students who just pass it (4.5 percentage points, $p=0.089$, estimated parameter sum $\beta_2+\beta_3$ from Equation (6)). In mathematics, we find smaller

<center>13</center>

effects of failing, but the pattern of the point estimates is similar to that in ELA. Thus, having to pass two examinations appears to be a greater hurdle than having to pass one. Viewing students who fail both the mathematics and ELA examinations as having the same "treatment" as students who fail just one of these tests ignores important differences in the consequences of failing these tests. Furthermore, for students near the joint cut score, barely passing the ELA examination appears to be somewhat more important than passing the mathematics examination. This could reflect the relative challenge of the two tests; the ELA examination is easier for Massachusetts students. Approximately 12% of all test-takers fail the mathematics examination, while only 7% fail the ELA test.

Importantly, our approach enables us to draw inferences about the causal effect of passing the ELA examination across all levels of students' mathematics scores. We can estimate these effects simply by fitting our full model from equation (6) using students with different mathematics scores near the ELA cut score. We can envision "sliding" the two-dimensional bandwidth smoothly along the ELA cut score, estimating the effect of passing the ELA examination for students at different levels of mathematics proficiency.

We find that the effect of passing the ELA examination appears to be greater for students scoring close to the mathematics cut score than for students with higher or lower mathematics scores. In particular, for students with relatively high scores on the mathematics test, whether they pass or fail the ELA examination does not affect whether they graduate from high school. In Figure 4, we present the results from this analysis, plotting the estimated causal effect of passing the ELA examination by mathematics score.[8] We see that the effect of barely passing is greatest

---

[8] In the tails of the distribution, these estimates become incredibly imprecise. As a result, we only present estimates where more than 3,000 students contributed to the estimate.

for students near the joint cut-off. Among students who fail the mathematics examination (on the left of the cut score), barely passing the ELA examination has an important effect on their graduation rates across a wide range of scores. The estimated effect is 3.0 percentage points, even for very low-performing students scoring 11 points below the cutoff (the 1st percentile statewide). For students who are relatively higher-performing in mathematics, though, barely passing the exit examination matters much less. While these estimates are not sufficiently precise to draw clear conclusions, they do suggest important patterns of heterogeneity in the student population.

INSERT FIGURE 4 ABOUT HERE

Section V: Threats to Validity

Our ability to draw causal inferences from our regression-discontinuity design requires two important conditions to be met. First, treatment assignment – here whether students are classified as failing one or more of the examinations – must be exogenous and applied rigidly to all students. All student characteristics, both observed and unobserved, must differ smoothly as a function of the forcing variables around the cut scores. In other words, students must not be able to manipulate their positions relative to the cut scores on the forcing variable. In Massachusetts, this assumption holds because all students who score below the cut-offs fail the tests and all students who score above the cut-offs pass them. Furthermore, the minimum passing scores differ from year-to-year based on a complicated scaling formula and are determined *after* students take the tests; thus, students cannot score intentionally just above the cut-offs.

In addition to this *prima facie* evidence, however, we can assess whether the cut-scores were imposed exogenously in several ways. First, we examine the density of students falling on either side of the cut scores and find no discontinuity. For example, 74 students score right at the

cut-scores on both tests, while 71 students score one point below the cut-offs. Second, we examine whether there are apparent discontinuities at the joint cut-scores in the average values of each of ten student-level covariates. Treating each covariate as an outcome, we use the same regression-discontinuity approach to estimate five relevant "effects."[9] Among these fifty estimates, we find only two statistically significant results at $p<.05$ and five at $p<.10$, exactly what we would expect from chance, with a five- or ten-percent Type I error. Thus, we find no reason to doubt that the state has imposed the cut-scores exogenously and consistently.

The second key assumption underpinning our regression-discontinuity analyses is that we are able to estimate the relationship between the outcome and forcing variable adequately, at least in the immediate vicinity of the cut-score. Here, we assess the sensitivity of our results to bandwidth selection. In Table 2, we present our key results as we manipulate the bandwidth across a range of credible values, along both forcing variables. Importantly, we see that the general relationships described above are preserved across these relevant bandwidths; in all cases, the effects of passing the ELA examination are positive. In the full model, just passing the ELA examination is associated with a greater jump in the fitted probability of on-time graduation for students failing the mathematics examination than for students who pass it. This result holds

---

[9] The effect of passing both tests (1), passing ELA for students who pass mathematics (2) and who fail mathematics (3), and passing mathematics for students who fail ELA (4) and pass ELA (5).

across a wide range of bandwidths. In all cases, the effect of passing both tests, compared to failing both, is large and substantively important (ranging from 4.9 to 8.5 percentage points).[10]

INSERT TABLE 2 ABOUT HERE

Finally, we must be concerned about type I error in our analysis, particularly given that several of our results do not reach traditional levels of statistical significance. Although our full model allows us to estimate a wide range of interesting effects and to present a more nuanced picture of the causal effect of failing an exit examination, it also requires great density of data in the vicinity of the multiple cut scores in order to estimate effects with sufficient power. Estimates of the slopes for several of the surfaces are particularly imprecise because relatively few students fall in some of the regions. For example, for most students in Massachusetts, passing the ELA examination is substantially easier than passing the mathematics test (only 2.0 percent of test-takers pass the mathematics examination but fail the ELA test; by contrast, 6.6 percent pass the ELA test but fail the mathematics test). As seen in Figure 5, because the cut scores fall in the tails of the distribution, the density of data local to the joint cutoff is restricted substantially. As a result, this analysis is underpowered and we would need to find quite large effects, as we do for failing both examinations, for all of our estimates to reach traditional levels of statistical significance. Nonetheless, these results suggest important heterogeneity in the effect of failing an exit examination that are not revealed by traditional methods of analyzing data from situations in which multiple variables assign individuals to treatments. As a result, despite the power trade-off from fitting this complicated model, we believe these results are instructive.

---

[10] We focus on the magnitudes of parameter estimates rather than the results of hypothesis tests. As bandwidths grow smaller, estimates will necessarily become less precise, but our substantive story remains the same.

INSERT FIGURE 5 ABOUT HERE

Section VI: Discussion

The factors that influence whether students decide to remain in school beyond the minimum school-leaving age are not well understood. In this paper, we examine one possible determinant: the increased costs of remaining in school that arise when students fail a high-school exit examination. Students who fail must spend time and effort studying for and retaking the examination, and failing may impose psychological costs on students because they may feel less confident about their academic abilities. We find that these costs are important, particularly for relatively low-performing students who score near the cutoffs on both the mathematics and ELA examinations. For these students, the decision to stay in school appears to be quite fragile: barely passing both exit examinations increases the probability of graduating from school by 7.6 percentage points. In other words, a one-point difference on these two tests matters a great deal. Given the substantial returns to a high-school diploma, this effect is substantively quite large.

There are two important clarifications in how we interpret these findings. First, we can say nothing about the overall effect of introducing an exit-examination policy. Simply having the requirement could improve student outcomes by making all students work harder throughout their school careers. On the other hand, students who believe they will never pass may drop out before even taking the first test, thereby reducing educational attainments. We focus on the consequences of passing or failing the exit examination in a state that has already implemented the policy. Second, we do not know whether these results represent the positive effect of barely passing the examination, the negative effect of barely failing it, or (more likely) some combination of the two. Regardless, students on either side of the cut scores face different treatments – those who fail must retake and pass one or both examinations, while those who pass

18

do not – and these treatments matter to students.

Students who fail face several challenges, which may affect their educational investment decisions. They may be placed in classes centered on test preparation, which may strike some students as remedial and as less relevant than the traditional curriculum. They may be required to take multiple mathematics or ELA courses, rather than having the flexibility to pursue electives. In order to graduate, they need to go through the process of retaking (and passing) the exit examination, which not only takes time but also may be seen as an additional burden. Finally, being labeled as a failure may reduce their perception of their academic abilities, which may affect their decisions directly.[11]

Unfortunately, we cannot draw clear conclusions about the relative contribution of these processes. However, the fact that retaking an examination requires a relatively small time investment suggests that there may be other mechanisms at play. In particular, being labeled as a failure on a state-administered examination may cause students to lose confidence in their academic abilities. Being identified as a failure in two subjects likely compounds this effect, and we find such a compounded effect of failing both examinations. In other words, students' decisions about whether to continue in school may not be only about balancing the long-term returns of graduating against the short-term costs of continuing in school and retaking the test. Instead, they may reflect the apprehension, disappointment, and lack of confidence among students who fail. Regardless of the mechanisms at play, students scoring on either side of the cut-off have dramatically different outcomes. That such minor differences in test performance produce such substantial consequences should be an area of concern for policymakers.

---

[11] Students' perceptions of their academic abilities affect not only the costs of schooling, but also students' ideas about the returns to high school graduation.

Importantly, we find that passing each examination matters more for students near the joint cut scores than for students far above or below these cutoffs on the other test. For example, barely passing the ELA test increases the probability of graduation by more than four percentage points for students scoring just above the mathematics passing score, but it has no effect on students with higher mathematics scores. This result makes sense, because being high-performing in one subject likely buffers students from the effects of passing or failing the other. In other words, for students with high mathematics scores, barely passing or failing the ELA examination may not matter as much because they have a sense of their stronger abilities in mathematics. What is more surprising is that barely passing an examination appears to have a substantial effect even on students who earn quite low scores on the other test. These students face a difficult challenge in raising their performance in the other subject substantially in order to pass on retest. For such students, passing one examination may provide some encouragement to sustain them through the retesting process (or failing both examinations may provide some additional discouragement preventing them from continuing).

These results not only hold substantive lessons about the importance of exit examination performance in student dropout decisions, but they also have important implications for research. Natural experiments in which units of analysis are assigned to several different treatment conditions based on values of multiple forcing variables are quite common in education as well as in other sectors.[12]   In this paper we show how the multi-dimensional regression-discontinuity

---

[12] For example, Leuven et al. (2007) examine the effects of a policy in the Netherlands in which schools receive extra personnel funding for having at least 70% minority students and extra computer funds for having at least 70% of students from any single minority group. Outside of

approach provides more information than more conventional strategies such as creating a single composite forcing variable or focusing on one of the assignment variables.

The approach we describe does have limitations. First, like any regression-discontinuity design, it has limited external validity: causal effects are only identified for observations in the immediate vicinity of either cut score. Second, statistical power is a key issue. Estimating these effects precisely requires a substantial density of data points near the joint cut scores. Even with more than 200,000 observations available in Massachusetts datasets, our analysis produces estimates with relatively large standard errors because few observations are local to both cutoffs and those are not distributed equally across each of the four treatment conditions. However, the severity of this limitation depends in large part on the location of the cut score in the joint distribution of the forcing variables. For example, fitting our full model at a pseudo-discontinuity declared at the mean of each of the assignment variables, where the distributions are much denser, produces much more precise estimates, with standard errors reduced by a factor of three. Thus, the particular data burdens of our question arise in large part because the cut scores are in the tails of the forcing variable distributions. These challenges may be less of a concern in other contexts where researchers may want to apply this approach.

Despite these limitations, the approach that we describe in this paper has a number of advantages over more conventional methods for addressing causal questions in situations in which multiple forcing variables assign individuals exogenously to different treatments. In particular, with sufficient data, the method provides a more complete picture of the relationship between different combinations of treatments and the outcome of interest.

---

education, eligibility for different types of insurance or entitlement programs may also be driven by several criteria, such as family size or family income.

References

Center on Education Policy. (20108). *State high school tests: Exit exams and other assessments*. Retrieved April 28, 2011, from http://cep-dc.org/index.cfm?DocumentSubTopicID=8.

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.

Heckman, J.J., & LaFontaine, P.A. (2010). The American high school graduation rate: Trends and levels. *Review of Economics and Statistics, 92*(2), 244-262.

Heckman, J.J., Lochner, L., & Todd, P. (2008). Earnings functions and rates of return. *Journal of Human Capital, 2*(1), 1-31.

Imbens, G., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics, 142*(2), 615-35.

Katz, L.F., & Goldin, C. (2008). *The race between education and technology*. Cambridge, MA: Harvard University Press.

Lee, D.S., & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics, 142*(2), 655-74.

Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2007). The effect of extra funding for disadvantaged pupils on achievement. *Review of Economics and Statistics, 89*(4), 721-36.

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 June 26, 2008, from http://www.doe.mass.edu/mcas/2002/news/01techrpt.pdf.

Massachusetts Department of Education. (2005). *2004 MCAS technical report.* Retrieved June

26, 2008, 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/

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

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

Papay, J.P., Willett, J.B., & Murnane, R.J. (2011). Extending the Regression-Discontinuity Approach to Multiple Assignment Variables. *Journal of Econometrics, 161*(2), 203-207.

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

Reardon, S.F., N. Arshan, A. Atteberry and M. Kurlaender. (2010). Effects of Failing a High School Exit Exam on Course Taking, Achievement, Persistence, and Graduation. *Educational Evaluation and Policy Analysis, 32*(4), 498-520.

Stillwell, R. (2010). Public school graduates and dropouts from the Common Core of Data: School year 2007–08 (NCES 2010-341). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved April 28, 2011 from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010341.U.S.

Census Bureau (2009). *The 2010 statistical abstract: The national data book*. Washington, DC: Author. Retrieved October 28, 2010 from http://www.census.gov/compendia/statab/rankings.html.

Table 1. Estimated causal effects for students near the joint cut scores, from the full multi-dimensional regression-discontinuity model in equation (7) using the optimal bandwidth (h^*_{math} = 7; h^*_{ELA} = 10).

Effect of interest	Parameter Sum	Estimate	Standard Error	*P*-Value*
Fail both exams vs. Pass both exams	$\beta_1 + \beta_2 + \beta_3$	0.076	0.033	0.011
Fail vs. Pass ELA for students who fail math	β_2	0.054	0.035	0.063
Fail vs. Pass ELA for students who pass math	$\beta_2 + \beta_3$	0.045	0.033	0.089
Fail vs. Pass Math for students who fail ELA	β_1	0.032	0.040	0.214
Fail vs. Pass Math for students who pass ELA	$\beta_1 + \beta_3$	0.022	0.027	0.208

*p-values are reported from one-sided hypothesis tests.

Table 2. Estimated effect of barely passing both exams and of just passing the ELA exam, for students who just failed and just passed the mathematics exam, from the full model in equation (7), at selected bandwidths (standard errors in parentheses).

h_{math}	h_{ela}	N	Estimated Effect Of Passing Both Exams	Estimated Effect of Passing ELA Exam	
				Fail Math	Pass Math
5	9	15,588	0.053 (0.050)	0.040 (0.054)	0.020 (0.040)
5	10	17,276	**0.061** (0.047)	0.045 (0.051)	0.017 (0.038)
5	11	18,865	0.049 (0.044)	0.031 (0.049)	0.026 (0.036)
5	12	20,405	**0.058** (0.043)	0.048 (0.047)	0.022 (0.035)
6	9	18,010	**0.068** (0.044)	0.046 (0.048)	0.024 (0.038)
6	10	19,963	***0.085*** (0.042)	**0.064** (0.045)	0.024 (0.036)
6	11	21,834	***0.069*** (0.040)	0.050 (0.043)	0.034 (0.034)
6	12	23,657	***0.080*** (0.038)	**0.066** (0.042)	0.030 (0.033)
7	9	20,367	**0.058** (0.041)	0.034 (0.044)	0.040 (0.036)
7	10	22,569	***0.076*** (0.038)	**0.054** (0.041)	**0.045** (0.034)
7	11	24,691	**0.059** (0.037)	0.040 (0.040)	**0.047** (0.033)
7	12	26,783	***0.064*** (0.035)	**0.049** (0.038)	**0.045** (0.031)
8	9	22,360	***0.072*** (0.038)	0.050 (0.041)	0.025 (0.035)
8	10	24,822	***0.083*** (0.036)	**0.061** (0.038)	0.030 (0.033)
8	11	27,222	***0.072*** (0.034)	**0.057** (0.037)	0.032 (0.032)
8	12	29,546	***0.075*** (0.033)	***0.064*** (0.035)	0.030 (0.030)
9	9	24,078	**0.056** (0.036)	0.039 (0.038)	0.026 (0.034)
9	10	26,774	***0.059*** (0.034)	0.042 (0.036)	0.031 (0.032)
9	11	29,392	**0.051** (0.032)	0.042 (0.034)	0.034 (0.030)
9	12	31,987	***0.054*** (0.031)	**0.048** (0.033)	0.032 (0.029)

NOTE: Bold and italic, $p<0.05$; Bold, $p<0.10$. P-values are reported from one-sided hypothesis tests.

Figure 1. Surface plot of the estimated values of the cross-validation criterion from equation (5) for the trimmed sample, by possible bandwidths in mathematics (h_{math}) and ELA (h_{ELA}).
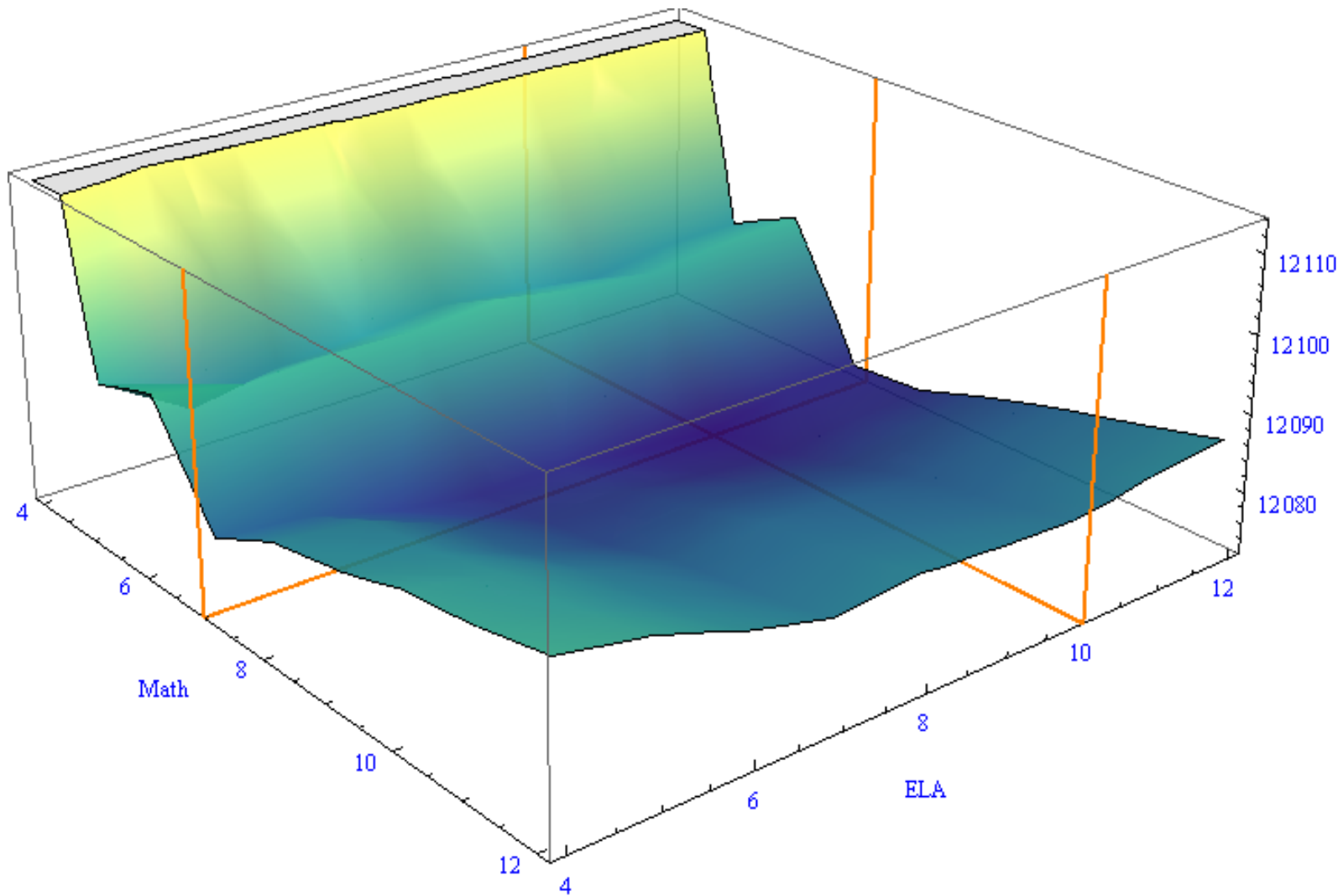
Figure 2. Graphical representation of the estimated causal effect of just passing the Massachusetts mathematics and/or ELA exit examinations, for students near the joint cut scores, from the full model in equation (7).
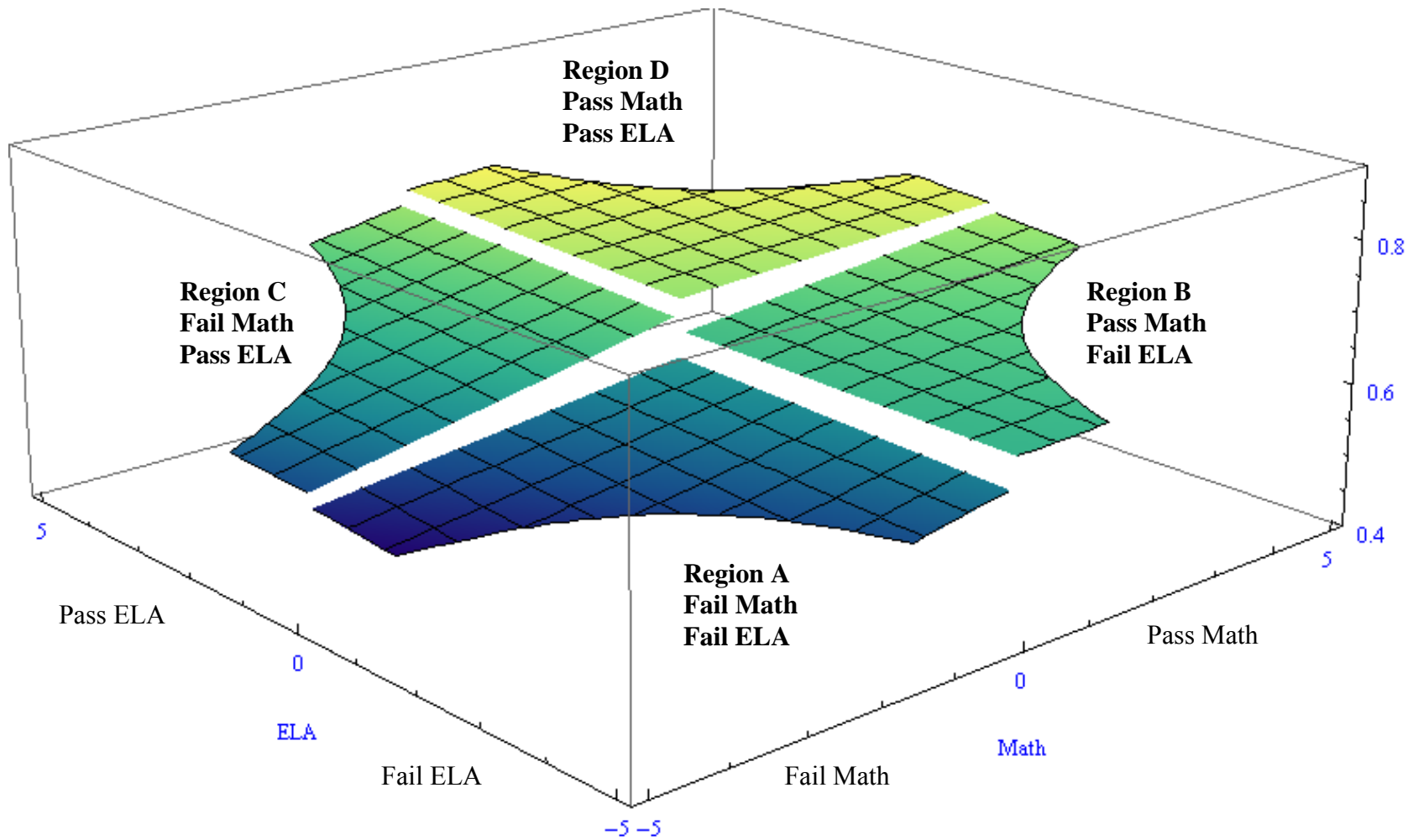
Figure 3. Graphical representation of the fitted probability of on-time graduation for students scoring at the ELA cut score who pass (solid line) and fail (dotted line), by mathematics test score (top panel), with the estimated causal effect of just passing the ELA exit examination by mathematics score for these students near the joint cut scores (bottom panel)
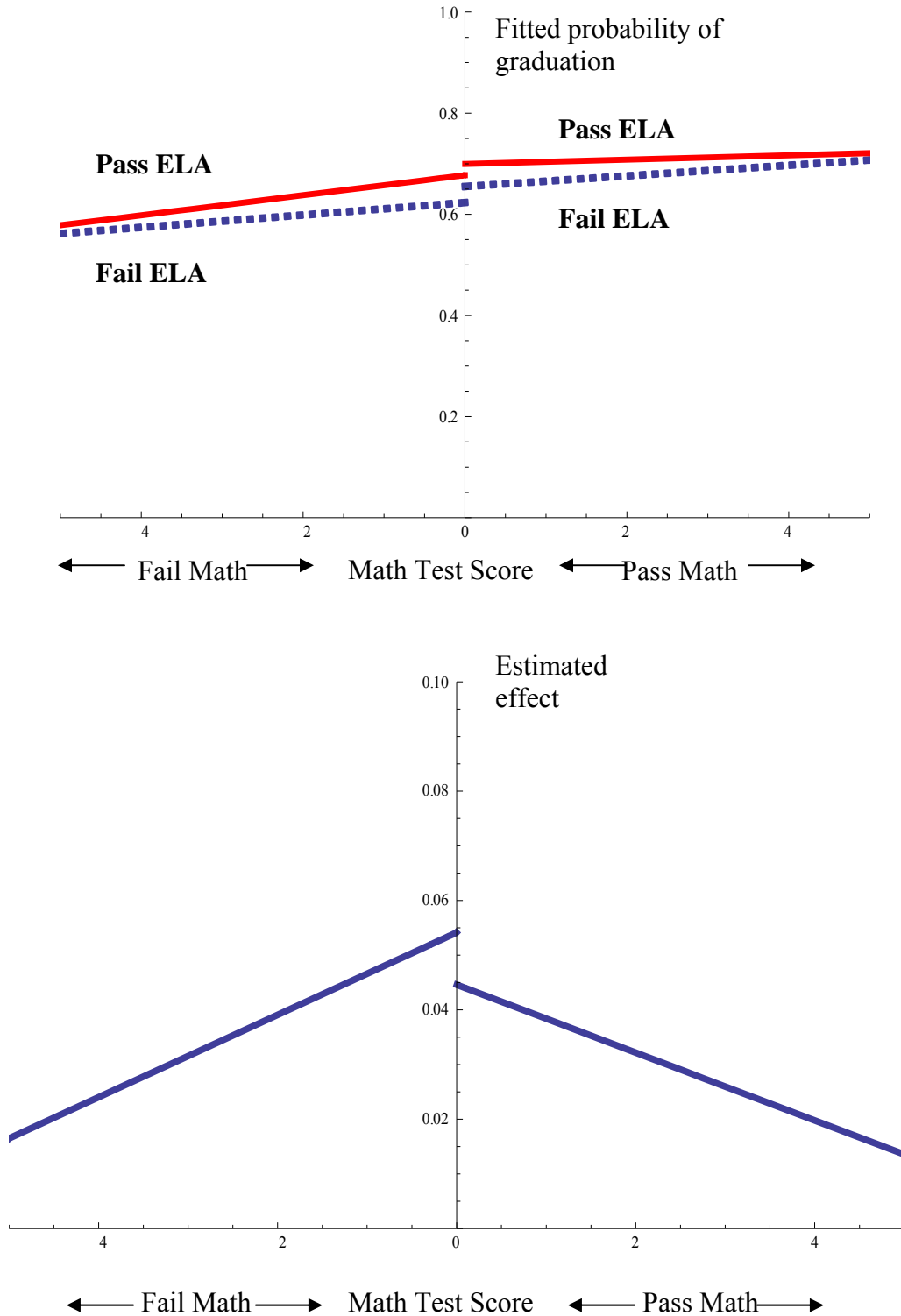
Figure 4. Graphical representation of the estimated causal effect of just passing the ELA exit examination for students near the ELA cut score, by mathematics score.
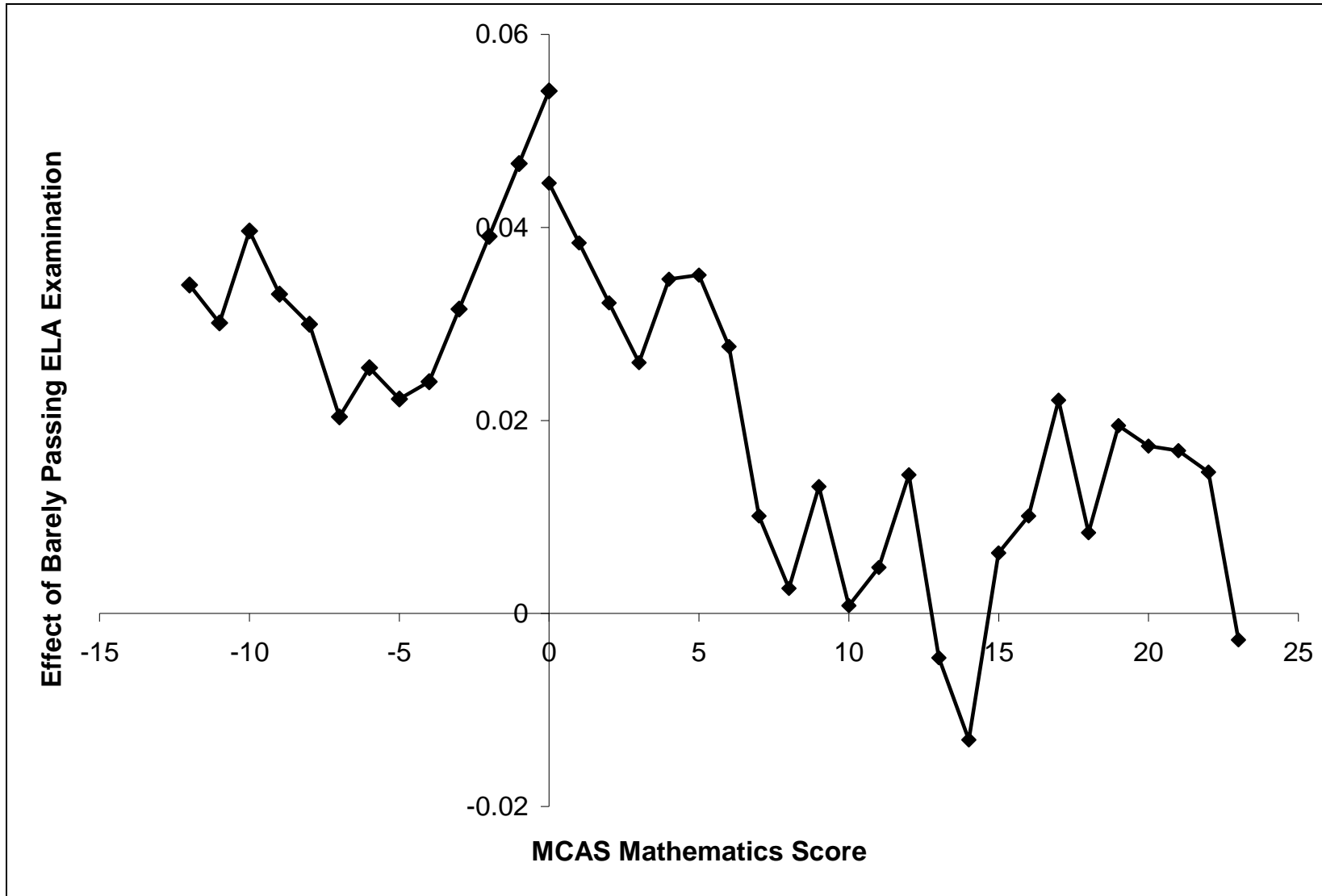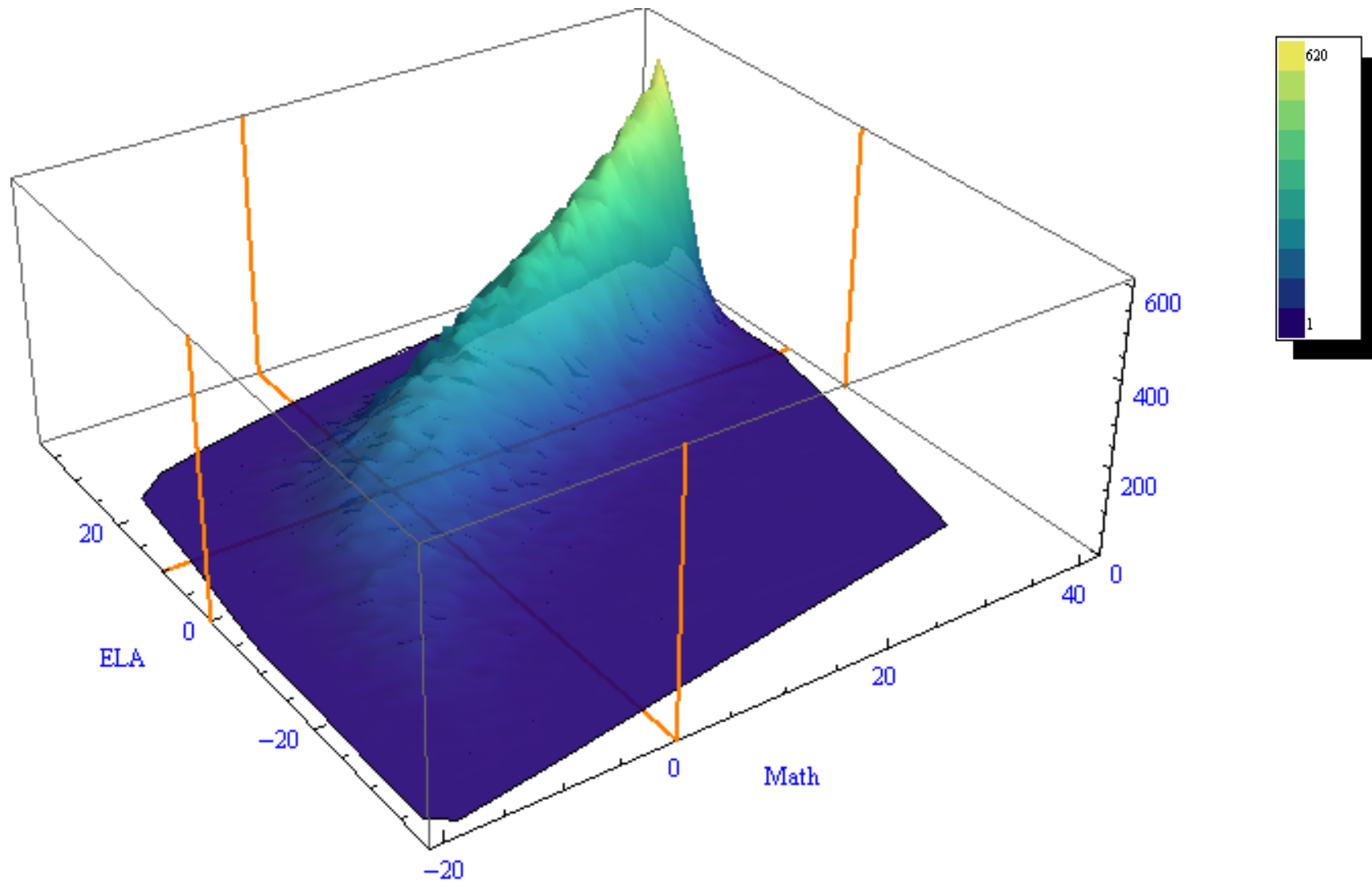
Figure 5. Surface plot of the empirical bivariate distribution of the sample data, by mathematics and ELA scores, with lines at the pass-fail cut scores.

Appendix
Implementing the Multi-Dimensional Regression-Discontinuity Approach

In any regression-discontinuity approach, we seek to estimate the mean outcomes for individuals at the cutoff in the treatment and control group (see equation (1)). Estimating these conditional means requires us to make assumptions about the relationship between our outcome and our forcing variables near the cut scores. Researchers have recently begun to explore "nonparametric" or "semi-parametric" approaches that do not rest on strong functional form assumptions (e.g., Ludwig & Miller, 2007; Imbens & Lemieux, 2008; Lee & Lemieux, 2010). As our parameters of interest are at the boundary of support on either side of the cutoff, we estimate these limits with local linear regression (Fan, 1992; Hahn, Todd, & Van der Klaauw, 2001; Porter, 2003). We first choose a bandwidth to govern the smoothing and then fit a single (locally) linear regression of the following form:

$$p[GRAD_i = 1] = \alpha_0 + \alpha_1 PASS\_MATH_i + \alpha_2 MATH_i^c +$$
$$\alpha_3(MATH_i^c \times PASS\_MATH_i) + \gamma'Z_i + \varepsilon_i \qquad \text{(A-1)}$$

using only observations that fall within one bandwidth on either side of the cutoff.[13] We can interpret α_1 as the effect of passing the mathematics examination, instead of failing it, for students near the cutoff, in the population.

In our case, however, these mathematics and ELA examinations define four different "treatment" conditions because students can either pass or fail each test. We are interested in all of the pairwise comparisons among these treatments. Estimating separate effects, as in equation (A-1) implicitly blurs important distinctions among these different treatment conditions. Understanding the relationship between exit examinations and graduation in a more nuanced manner requires us to examine simultaneously the consequences of all four "treatment"

---

[13] We describe this approach in detail in Papay, Murnane, and Willett (2010).

conditions.

In Papay, Willett, and Murnane (2011), we propose a general approach to do so by incorporating discontinuities in multiple forcing variables into a single analysis. Here, we describe in detail how we implement that strategy for this paper. The four treatment conditions define four separate regions in the two-dimensional space spanned by the forcing variables, ($MATH^C$, ELA^C). Again, as in the case with a single forcing variable, our parameters of interest are the conditional mean probabilities of graduation for individuals in each treatment condition, at the cutoff. To estimate these limits, and the difference between them, we use a two-step process in which we first choose an "optimal" joint bandwidth around the cut-off on the forcing variables (labeled h_1^*, h_2^*) and then estimate the causal effect by conducting a local linear regression analysis using this optimal bandwidth.

Bandwidth Selection

The primary challenge in implementing our approach comes in choosing the appropriate bandwidths (h_1^*, h_2^*) for our analysis. For each observation, at each point on the ($MATH^C$, ELA^C) grid, we fit a linear regression model – within an arbitrary bandwidth (h_1, h_2) – to estimate a fitted value of the outcome at that point:

$$\hat{\mu}(MATH_i^C, ELA_i^C, h_1, h_2) = \hat{\gamma}_0 + \hat{\gamma}_1 MATH_i^C + \hat{\gamma}_2 ELA_i^C + \hat{\gamma}_3 (MATH_i^C \times ELA_i^C) \quad \text{(A-2)}$$

In each case, we attempt to mirror the regression-discontinuity approach by only using observations that fall within the appropriate region, and estimating $\hat{\mu}(MATH_i^C, ELA_i^C, h_1, h_2)$ as if it were a boundary point. In other words, for a student who fails both tests and for whom $MATH^C$=-3 and ELA^C=-5, we use only observations for which $MATH^C$<-3 and ELA^C<-5. By contrast, for students who pass both tests and for whom $MATH^C$=10 and ELA^C=15, we use only observations for which $MATH^C$≥10 and ELA^C≥15.

For a given bandwidth (h_1, h_2), we thus estimate a fitted probability of graduation for each observation. We compare these fitted values to the observed values, across the entire sample, using a generalized Imbens & Lemieux (2008) cross-validation criterion:

$$CV_{GRAD}(h_1, h_2) = \frac{1}{N} \sum_{i=1}^{N} (GRAD_i - \hat{\mu}(MATH_i^C, ELA_i^C, h_1, h_2))^2 \qquad \text{(A-3)}$$

Our optimal joint bandwidth, h_1^* and h_2^*, is the pair of bandwidths that minimizes the CV criterion.

Because data are less dense in the tails of the distribution, including observations far from the cut score in our bandwidth estimation may lead us to select a larger bandwidth than necessary. As a result, Imbens & Lemieux recommend deleting observations selectively that fall beyond a certain quantile (δ) on either side of, and most remote from, the discontinuity before implementing the cross-validation procedure described above. Here, we set δ to 25% because the minimum passing scores are in the tails of the joint distribution. This is the key parameter that we are choosing in our analysis; as a result, it is important to assess the sensitivity of our findings to the choice of δ. In practice, smaller values of δ (i.e., excluding fewer observations) will produce somewhat larger optimal bandwidths. Thus, we investigate whether key findings are consistent across a range of bandwidths.

We conduct this trimming process separately at each value of the forcing variables. In other words, at each value of $MATH^C$, we determine the 25th percentile of ELA scores for observations below the ELA cut score and the 75th percentile for observations above the ELA cut score. We follow a similar process, estimating relevant quantiles of mathematics score at each value of ELA^C. Then we exclude simultaneously all observations with $MATH^C$ or ELA^C scores more extreme than either of these estimated quantiles, on either side of the cut score.

Given that students can only earn integer scores on the tests, our forcing variables have discrete values and therefore we can define a finite set of plausible bandwidths. Thus, instead of using a more complicated multidimensional optimization routine, we simply estimate the values of the cross-validation criterion for each possible pairwise combination of bandwidths. We estimate an optimal bandwidth of $h^*_{MATH} = 7$ and $h^*_{ELA} = 10$.

Estimation

This procedure uses the entire sample to develop optimal bandwidths for an analysis in which the objects of interest are the parameters at the joint cut scores. The two cut scores define four regions of interest: students can either pass or fail both the mathematics and ELA examinations. Although we could fit four separate models, one for each of these treatment regions, we simply fit the requisite regression models in each region simultaneously, by specifying a single statistical model with 16 parameters – an intercept and slope parameters to accompany all 15 possible interactions among $MATH^C$, ELA^C, $PASS\_MATH$, and $PASS\_ELA$. We then fit locally linear probability models[14] of the following form, using observations whose mathematics and ELA scores fall within one optimal bandwidth on either side of the relevant cut-scores:

---

[14] Given our dichotomous outcome, we could use a local probit functional form to model these effects. However, with our narrow bandwidth and the fact that on-time graduation rates are approximately 60% for students near the joint cutoff, results obtained under the linear and nonlinear models are almost identical. As a result, we present the linear probability model results for ease of interpretation.

$$p[GRAD_i = 1] = \beta_0 + \beta_1 PASS\_MATH_i + \beta_2 PASS\_ELA_i + \beta_3(PASS\_MATH_i \times PASS\_ELA_i)$$
$$+ \beta_4 MATH_i^c + \beta_5 ELA_i^c + \beta_6(MATH_i^c \times ELA_i^c) + \beta_7(MATH_i^c \times PASS\_MATH_i)$$
$$+ \beta_8(ELA_i^c \times PASS\_ELA_i) + \beta_9(MATH_i^c \times PASS\_ELA_i) + \beta_{10}(ELA_i^c \times PASS\_MATH_i) \qquad \text{(A-4)}$$
$$+ \beta_{11}(MATH_i^c \times ELA_i^c \times PASS\_MATH_i) + \beta_{12}(MATH_i^c \times ELA_i^c \times PASS\_ELA_i)$$
$$+ \beta_{13}(MATH_i^c \times PASS\_MATH_i \times PASS\_ELA_i) + \beta_{14}(ELA_i^c \times PASS\_MATH_i \times PASS\_ELA_i)$$
$$+ \beta_{15}(MATH_i^c \times ELA_i^c \times PASS\_MATH_i \times PASS\_ELA_i) + \gamma'Z_i + \varepsilon_i$$

Here, we include the set of student-level covariates described above (Z_i) to improve the precision of our estimation. We can interpret this single model *parametrically* for observations local to the cut score and use the standard errors to conduct appropriate statistical tests. Following Lee & Card's (2008) admonition to account for the discrete nature of assignment variables, we treat each combination of mathematics and ELA scores as a cluster in computing the standard errors. In all cases, we expect students who pass the test to have better outcomes than students who fail; as a result, we make use of one-sided tests throughout our analysis, reporting the corresponding *p*-values.

Appendix References

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

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.

Lee, D.S., & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics, 142*(2), 655-74.

Lee, D.S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature, 48*(2), 281-355.

Ludwig, J., & 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.

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

Papay, J.P., Willett, J.B., & Murnane, R.J. (2011). Extending the Regression-Discontinuity Approach to Multiple Assignment Variables. *Journal of Econometrics, 161*(2), 203-207.

Porter, J. (2003). Estimation in the Regression Discontinuity Model. Unpublished working paper: Author.