Economics of Education Review 38 (2014) 64–75

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

A regression discontinuity analysis of graduation standards and their impact on students' academic trajectories☆

Tom Ahn

Department of Economics, University of Kentucky, United States

ARTICLE INFO

Article history:
Received 18 July 2012
Received in revised form 3 November 2013
Accepted 16 November 2013
Available online 25 November 2013

JEL classification:
I24
J48
J62

Keywords:
Exit standards
Academic trajectory
Regression discontinuity

ABSTRACT

In 2006, North Carolina put in place high school exit standards requiring students to pass a series of high-stakes exams across several years. I use a regression discontinuity (RD) approach to analyze whether passing or failing one of these exams (Algebra I) impacts a student's decision between choosing a more rigorous college-preparatory math curriculum and an easier 'career' track math curriculum. I find a 5 percentage point gap in the probability of selecting the rigorous curriculum between 9th grade students who just passed and those who just failed the exam. RD results across two years (one year in which the graduation standards were not in place) suggest that the discontinuity arose due to fewer students opting into the college track as a result of the exam results.

1. Introduction

The impact of high-stakes examinations on future student outcomes has been extensively studied in the literature. One group of studies has analyzed the use of standardized exams as a means of strengthening graduation standards. Imposing or strengthening high school exit standards through the use of high-stakes exams has the potential to affect various student outcomes, including graduation and dropout rates, college matriculation, and labor market outcomes, with theoretical arguments for both positive and negative impacts. Introduction of strict standards may lead to negative outcomes. Marginal students may decide that the additional study required to clear the bar is not worth the effort, leading to lower graduation and college matriculation rates. On the other hand, the presence of standards may motivate marginal students and their teachers to additional exertions to make sure that they qualify for graduation. If additional effort results in real academic gains for students, this can lead to increase in graduation and college matriculation rates, as well as better labor market outcomes.

In general, the empirical literature has failed to find consistently large observable impacts of these exams. Some studies have found negative impacts of failing an exit exam on on-time graduation and increased drop-out.[1] Some tie the increase in the number of dropouts to other negative externalities in society, such as increased crime rate arising from poverty.[2] Many studies have found small positive impacts on academic achievements and earnings.[3] Indeed, many papers find that exit standards have not

☆ I am grateful to the North Carolina Education Research Data Center at the Sanford School of Public Policy at Duke University for access to the NCDPI data set. I thank two anonymous referees for helpful comments. All remaining errors are my own.

E-mail address: thomas.ahn@uky.edu.

[1] See for example, Jacob (2001), Dee (2003), among others.
[2] See Larson (2011).
[3] See for example, Woessmann, Ludemann, Schutz, and West (2007), and Bishop and Mane (2001).

0272-7757/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.econedurev.2013.11.005

caused large changes in graduation, dropout, or college attendance, in either direction.[4] Recent work by Papay, Murnane, and Willett (2011) using regression discontinuity find that barely failing an exit exam results in a substantial reduction in the likelihood of graduating on time. Another paper by the same authors (2010) find, again using regression discontinuity, that being labeled as ''advanced'' results in higher on-time high school graduation and college enrollment.[5]

Many studies of graduation standards analyze the impact of a graduation exam that is usually administered toward the end of a student's senior year. This study examines a different graduation standard that was briefly in place in North Carolina. A series of End-of-Course (EOC) standardized exams on various subjects were administered throughout a student's high school career as part of the state's accountability system. From 2006 to 2009, the state used the results from some of these exams as its high school graduation standard: a student needed to score above a cut-off value on five exams (or be granted a waiver) to graduate.

While this system sounds demanding, in actuality, a student was given multiple chances to pass each exam, with extensive tutoring offered for those who needed help.[6] Passing EOC exams did not serve as a prohibitive barrier for students who were marginal between passing and failing these exams.[7]

Instead of (or in addition to) serving to sharpen the signal sent by a high school degree to post-secondary institutions and future employers, the results from these EOC exams may have helped the student and the school adjust future academic plans. If student performance on a high-stakes exam was viewed as a good measure of his or her academic ability, the test outcome may have directly impacted a student's selection of courses during his or her high school career and post-high school plans, even if poor performance on the exam did not automatically prevent graduation. A good outcome may encourage students to try their hands at tougher courses, while a bad outcome may deter students from pursuing a more rigorous curriculum. This study investigates whether 10th grade students were influenced by the results from an Algebra I EOC exam in selecting between a college preparatory math track and a career focused math track.

For the North Carolina high-stakes exams, the signal is most useful to students near the pass/fail cut-off. Students who are far below the bar may use the multiple opportunities and extra help to pass the exam and graduate. The bad outcome merely confirms for the students and the schools that they are located in the left-tail of the grade distribution. Similarly, students who are far above the bar do not change their academic plans as a result of passing the exams. The discrete pass/fail signal is most useful for students who may be unsure whether they should or should not pursue a more rigorous curriculum that can prepare them for post-secondary education.

This study focuses on students that are on the edge between passing or failing a high-stakes exam on their first try. The academic trajectories of these students may be especially malleable for two reasons. The gains to taking career preparatory math (which includes expected labor market outcomes and disutility from taking more difficult math courses) for these marginal students may be comparable to the expected gains from taking math courses designed for post-secondary education preparation. Alternatively, the lack of information on their academic potential may make valuation of taking tougher courses more difficult. In either case, the outcome of a standardized exam that is regarded as important by the state may strongly impact whether the student is placed in a more rigorous academic environment.[8]

If there is an impact on the likelihood that a student will move into a more rigorous math sequence based on the Algebra I exam results, this may also be attributable to the school's response to the signal. Students may be tracked into different math sequences, at least partially based on whether they passed or failed the exam. Students who failed may be counseled away from tougher courses, and learning resources may be diverted to lower achieving students (to get them closer to passing the Algebra I EOC exam) or higher achieving students (to strengthen their academic portfolio). Whether the response comes from the student or the school, a student who may be capable of being successful in a more difficult math sequence may be pushed away due to the outcome of the standardized exam.[9]

A regression discontinuity (RD) framework is particularly appealing here, as the high-stakes standardized exam is based on a continuous scale, yet whether a student passes or fails is determined by a sharp cutoff in exam scores. Two students with virtually identical abilities may send a positive or negative signal to the school

[4] See for example, Grodsky, Warren, and Kalogrides (2009), Reardon, Atteberry, Arshan, and Kurlaender (2009), Warren and Edwards (2005), Dee and Jacob (2006), Carnoy, Loeb, and Smith (2001), Clark and See (2011), Jurges, Schneider, Senkbeil, and Carstensen (2012) among many others.

[5] For a complete review of the literature, see Holme, Richards, Jimerson, and Cohen (2010).

[6] Even after multiple (sometimes as many as five or six) failures, a student was not automatically prevented from graduating. He or she was evaluated by a committee of teachers and administrators and given a final recommendation, where one of the possible outcomes was being granted a waiver on passing the exams. School administrators may have had some motivation to push students through, as low graduation rates negatively impacted a school's adequate yearly progress (AYP) status and reflected badly on the principal's professional capabilities.

[7] Over 95% of 9th grade students in the sample require at most two attempts to pass the Algebra I exam.

[8] It is worth noting that students and schools may *perceive* passing or failing in different ways. A pessimistic student (or school counselor) may perceive the signal as bad news and move away from taking a rigorous course-load, while an optimist may perceive the exam result as a wake-up call and be motivated to work harder, pushing up (or at least not decreasing) the likelihood of opting into the rigorous curriculum. Both pessimistic and optimistic students who pass will respond by taking the more rigorous sequence. An estimated decline in the probability of taking a rigorous course-load in response to the *perception* from failing will be underestimated for pessimistic students and overestimated for optimistic students. Therefore, the treatment is the pass/fail signal itself and not the student's (or school counselor's) perception of the signal.

[9] Conversely, if the student was unlikely to do well in higher math courses, the student may have been dissuaded to his or her 'benefit.'

(or themselves) about their capacity to take on challenging math courses, based on whether they achieve 'sufficient mastery' on an EOC exam.

The results of the RD analysis show that there is indeed a sizable impact of failing the standardized exam. Compared to those who just pass the high-stakes exam, students who just fail are about 5 percentage points less likely to pursue an academically challenging math curriculum that would, according to North Carolina's Department of Public Instruction (NCDPI), prepare them for post-secondary education.

While other studies of exit standards have demonstrated impact in the positive or negative direction in other states (see Martorell, 2004; Ou, 2010; Papay, Murnane, & Willett, 2010; Reardon et al., 2009), the results in this study are buttressed by two important supplemental RD results.

Insignificant RD results with the same standardized exam in the year prior to the implementation of exit standards suggest that the impact does not arises from the existence of the exam itself. Instead, it seems to come from the signal that students or schools receive about how the state views the fail/pass outcome of the exam.

In addition, with the year before the graduation standard implementation as a base line, it may be possible to divide the discontinuity between negative impact from failing and the positive impact from passing. RD results suggest that most of the discontinuity is generated by the negative signal from failure. However, the temporal stability assumptions required for interpreting the results in this way is daunting, and testing for the robustness of these assumptions proved inconclusive. Caution should be exercised in interpreting the results too strongly.

Section 2 describes the administrative details of the North Carolina standardized exams and exit standards. Section 3 presents the data. Section 4 explains the results and Section 5 concludes.

2. North Carolina End-of-Course exams

North Carolina public school high school students enrolled in certain classes must take EOC exams to evaluate whether they have an 'adequate' level of knowledge as specified in the North Carolina Standard Course of Study. The exam scores are used for state accountability program evaluations as well as student grades. North Carolina's ABC accountability program uses EOC scores to hand out cash bonuses to teachers at high schools.[10] The scores are aggregated at the school level to gauge year-over-year academic achievement growth. If the school had a larger gain than the state specified bar, the school is paid a cash bonus. The system is two-tiered, with teachers in schools making 'expected growth' earning $750, and teachers in schools making 'high growth' being paid $1500.[11]

EOC exam results are also used in student assessment. Test scores are included in students' permanent records and high school transcripts. In addition, the exam scores make up 25% of the final grade in the respective course. Students who first entered high school in or after 2006 are also subject to a graduation requirement, part of which requires performing at a 'consistently demonstrated level of achievement' in five EOC exams (Algebra I, Biology, English I, Civics and Economics, and U.S. History). This equates to attaining level III or higher on a four-level scale. In this sense, the stakes in EOC exams are high not only for schools and teachers, but students as well.

Should a student fail to achieve level III or above on an EOC exam, he or she must be re-tested within three weeks of the receipt of the exam results. As the exam is usually given toward the end of a semester (whether fall or spring), the re-test is usually given during the 'next' semester (spring semester for students who took the exam in the fall, and summer for those who took the exam in the spring). Should the student fail to achieve level III or above again, he or she is offered another chance after tutoring/remediation. If the student should fail again, a committee consisting of teachers, principals, and district staff members decides whether to grant a waiver or force the student to retake the course.

2.1. Achievement levels

EOC exam results, in addition to a continuous score, are categorized into one of four achievement levels. Level I indicates that the student does not have sufficient mastery of the subject. Level II indicates inconsistent mastery. Level III represents consistently demonstrated mastery, and Level IV shows consistent performance in a superior manner. In general, a student is judged to have 'passed' the exam if he or she achieves level III or above. In particular, a student cannot graduate unless he or she has attained level III or higher on the Algebra I EOC exam. It should be noted that failing the EOC exam is not equivalent to failing the course. For the 2006 sample of 9th grade students who are within 5 points of the cut-off on either side (out of a total of 80 possible points), their modal final grade in Algebra I was a C.

2.2. High school math sequence in North Carolina

The high school math curriculum is broadly defined into three tracks of studies. The most rigorous track (College/University Prep) requires at least 4 math classes, with one of these courses being an upper level math course that has Algebra II as a prerequisite (usually Pre-Calculus or Calculus). The middle track (College Technical Prep) requires at least 3 courses, with Algebra II, Technical Math II, or Integrated Mathematics III as the top advanced course. Finally, the least rigorous track (Career Prep) requires any 3 math courses, as long as Algebra I is included. Of course, as the graduation requirement, all three tracks require the student to pass the Algebra I EOC exam.

NCDPI-provided guidance for the math tracks states that the College/University Prep track is the minimum appropriate preparation for matriculation to a 4-year college or university. The College Technical Prep track has the minimum appropriate courses to move on to a 2-year post-secondary school, and the Career Prep track is the

[10] The acronym stands for Accountability, teaching the Basics, and emphasis on local Control.

[11] For a more complete description of the accountability system, see Ahn (2013).

Table 1
Summary statistics.

Variable	2006 sample (Std. Dev.)	2005 sample (Std. Dev.)	Relevant group (Std. Dev.)
Minority	0.370 (0.483)	0.348 (0.476)	0.491 (0.500)
Female	0.520 (0.500)	0.518 (0.500)	0.526 (0.499)
Alg I score	7.846 (8.428)	8.888 (8.676)	1.116 (3.007)
Alg I level	3.211 (0.738)	3.283 (0.726)	2.709 (0.454)
Best Alg I score	8.105 (8.192)	9.347 (8.234)	1.386 (3.152)
Best Alg I level	3.234 (0.714)	3.325 (0.681)	2.737 (0.455)
# of Times Took Exam	1.102 (0.352)	1.085 (0.317)	1.136 (0.396)
Career Prep/No Math	0.070 (0.254)	0.052 (0.221)	0.104 (0.305)
Technical Prep Math	0.144 (0.351)	0.125 (0.331)	0.236 (0.425)
University Prep Math	0.786 (0.410)	0.823 (0.381)	0.660 (0.474)
Observations	40,686	34,488	13,319

minimum recommended coursework to enter the labor force directly out of high school.[12]

3. Data

I use an administrative data set for the North Carolina public school system from 2005 to 2007. The data set contains information on all public schools, students, teachers, and administrators in North Carolina. Because the data is collected annually, and individuals can be matched across years, a relatively complete panel data set of the public school system in North Carolina emerges.

There are two unique features of the data set that I take advantage of to identify the impact of the EOC exams on a student's academic trajectory. The first is that each student who takes an EOC exam has a unique identifier that can be linked to a transcript file in subsequent years. This allows me to track the course selection of students after the results of the exams. For this study, I focus on math courses the student takes after the Algebra I exam.

The second unique feature of the data is that the high school exit standard was implemented for all students entering 9th grade in 2006, but students who entered high school prior to this year were exempt from this requirement. This allows me to use 9th graders in 2005 to further test whether the pass/fail impact is real, as performance on EOC exams for these students only matter for current course grade.

I restrict Algebra I test takers in the data to those students who took their first exam in 9th grade.[13] This effectively eliminates students who took the exam in 8th grade or prior. These students represent high achievers, and as such, the high school exit requirement is not a difficult standard to satisfy. Summary statistics for 9th grade students in the 2006 and 2005 samples are shown in the first two columns in Table 1, respectively.[14]

"Best" score and level are the maximum of all Algebra I test score results for the student, as he or she is allowed to take the exam multiple times. The last three rows indicate the fraction of students who took 'Career Prep,' 'Technical Prep,' or 'University Prep' math during their sophomore year.

3.1. Description of the relevant student sub-population

Who are the students at the border of achievement levels II and III? As Fig. 1 shows, where the scaled score of 0 indicates the demarcation between passing and failing, these students are between the 10th and 40th percentile of the exam score distribution. As the last column in Table 1 shows, among students near the cutoff, over 65% take University Prep math courses and over 23% take Technical Prep math courses.[15] This seems to indicate that many of these students are weighing taking a more rigorous math course sequence with an eye toward preparation for post-secondary education.[16]

3.2. Do math EOC exam results and subsequent course rigor matter?

The assignment variable used in the analysis is the result of the Algebra I EOC exam. While the regression discontinuity analysis generates a statistically significant gap, showing that the exam results do matter, it would be more meaningful if there was true confirmation that schools, teachers, and students take the signal from the pass/fail outcome seriously.

From discussions with high school teachers in North Carolina as well as reports produced by NC WISE (North Carolina Window of Information on Student Education), the arm of NCDPI that manages student information including test scores and graduation rates, it is clear that

[12] Many students in the Career Prep track choose to take math courses more oriented toward vocational training.

[13] The 10th grade student sample used for robustness checks are also limited to those who take the Algebra I exam for the first time in 10th grade.

[14] The smaller number of observations in the 2005 year sample reflects the lower success rate in matching students to their transcripts. The mismatch was essentially random across geographic and demographic characteristics. For further details, see North Carolina Education Research Data Center (NCERDC: www.pubpol.duke.edu/centers/child/nceddata-center.html).

[15] This subsample is composed of students with first-time EOC standardized score between −5 and 5.

[16] Overall, of the students who take the Algebra I exam for the first time in 2006, 20.1% are 8th graders, 58.7% 9th graders, 16.4% 10th graders, 3.7% 11th graders, and 1.0% are 12th graders. Since the sample excludes 10th, 11th, and 12th grade exam takers as well as those who never attempt the exam, the unconditional distribution of ability would probably place them further to the middle (or right) of an overall hypothetical test score distribution (if it were possible to force all 9th graders to take the Algebra I exam at the same time).

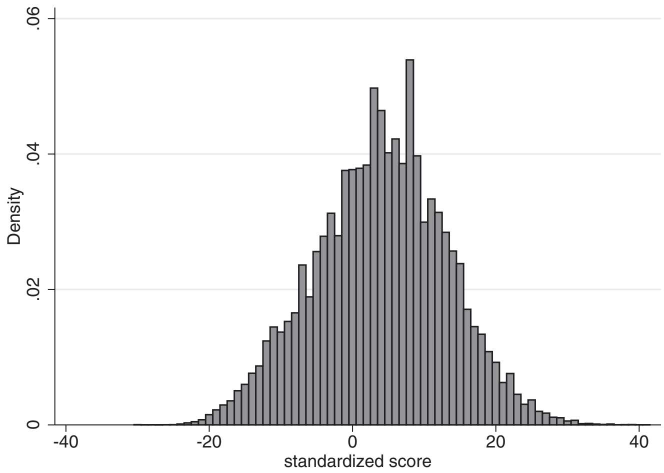

Fig. 1. Assignment variable density.

Algebra I results are regarded as an important indicator of risk in a student's academic career.[17] In fact, along with low GPA and a high rate of absence, low Algebra I test results (levels I and II) for 10th, 11th, and 12th grade students are highlighted as the best signals to identify students at risk of dropping out.[18] With so much attention paid to Algebra I, it is not surprising that students (or their teachers) may reevaluate their academic plans based on exam performance.

The outcome variable under consideration in the RD analysis is whether the student chooses to pursue a rigorous math course in his or her sophomore year that would put the student on track for post-secondary education. As this variable is not the standard outcome variable considered in many graduation exam studies, it is fair to ask if this is an outcome worth using.

There are limitations in using traditional outcome measures, such as standardized test scores in the 'next' year, graduation, college matriculation, or intent to graduate/matriculate to college. For North Carolina high school students, the next math exam after the Algebra I EOC exam is the Algebra II EOC exam. Because different students take Algebra II at different times, using the Algebra II EOC exam performance as the outcome measure is problematic. In addition, Algebra II is not a required course for students unless they plan to pursue post-secondary education. Excluding students who never take Algebra II from analysis would severely truncate the distribution of students due to self-selection. For the same reason, intent to pursue further education is impossible to capture consistently for all students. Besides students who drop out (and would not be captured), future academic plans are only recorded if the student takes an EOC exam. Besides the English I exam (which most students take in 9th grade), the other exams may be taken at anytime during a student's high school career. This makes it difficult to consistently capture academic plans for the 9th grade sample at the same time in the future.

Data issues preclude the analysis of longer term outcomes, such as graduation, college matriculation, or labor market outcomes. The data set does not contain post-high school outcome information and attempting to gauge the impact on graduation proved problematic. Linking across three more years led to a loss of approximately 40% of the original sample of 9th graders. This includes both actual attrition (drop out) as well as loss from the data set (due to out of state or private school transfers and data matching problems across 4 years).[19]

Outside of being able to ask students *after* the exam results are revealed how and why their academic and future plans have changed based on this new information, a plausible way to get at the students' response is to look at their math course trajectory.[20] The student's choice of math curriculum may serve as an observable indicator of his or her *intention* to prepare for and possibly pursue post-secondary education. Specifically, there are no different requirements across the three tracks in English, science, or social studies. There is a foreign language requirement for the four-year college track (not for the community college track), but only two semesters in any one language is required. Mathematics is the only subject where courses are specifically divided into the different college readiness level tracks.[21] Among 9th grade students who went on to take a rigorous math sequence in 10th grade, 79% had plans

[17] For instance, see http://www.ncpublicschools.org/docs/graduate/resiliency/materials/20101119ncwise.pdf.

[18] It is worth emphasizing that failing the Algebra I EOC exam for the first time in 9th grade is *not* a risk-factor for drop out. A student with an achievement level of I or II in 10th grade (who had taken Algebra I in 9th grade) would imply that he or she had failed the exam at least twice.

[19] A plot of attrition rate at each test score shows that attrition does not sharply change at the discontinuity, indicating that the RD estimates are valid with the attrition. See online appendix at http://sites.google.com/site/tomsyahn/.

[20] It should be noted that there is information about a student's post-high-school plans before the exam (employment, community college, four-year college), but there is no information on how those plans change post-exam (unless they take the Algebra II exam).

[21] It is also possible that the test outcome is only interpreted as a signal on the student's aptitude for math, independent of college aspirations.

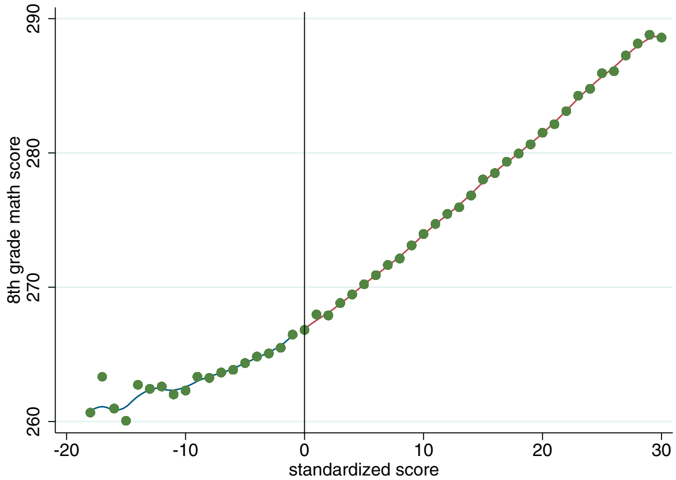

Fig. 2. Placebo 8th grade math score. Figure is generated with local polynomial of degree zero. Covariates are excluded. Observations for standardized scores less than −20 or greater than 30 (which compose approx. 5.7% of observations) are dropped for presentation purposes.

(prior to taking the Algebra I EOC exam) to matriculate to a 2-year or 4-year post-secondary institution. Among 9th graders who did not take a rigorous math sequence in their sophomore year, 61% indicated (prior to taking the Algebra I EOC exam) that they planned go onto college.[22]

3.3. Validity checks for RD

Following Imbens and Lemieux (2008), validity checks for RD analysis are presented below.

Fig. 1 shows that the density of the assignment variable is relatively smooth, and importantly, does not exhibit any sharp changes near the cut-off value. The score of 'zero' is the sharp cut-off point that demarcates levels II and III (100% of test scores at or below zero are assigned in the data as level I or II, and 100% of test scores above zero are assigned in the data as level III or IV).

Fig. 2 plots the average math test score in 8th grade at each test score, showing that there are no sharp changes in last year's standardized test outcome (or student grades) around the cut-off point that could be driving the observed treatment impact. Similar plots for average school percent of females, minorities, or average attrition rate also show no discontinuity near the cut-off.[23] This lends weight to the assertion that the discontinuity in math course selection is the result of the policy itself and not due to sharp differences in student demographic characteristics or data concerns for those who just fail and those who just pass the exam.

4. Results

Regression discontinuity analysis can be performed either parametrically or non-parametrically. This study uses the non-parametric specification from Hahn, Todd, and van der Klaauw (2001). I estimate a local linear regressions to fit a smooth function to either side of the cut-off. The shape of the function is heavily influenced by the size of the bandwidth in the local linear regression. As such, I reports results for a variety of bandwidths.[24] As the probability of receiving the 'treatment' (that is, receiving the pass/fail signal based on exam results) is one/zero conditional the student's standardized score being above/below zero, a sharp regression discontinuity framework is appropriate. The 'optimal' bandwidth is calculated following the procedure proposed by Imbens and Kalyanaraman (2010).

Fig. 3 presents a graphical representation of the basic RD estimates, and Table 2 shows the impact and standard errors.[25] RD parameter estimates indicate that students who just failed the Algebra I EOC exam sort into less rigorous math courses compared to students who just passed the exam.[26] While all levels of math curriculum rigor were considered, the largest discontinuity was seen for students choosing between 'Career Prep' (no math course more advanced than Algebra I including not taking any math) and all post-secondary preparation math courses.[27] The results are fairly robust to bandwidth choices and specifications. Compared to students who just passed the Algebra I EOC exam, those who just failed are

[22] For 10th graders who took the Algebra I exam for the first time, the trend is similar, although a smaller fraction of students had post-secondary plans.

[23] See online appendix for figures.

[24] See Imbens and Lemieux (2008) for details.

[25] The preferred RD specification (labeled (5)) contains the following covariates: student minority status and gender, teacher minority status and gender, first year teacher indicator, and school dummy variables. Errors are clustered at the school level. See the tables for robustness checks for additional covariates including quadratic and cubic controls for normalized test scores.

[26] Although it is not a focus of this paper, a negative RD estimate would have been consistent with the story that just failing the exam may serve as an extra motivator for students to try harder to achieve a better outcome in the next period.

[27] That the largest RD estimates would be at the College Technical Prep track or above is consistent with the students on the border of passing and failing the Algebra I EOC exam being between the 10th and 40th percentile of the exam score distribution.

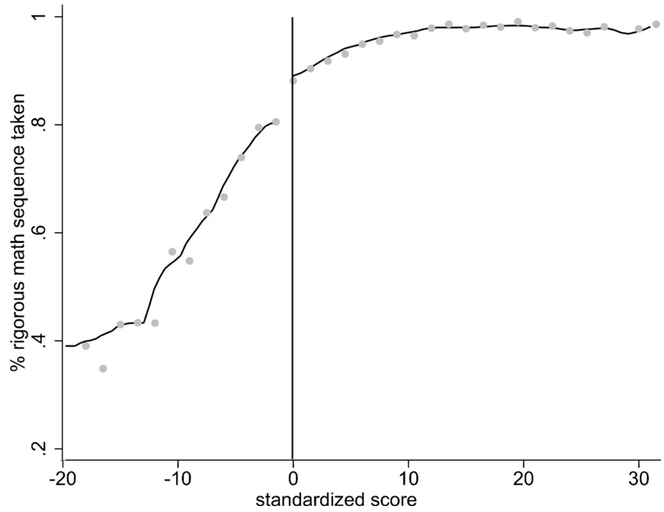

Fig. 3. 2006 sample RD. Figure is generated with local polynomial of degree zero. Covariates are excluded. Observations for standardized scores less than −20 or greater than 30 (which compose approx. 5.7% of observations) are dropped for presentation purposes.

Table 2
Regression discontinuity results: 2006 sample.

Bandwidth	(1)	(2)	(3)	(4)	(5)
Career Prep vs. Technical Prep					
1.9777	0.0561 (0.0159)***	0.0791 (0.0171)***	0.0449 (0.0199)**	0.0564 (0.0159)***	0.0519 (0.0145)***
0.9889	0.0559 (0.0159)***	0.0395 (0.0373)	0.0027 (0.0493)	0.0567 (0.0159)***	0.0598 (0.0139)***
3.9554	0.0427 (0.0227)*	0.0508 (0.0201)**	0.0574 (0.0280)**	0.0430 (0.0227)*	0.0471 (0.0190)**
6.9777	0.0338 (0.0154)**	0.0392 (0.0157)**	0.0619 (0.0239)***	0.0341 (0.0154)**	0.0348 (0.0185)*
Technical Prep vs. University Prep					
2.2106	0.0122 (0.0408)	0.0092 (0.0522)	0.0508 (0.0266)*	0.0110 (0.0405)	0.0374 (0.0377)
1.1053	0.0632 (0.0214)***	0.0738 (0.0242)***	−0.0084 (0.0632)	0.0647 (0.0213)***	0.0823 (0.0229)***
4.4212	0.0284 (0.0261)	0.0070 (0.0283)	0.0038 (0.0456)	0.0290 (0.0260)	0.0433 (0.0292)
7.2106	0.0199 (0.0195)	0.0157 (0.0191)	0.0168 (0.0282)	0.0216 (0.0194)	0.0263 (0.0235)

Specification (1) has linear normalized score controls. Specification (2) controls for normalized score squared. Specification (3) controls for normalized score cubed. Specification (4) controls for student ethnicity and gender and teacher ethnicity, gender, and experience, and school fixed-effects. Specification (5) controls for all covariates in Specification (4) and adds clustered errors at the school level. Specifications (4) and (5) have linear normalized controls. Specification (5) is the preferred specification. Bandwidths are as follows: 1st row (optimal bandwidth), 2nd row (half of optimal bandwidth), 3rd row (twice optimal bandwidth), and 4th row (optimal bandwidth + 5). Standard errors are in parentheses.

* An estimate significant at the 10% level.

** An estimate significant at the 5% level.

*** An estimate significant at the 1% level.

about 5 percentage points less likely to choose a rigorous math sequence.

RD results using the 2005 sample in Table 3 show that the discontinuity did not exist prior to the implementation of the graduation standard.[28] Although students were taking the exam, and the results were known to the school and parents (as it was in 2006), students were not sharply divided into different math tracks based on the pass/fail results. A plot of the RD in Fig. 4 shows a smooth increasing function in the scale score, indicating that students were not pushed into different math sequences at the point between level II and level III. This provides some evidence that the discontinuity observed in 2006 is a real reaction to the graduation standard implementation.

RD results from the years surrounding the two comparison years provide supporting evidence that the 2006 result is not an anomaly in the data. Fig. 4 plots the RD results for 2004 and 2005. RD results from 2004 closely follow those of 2005, showing no discontinuity at the cut-off value. Fig. 5 plots RD results from 2006, 2007, and 2008.[29] RD results from the two subsequent years continue to exhibit discontinuity at the border between pass/fail, confirming that the treatment is felt beyond 2006.

[28] Fig. 3 shows relatively steep drops at 3 other standardized scores below zero, and it is unclear where these 'discontinuities' are from. The lack of a treatment effect at the cut-off in Fig. 4 does show that the discontinuity observed at score zero in 2006 is indeed due to the high stakes exam.

[29] Scatter plots (of the conditional probability) make the graph difficult to read and are omitted.

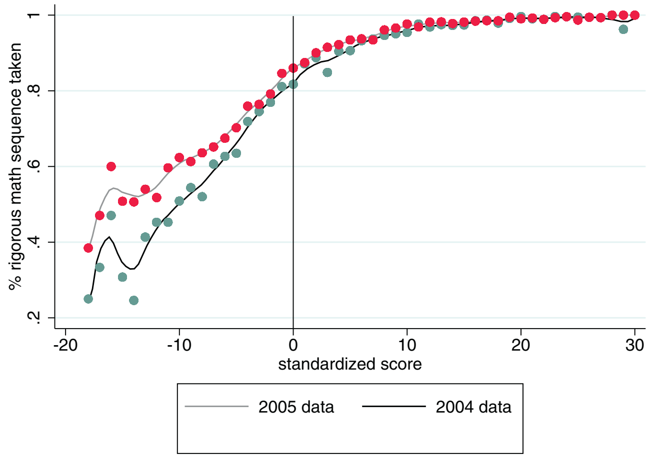

Fig. 4. Pre-graduation standard years. Figure is generated with local polynomial of degree zero. Covariates are excluded. Observations for standardized scores less than −20 or greater than 30 (which compose approx. 5.7% of observations) are dropped for presentation purposes.

Table 3
Regression discontinuity results: 2005 sample.

Bandwidth	(1)	(2)	(3)	(4)	(5)
Career Prep vs. Technical Prep					
2.3315	−0.0149 (0.0205)	0.0203 (0.0189)	0.0050 (0.0267)	−0.0151 (0.0205)	−0.0046 (0.0195)
1.1658	0.0127 (0.0172)	−0.0093 (0.0189)	−0.0205 (0.0455)	0.0129 (0.0172)	0.0149 (0.0200)
4.6630	−0.0442 (0.0338)	−0.0336 (0.0276)	−0.0291 (0.0302)	−0.0439 (0.0336)	−0.0473 (0.0333)
7.3315	−0.0176 (0.0170)	−0.0123 (0.0172)	−0.0274 (0.0200)	−0.0173 (0.0169)	−0.0052 (0.0220)
Technical Prep vs. University Prep					
2.3999	0.0445 (0.0471)	0.0146 (0.0593)	0.0470 (0.0308)	0.0474 (0.0465)	0.0412 (0.0258)
1.2000	0.0419 (0.0248)*	0.0606 (0.0274)**	0.0233 (0.0389)	0.0460 (0.0244)*	0.0367 (0.0466)
4.7998	−0.0050 (0.0289)	0.0019 (0.0310)	0.0010 (0.0354)	−0.0034 (0.0285)	−0.0224 (0.0281)
7.3999	−0.0124 (0.0223)	−0.0056 (0.0231)	−0.0182 (0.0272)	−0.0100 (0.0220)	−0.0104 (0.0184)

Specification (1) has linear normalized score controls. Specification (2) controls for normalized score squared. Specification (3) controls for normalized score cubed. Specification (4) controls for student ethnicity and gender and teacher ethnicity, gender, and experience, and school fixed-effects. Specification (5) controls for all covariates in Specification (4) and adds clustered errors at the school level. Specifications (4) and (5) have linear normalized controls. Specification (5) is the preferred specification. Bandwidths are as follows: 1st row (optimal bandwidth), 2nd row (half of optimal bandwidth), 3rd row (twice optimal bandwidth), and 4th row (optimal bandwidth + 5). Standard errors are in parentheses.

* An estimate significant at the 10% level.

** An estimate significant at the 5% level.

***An estimate significant at the 1% level.

This sizable difference in the propensity of students to select into more difficult math courses may be due to one (or some combination) of three possible mechanisms.

One possible reason for the discontinuity is that there may be a tracking system pushing students into the career, community college, or university preparatory curricula. The state and schools may feel that having 'sufficient mastery' of Algebra I is simply a prerequisite for college-preparatory work, and those who demonstrate proficiency by attaining level III or above on the EOC exam (as well as satisfying other criteria) are mechanically sorted into the more rigorous math tracks. If the complete formula for placing students into different tracks was known, the data was available, and sorting was purely mechanical, an assignment score could be constructed that would generate close to a 100 percentage points difference at the point of treatment.[30] Thus, the observed gap may be an imperfect representation of the statewide tracking policy for math courses.

Another possibility is that the exam results serve as an easy to interpret signal of the student's readiness for tougher math classes to the student, parents, and teacher. The pass or fail outcome may be perceived by students who do not have good information on their academic potential as an important indicator of their ability to succeed in a

[30] The gap may not be perfect because some students may insist on taking a different course sequence regardless of their performance on the exam and other criteria.

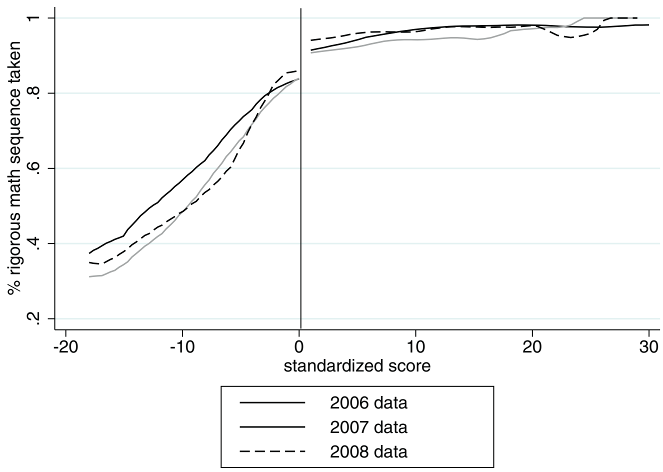

Fig. 5. Post-graduation standard years. Figure is generated with local polynomial of degree zero. Covariates are excluded. Observations for standardized scores less than −20 or greater than 30 (which compose approx. 5.7% of observations) are dropped for presentation purposes.

Table 4
RD results: 2006 sample with plans to pursue 2-year college or no college.

Bandwidth	(1)	(2)	(3)	(4)	(5)
Career Prep vs. Technical Prep					
2.1294	0.0646 (0.0344)*	0.0680 (0.0373)*	0.0384 (0.0601)	0.0653 (0.0343)*	0.0528 (0.0323)*
1.0647	0.0320 (0.0528)	0.0807 (0.0636)	0.0611 (0.0336)*	0.0309 (0.0528)	0.0248 (0.0432)
4.2588	0.0621 (0.0267)**	0.0623 (0.0277)**	0.0723 (0.0411)*	0.0620 (0.0266)**	0.0432 (0.0289)
7.1294	0.0602 (0.0257)**	0.0607 (0.0265)**	0.0732 (0.0399)*	0.0602 (0.0256)**	0.0417 (0.0259)*
Technical Prep vs. University Prep					
1.8894	0.0100 (0.0452)	−0.0257 (0.0542)	0.0649 (0.0521)	0.0123 (0.0446)	0.0290 (0.0442)
0.9447	0.0495 (0.0329)	0.0341 (0.0382)	−0.1070 (0.0936)	0.0503 (0.0325)	0.0917 (0.0440)**
3.7788	0.0196 (0.0340)	0.0027 (0.0360)	−0.0020 (0.0548)	0.0208 (0.0336)	0.0386 (0.0327)
6.8894	0.0188 (0.0323)	0.0036 (0.0340)	0.0044 (0.0499)	0.0201 (0.0320)	0.0371 (0.0319)

Specification (1) has linear normalized score controls. Specification (2) controls for normalized score squared. Specification (3) controls for normalized score cubed. Specification (4) controls for student ethnicity and gender and teacher ethnicity, gender, and experience, and school fixed-effects. Specification (5) controls for all covariates in Specification (4) and adds clustered errors at the school level. Specifications (4) and (5) have linear normalized controls. Specification (5) is the preferred specification. Bandwidths are as follows: 1st row (optimal bandwidth), 2nd row (half of optimal bandwidth), 3rd row (twice optimal bandwidth), and 4th row (optimal bandwidth + 5). Standard errors are in parentheses.

* An estimate significant at the 10% level.
** An estimate significant at the 5% level.
***An estimate significant at the 1% level.

academically rigorous environment. Alternatively, students who are truly marginal between taking tougher math courses and opting for classes more appropriate for those directly entering the labor force may be tipped toward the higher or lower track based on the outcome. If teachers are guiding students, the dichotomous pass/fail outcome and the emphasis placed on this result by the state may make math sequence recommendation for the marginal student easier.

In an alternative RD analysis focusing on the sample of students with modest academic goals, who (prior to taking the Algebra I exam) indicated a desire to go to a 2-year college or pursue some type of employment out of high school, we see that just failing the exam will make a student about 5–7 percentage points more likely pursue a 'Career Prep' math sequence, compared to a student who just passed. See Table 4. This gap, which is similar in magnitude to the discontinuity estimated with the full sample, would seem to point toward a more mechanistic sorting of students into tracks. At the same time, the lack of a discontinuity in the pre-graduation standard year (2005) sample seems to indicate that passing or failing the EOC exam did not serve as a strong sorting mechanism in the previous year.

To explore further, a RD analysis was done with 10th grade students in 2006. These students are not subject to the graduation requirement, but are the first non-9th grade cohort to be exposed to the policy in their schools.[31] They are grandfathered in to the old regime. As a consequence, they would not experience any mechanical changes in how they are pushed into math classes depending on their EOC achievement level, unless the tracking system was adopted for 10th grade students as well. However, if the graduation requirement policy was known and implemented in the

[31] The 10th grade sample of students is restricted to those who take the Algebra I EOC exam for the *first* time in 10th grade.

Table 5
Regression discontinuity results: 10th grade results.

Bandwidth	(1)	(2)	(3)	(4)	(5)
2005 sample					
2.3315	−0.0048 (0.0652)	−0.0017 (0.0843)	0.0108 (0.0431)	0.0030 (0.0638)	0.0265 (0.0602)
1.1658	0.0166 (0.0348)	0.0333 (0.0392)	−0.0001 (0.0557)	0.0229 (0.0345)	0.0320 (0.0241)
4.6630	0.0019 (0.0401)	−0.0146 (0.0441)	−0.0343 (0.0711)	−0.0003 (0.0391)	0.0272 (0.0374)
7.3315	0.0093 (0.0300)	0.0189 (0.0313)	0.0012 (0.0445)	0.0082 (0.0295)	0.0273 (0.0330)
2006 sample					
2.0768	0.0606 (0.0389)	0.0749 (0.0436)*	0.0706 (0.0670)	0.0578 (0.0386)	0.0594 (0.0549)
1.0384	0.0605 (0.0430)	0.0847 (0.0492)*	0.0703 (0.0886)	0.0579 (0.0426)	0.0448 (0.0335)
4.1536	0.0526 (0.0328)	0.0637 (0.0353)*	0.0881 (0.0550)	0.0501 (0.0326)	0.0582 (0.0378)
7.0768	0.0286 (0.0285)	0.0457 (0.0297)	0.0901 (0.0440)**	0.0272 (0.0283)	0.0132 (0.0378)

Specification (1) has linear normalized score controls. Specification (2) controls for normalized score squared. Specification (3) controls for normalized score cubed. Specification (4) controls for student ethnicity and gender and teacher ethnicity, gender, and experience, and school fixed-effects. Specification (5) controls for all covariates in Specification (4) and adds clustered errors at the school level. Specifications (4) and (5) have linear normalized controls. Specification (5) is the preferred specification. Bandwidths are as follows: 1st row (optimal bandwidth), 2nd row (half of optimal bandwidth), 3rd row (twice optimal bandwidth), and 4th row (optimal bandwidth + 5). Standard errors are in parentheses.

* An estimate significant at the 10% level.

** An estimate significant at the 5% level.

***An estimate significant at the 1% level.

school, these 10th grade students (or their teachers) would realize that a bad performance in Algebra I EOC exam is viewed by the state as an indicator that they are ill-prepared for more rigorous math courses.

RD results for 2006 10th graders in the bottom portion of Table 5 show that while treatment impact is weak, many of the estimates of the discontinuity are barely insignificant. In specification (1), the discontinuities in 3 of the 4 bandwidths are 'statistically significant at the 11–16% level.' Specification (2) has 3 of 4 discontinuities significant at the 10% level, and specification (3) has 1 of 4 discontinuities significant at the 5% level. In specifications (4) and (5), the discontinuities are 'statistically significant at the 13–18% level' for 3 of the 4 bandwidths. In contrast, the 10th grade results from 2005 in the top portion of Table 5 clearly show that there is no discontinuity. Together, I interpret the results as suggestive of 10th graders in 2006 also responding to the signal sent from failing or passing the Algebra I exam. A 10th grade student was about 5 percentage points less likely to select into a more rigorous math track if he or she just fails the exam.[32] The 'noise' that leads to weaker discontinuity results may be interpreted as many 10th graders and their teachers responding more weakly in response to the signal, since they have had an additional year of high school course work and academic advising.

On balance, whether formal tracking exists or not, the similar results across 9th and 10th grade samples seem to indicate that schools (and teachers or counselors) exercise some influence over which math track students select. At the same time, the weaker results for 10th graders suggest that students have input and are not mechanically placed in tracks.

Finally, another possible source of the discontinuity is the possibility of strategic test taking. If a large number of students delayed taking college-preparatory courses after failing their initial Algebra I exam to prepare for the re-take, this may drive the observed discontinuity. There is some reason to expect that the effect would not be large. As the re-take occurs within three weeks of the first exam results (which must be distributed before the end of the semester as part of the student's final course grade), a student will not be able to use the math course in the next semester for review and preparation. As seen in Table 1, the vast majority (over 95%) of students near the threshold will take the exam once or twice, indicating that whatever effort was being made (by the student and teachers) in the three weeks, it is usually enough to get the marginal student over the bar. This form of strategic course-selection would be more beneficial for students who are further below the bar, and expectations are that they will require more than two attempts to pass the exam. These students will not be part of the relevant population for regression discontinuity analysis.

4.1. Do high-stakes exams increase or decrease the number of students who select the rigorous math sequence?

An issue in interpreting the RD results is that it is unclear whether the treatment is pushing the failed students down, pulling the passed students up, or some combination of both. Researchers were left to make arguments in support of one or the other as the dominant reason for the discontinuity.[33]

[32] Of course, one other possibility is that graduation requirements were instituted only for 9th graders, but math tracking was instituted for both 9th and 10th graders. I have been unable to find any documentation (or corroboration from talking with school administrators) of explicitly forcing 10th graders into a math track as a result of their EOC exam results.

[33] Many RD studies fall under this category. Treatments where assignment is not random, but the result of competition or tournament may have this feature, such as any study with a discontinuity of failing an exam, failing to pass the bar in accountability studies, or election results. See Jacob and Lefgren (2004), Matsudaira (2008), and Lee (2008), among others.

Table 6
Division of discontinuity: 9th grade.

Bandwidth	(1)	(2)	(3)	(4)	(5)
τ_{pass}					
1.5343	0.0283 (0.0268)	0.0201 (0.0159)	0.0201 (0.0159)	−0.0053 (0.0279)	0.0227 (0.0200)
0.7672	−0.0027 (0.0146)	0.0464 (0.0334)	0.0464 (0.0334)	−0.0340 (0.0180)*	−0.0057 (0.0118)
3.0686	0.0259 (0.0245)	0.0374 (0.0287)	0.0409 (0.0304)	−0.0048 (0.0247)	0.0192 (0.0186)
6.5343	0.0002 (0.0122)	0.0067 (0.0131)	0.0076 (0.0136)	−0.0146 (0.0122)	−0.0043 (0.0095)
τ_{fail}					
2.0579	0.0411 (0.0220)*	0.0587 (0.0446)	0.0587 (0.0446)	0.0546 (0.0227)**	0.0531 (0.0272)*
1.0290	0.0513 (0.0167)***	0.0385 (0.0270)	0.0524 (0.0292)*	0.0645 (0.0221)***	0.0630 (0.0212)***
4.1158	0.0357 (0.0173)**	0.0483 (0.0203)**	0.0394 (0.0218)*	0.0518 (0.0179)***	0.0493 (0.0216)**
7.0579	0.0317 (0.0166)*	0.0480 (0.0192)**	0.0388 (0.0207)*	0.0493 (0.0172)***	0.0456 (0.0204)**

Specification (1) has linear normalized score controls. Specification (2) controls for normalized score squared. Specification (3) controls for normalized score cubed. Specification (4) controls for student ethnicity and gender and teacher ethnicity, gender, and experience, and school fixed-effects. Specification (5) controls for all covariates in Specification (4) and adds clustered errors at the school level. Specifications (4) and (5) have linear normalized controls. Specification (5) is the preferred specification. Bandwidths are as follows: 1st row (optimal bandwidth), 2nd row (half of optimal bandwidth), 3rd row (twice optimal bandwidth), and 4th row (optimal bandwidth + 5). Standard errors are in parentheses.

* An estimate significant at the 10% level.

** An estimate significant at the 5% level.

*** An estimate significant at the 1% level.

Using 2005 RD results, it is possible to assign portions of the 2006 discontinuity between the response to failing and passing. Intuitively, the intercept at the cut-off *in* 2005 is the probability of pursuing more rigorous coursework conditional on scoring 'zero.' Then, the difference between the right side of the discontinuity *in* 2006 and the intercept in 2005 is the increase in the likelihood of pursuing more rigorous coursework due to the signal from achieving level III. The difference between the left side of the discontinuity and the intercept is the decrease in the likelihood of taking more difficult classes due to the signal from failing to achieve level III.[34]

As Table 6 shows, almost all of the discontinuity is generated by the signal of failing (τ_{fail}). I remain agnostic about why the signal from passing does not result in more students opting into the tougher math sequence. It may be because most students were already intending to take college preparatory math and 'knew' they would pass. Those who fail may be shocked into reevaluating academic plans. Alternatively, the school may assign marginal students to the rigorous math course (across 2005 and 2006) by default, and the new standard pushes some students who fail to achieve level III into the career preparatory track. It should be noted that very strong assumptions about temporal stability across 2005 and 2006 are required to reinterpret the discontinuity. This is briefly addressed below.

4.2. Temporal stability across 2005 and 2006

An important assumption for the interpretation of the division of discontinuity in this way is that the population and exam must be stable across 2005 and 2006. The samples in Table 1 are similar. The main difference is that the average score in 2006 is slightly lower, allaying some concerns that students are delaying taking the exam (say, in the hope that the standards would be discontinued).[35] The average scores in the exams differ by one point, which could be indicative of the exam becoming more difficult. The exam is 80 multiple choice questions, and a one point difference is equivalent to getting one fewer question correct. Historic score averages from 2001 to 2006 suggest that the mean does move around, which may be attributable to random variance in exam difficulty.[36]

It is also critical that the intercept at the 'zero' point be stable across years prior to the graduation standard. Even if prior years showed no discontinuity, if the conditional probability at the 'zero' point varies wildly across years, it would be difficult to justify assigning portions of the discontinuity using 2005 intercept as the base. Similarly, the discontinuity should be stable after the imposition of the standard. Even if discontinuity exists in several years, if the magnitude and location of the discontinuity move around, it would be difficult to justify using the 2006 probabilities in the calculations. As discussed previously, Fig. 4 graphs results for 2004 and 2005. Beyond confirming the lack of discontinuity, the figure shows that the conditional probabilities across the two years are relatively close together.[37] In Fig. 5, the

[34] See online appendix.

[35] There are differences in the highest score/level and courses taken afterward, which is the impact of the treatment.

[36] I have been unable to find documentation from the state that would indicate that the exam was deliberately made more difficult in 2006. To deliberately change the difficulty of a test by only about 1/8 of a standard deviation seems strange, but it is possible that the exam was inadvertently made more difficult. Yearly densities of the test score are relatively stable. See online appendix.

[37] The difference in the 'intercepts' at the cut-off point between the two years is approximately 0.015. Since the discontinuity estimated in the preferred specification is approximately 0.052, this is approximately 29% of the size of the discontinuity. Ideally, I would want more than one additional year to check stability. However, the transcript file, which is necessary to capture the math course that students take, does not exist prior to 2005.

three post-graduation standard years line up near the cut-off value. However, the divergence in the conditional probabilities for both the pre- and post-graduation standard years, especially toward the left tail, means that the interpretation is not valid everywhere along the range of scores. While the relatively stable pre- and post-graduation standard years estimates at the cut-off lends some support to dividing the discontinuity this way, the lack of additional corroborating years in the pre-graduation standard figure means the test is ultimately inconclusive. Caution should be exercised in interpreting the results too broadly.

5. Conclusion

Having a simple thumbs up or thumbs down result on a high-stakes exam at an early point in a student's high school career will make it easier for students, teachers, and counselors to plot out an academic trajectory for a student. However, this simplistic approach to student evaluation (and tying the results to a substantive high-stakes outcome, such as graduation) will invariably send too coarse of a signal to some students and their teachers.

North Carolina briefly used high-stakes exams as part of its graduation standard. 9th grade students who took the Algebra I EOC exam and just failed are 5 percentage points less likely to pursue a rigorous mathematics curriculum that would prepare them for post-secondary education, compared to those who just passed. Results suggest that more students moved into the easier math track as a result of the imposition of the graduation standard. Whether this impacts a student's self-perception or a school's expectation of his or her ability, the observed response is that students are steered away from rigorous coursework designated by the state as necessary for college. In contrast, passing the Algebra I exam (for 9th graders) does not seem to increase the likelihood of these students pursuing a college-preparatory math curriculum.

Whether this is harmful or beneficial for the average student affected is unclear. If the average student just at the point of passing or failing the exam *should* prepare for college and/or be challenged intellectually, the exam is clearly harmful. If, on the other hand, these students have a sufficiently low probability of succeeding in these tougher courses, pushing a random number of students away from tougher math courses may yield some benefits (such as increasing on-time graduation). In any case, from the analysis above, it is clear that a more nuanced understanding of the impact of high-stakes exams is required.

References

Ahn, T. (2013). The missing link: Estimating the impact of incentives on effort and effort on production using teacher accountability legislation. *Journal of Human Capital, 7*(3), 230–273.

Bishop, J. H., & Mane, F. (2001). The impacts of minimum competency exam graduation requirements on high school graduation, college attendance and early labor market success. *Labour Economics, 8*, 203–222.

Carnoy, M., Loeb, S., & Smith, T. (2001). *Do higher state test scores in Texas make for better high school outcomes? CPRE Research Report RR-047* Philadelphia, PA: Consortium for Policy Research in Education.

Clark, D., & See, E. (2011). The impact of tougher education standards: Evidence from Florida. *Economics of Education Review, 30*, 1123–1135.

Dee, T. S. (2003). The "first wave" of accountability in No Child Left Behind? In P. E. Peterson & M. R. West (Eds.), *No Child Left Behind? The politics and practice of school accountability* (pp. 215–241). Washington, DC: Brookings Institution.

Dee, T. S., & Jacob, B. A. (2006). *Do high school exit exams influence educational attainment or labor market performance? Working Paper 12199*.

Grodsky, E. S., Warren, J. R., & Kalogrides, D. (2009). State high school exit examinations and NAEP long-term trends in reading and mathematics, 1971–2004. *Educational Policy, 23*, 589–614.

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.

Holme, J., Richards, M., Jimerson, J., & Cohen, R. (2010). Assessing the effects of high school exit examinations. *Review of Educational Research, 80*, 476–526.

Imbens, G., & Kalyanaraman, K. (2010). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies, 79*, 933–959.

Imbens, G., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics, 142*(2), 615–635.

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

Jacob, B., & Lefgren, L. (2004). Remedial education and student achievement: A regression discontinuity analysis. *Review of Economics and Statistics, 86*(1), 226–244.

Jurges, H., Schneider, K., Senkbeil, M., & Carstensen, C. (2012). Assessment drives learning: The effects of central exit exams on curricular knowledge and mathematical literacy. *Economics of Education Review, 31*, 56–65.

Larson, M. (2011). *High bars or behind bars? The effect of graduation requirements on arrest records. Working Paper*.

Lee, D. (2008). Randomized experiments from non-random selection in the U.S. House elections. *Journal of Econometrics, 142*(2), 675–697.

Martorell, F. (2004). *Do high school graduation exams matter? A regression discontinuity approach. Working Paper*.

Matsudaira, J. (2008). Mandatory summer school and student achievement. *Journal of Econometrics, 142*(2), 829–850.

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*, 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*, 5–23.

Papay, J. P., Murnane, R. J., & Willett, J. B. (2011). *High-school exit examinations and the schooling decisions of teenagers: A multidimensional regression discontinuity analysis. NBER Working Paper No. 17112*.

Reardon, S. F., Atteberry, A., Arshan, N., & Kurlaender, M. (2009). *Effects of the California high school exit exam on student persistence, achievement, and graduation. Working Paper*.

Warren, J. R., & Edwards, M. R. (2005). High school exit examinations and high school completion: Evidence from the early 1990. *Educational Evaluation and Policy Analysis, 27*, 53–74.

Woessmann, L., Ludemann, E., Schutz, G., & West, M. R. (2007). *School accountability, autonomy, choice, and the level of student achievement: International evidence from PISA 2003. OECD Education Working Paper 13*.