Economics of Education Review 30 (2011) 1123–1135

Contents lists available at ScienceDirect

Economics of Education Review

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

The impact of tougher education standards: Evidence from Florida

Damon Clark[a,*], Edward See[b]

[a] University of Florida, NBER, IZA, United States
[b] University of Florida, United States

ARTICLE INFO

Article history:
Received 24 December 2010
Received in revised form 11 May 2011
Accepted 13 May 2011

JEL classification:
I21
I28
J24

Keywords:
Economic impact
Human capital
Productivity
Rate of return

ABSTRACT

Many of the policies that fall under the school accountability umbrella are designed to incentivize students. Prominent among these are high school exit exams, standardized tests that, in some states, students must pass to earn a high school diploma. Proponents of these tests argue that by incentivizing students, they induce them to work harder and, therefore, improve their high school performance and, perhaps, longer-run outcomes; some of these proponents argue that these exams would be even more helpful if they were set at a higher standard. Critics worry that these exams prevent some students from graduating and cause others to dropout; they contend that these effects are worse if standards are higher. In this paper we investigate the impacts of an increase in the exit exam standard in Florida. Using difference-in-difference methods, we show that this had few of the negative effects claimed by critics. We cannot detect any positive effects of the higher standard, although such effects may be too small to be picked up with our data.

1. Introduction

The "standards and accountability" movement in US education is often characterized as a set of policies designed to improve school effectiveness. But several policies that fall under the standards and accountability umbrella are designed to incentivize students. These policies, which include test-based promotion and high school exit exams, differ in detail.[1] The basic idea is, however, the same: standards are set, tests are used to measure students' progress against those standards and high-stakes consequences are tied to the outcome of those tests. In the case of high school exit exams, the focus of this paper, the standard is defined by particular thresholds on math and reading tests given to students in tenth grade. Students cannot graduate high school if they do not pass these tests.

There are several issues surrounding the practical implementation of high school exit exams. Which aspects of the curriculum should be covered by these tests (e.g., calculus or only algebra)? What format should the tests take (e.g., short answer or multiple choice)? Should any students be exempted (e.g., limited English proficient students)? How many chances should students have to retake the test if they fail first time (e.g., five, as in Florida, or more, as in some other states)? Where should the standard (i.e., passing score) be set?

In this paper we focus on this last question and consider whether students should be subjected to higher standards. Policy-makers are often perceived as setting standards that are too low. Moreover, economic models suggest this perception may be correct.[2] Yet these models are static, in

* Corresponding author.
E-mail address: damonc@princeton.edu (D. Clark).

[1] Greene and Winters (2009) discuss test-based promotion in Florida.

[2] In a formal model of standards setting, Costrell (1994) argues that policy-makers may set standards below those that would be preferred by a median voter. This is because egalitarian-minded policy-makers care more about lower-ability students than does the median voter. Betts (1998) argues that even policy-makers that care about low-ability students should set higher standards. Using a different theoretical set-up, he shows that higher standards make lower- and higher-ability students better off, with only middle-ability students losing out. The intuition is that higher standards reduce the fraction of students that pass, thereby

0272-7757/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.econedurev.2011.05.010

the sense that they consider a set of students that face a particular standard. In practice, students can drop out of education before they have to face the standard. For example, high school students can drop out in tenth or eleventh grade, before any exit exam standard must be met (the end of twelfth grade). As such, higher standards might increase drop out rates, perhaps generating a net human capital loss.

In this paper we aim to shed light on this question by examining the effects of the higher graduation standards introduced in Florida in 2001. These increased the scores required on both the math and reading portions of the high school exit exam such that 19 percent of the students that passed under the old lower standard would have failed under the new higher standard. We estimate these effects using a difference-in-difference approach that classifies students according to whether they reached tenth grade in the year before or the year after the higher standard was introduced (labeled "pre" and "post") and according to whether their initial exam score would pass under the lower and the higher standard. Assuming that the higher standard only affects behavior after the initial exam attempt (in tenth grade) and assuming that the higher standard did not affect students whose score would have passed under both regimes, pre-post comparisons of the students that would have passed under the higher standard can identify the effects of the higher standard.[3] We examine effects on dropout rates, graduation rates, postsecondary enrollment rates and earnings.

Our estimates point to two main results. First, the higher standard had no impact on dropout rates. Instead, the fraction of students that completed grade twelve appears unaffected. This is surprising, since one might expect students that failed the initial exam to become discouraged and dropout of school. Instead, it seems that the higher standard caused a higher fraction of students to complete grade twelve without passing the exam. The net result of these effects was a very small drop in graduation rates. Second, the higher standard had no apparent impact on postsecondary enrollment rates and earnings. The postsecondary enrollment effects are unsurprising considering that students do not need a high school diploma to enroll. The earnings effects are consistent with small standards effects on graduation rates and small earnings effects of graduation (Martorell & Clark, 2010).

2. High school standards in Florida

Florida was the first state to make high school graduation contingent on passing an exit exam (in 1978).[4] This exam was known as the High School Competency Test (HSCT). In 2000, by which time Florida had developed an accountability system based on FCAT tests in grades three through ten, the HSCT was replaced by the grade ten FCAT.

The grade ten FCAT exams are offered in reading and mathematics.[5] In general, all students enrolled in grade ten should take these exams.[6] The exams contain multiple choice items, gridded-response items and performance tasks.[7] Raw scores are scaled using item response theory methods. The scaled scores range from 100 to 500 and cover five pre-defined levels of achievement.[8] To pass the exit exam, students need to perform at the second level in both exams. For students in tenth grade in 2000–2001 (i.e., took the exam for the first time in spring 2001), the passing scores were 287 for reading and 295 for math. For students in tenth grade from 2001 to 2002 onwards (i.e., took the exam for the first time from spring 2002 onwards), the passing scores were 300 in both subjects. Students and parents receive a four-page report describing their exam performance. This lists the scores obtained, whether the student passed, the associated performance levels and the scores required to obtain the various performance levels and to pass.

Students that fail one or both of these exams have numerous opportunities to retake them. Students need only retake the exams they have not yet passed. Initially, these retakes took the same format as the initial exam. Starting fall 2004, the retake format changed to include only multiple choice and gridded response questions.[9] Students in tenth grade in 2000–2001 (i.e., took the first exam in spring 2001) had five chances to retake the exam: the fall of eleventh grade (October 2001), the spring of eleventh grade (February–March 2002), the summer of eleventh grade (June 2002), the fall of twelfth grade (October 2002) and the spring of twelfth grade (February–March 2003). Students in tenth grade in 2001–2002 onwards had an additional retake opportunity in the summer of grade 10. Since few students retake the exam in the summer of tenth and eleventh grades, students typically have five chances to pass before the end of grade twelve. School districts

improving the quality of both the group that fails and the group that passes the test. The losers are those that would have passed under the low standard but do not pass under the high standard.

[3] Without these additional assumptions, which we justify below, the impacts of the higher standard would have to be identified using an aggregate pre-post comparison. Since this type of estimate would be highly sensitive to the specification of cohort trends, we think that our approach is more appealing.

[4] This section draws heavily on several Florida Department of Education publications. These include Florida Department of Education (2005a,b).

[5] The reading exam is a 160-min exam that assesses students' reading comprehension. The test is composed of about 6–8 reading passages (informational or literary) with sets of 6–11 questions based on each passage. The math exam is a 160-min exam that assesses performance on five strands: (1) Number Sense, Concepts and Operations; (2) Measurement; (3) Geometry and Spatial Sense; (4) Algebraic Thinking and (5) Data Analysis and Probability. Students are allowed to use calculators.

[6] LEPt students can be exempted if they have received services in an LEP program for one year or less. An Exceptional Student Education (ESE) student may be exempted if he has an Individual Education Plan (IEP) and meets certain criteria.

[7] In gridded-response items, students answer questions that require a numerical response and they mark their numerical answers in response grids.

[8] These are "little success with the content on the FCAT" (level 1), "limited success with the content on the FCAT" (level 2), "partial success with the content on the FCAT"(level 3), "success with the content on the FCAT by answering most questions correctly, except for the challenging questions" (level 4), "success with the content of the FCAT by answering most questions correctly" (level 5).

[9] This change was designed to reduce grading costs. The DOE claimed that the new format was aligned with the old standards. Students will receive this if their IEP team determines that the FCAT accurately reflects the student's abilities.

are required to give “intensive remediation” to seniors that have not yet received a passing score on the FCAT. In practice, schools offer remediation to students before they reach their senior year. Students are given performance reports after each retake.

To obtain a public high school diploma, students must meet the exam requirement, maintain a 2.0 GPA and earn course credits in the required number and distribution. Some students that have met the other graduation requirements but not passed the FCAT can be exempted from the FCAT and can receive a diploma. Exemptions are provided for students with disabilities if the student’s IEP team determine in the student’s senior year that the FCAT does not accurately reflect the student’s abilities. Starting May 2004, students with ACT/SAT scores above certain thresholds are also exempt from the FCAT requirement. Students that have met the other graduation requirements but not passed the FCAT can receive a “Certificate of Completion”.[10]

Students that have met the other graduation requirements but have not passed the FCAT and are not able to receive an exemption can still obtain a diploma if they pass another administration of the FCAT after grade twelve. Students can retake the FCAT in the summer of grade twelve or in the following academic year. They can prepare for the FCAT by taking an additional semester or full year (“thirteenth year”) of high school education or by taking remediation and FCAT preparation classes available at adult and community colleges.

Students wishing to enroll in state universities must have a high school diploma and must have acquired course credits in specific amounts and types. Students wishing to enroll in college credit courses at community colleges must have a high school diploma or must have a certificate of completion and pass a college placement test. Students that do not pass the college placement test are placed in college preparatory courses. Students can enroll in adult education programs without a diploma or a certificate of completion.

3. Theoretical model and empirical strategy

In this section we discuss how higher standards might affect various outcomes and we describe the empirical strategy that we will use to identify them. We begin with the empirical strategy. This can identify the effects of higher standards under assumptions that we find plausible. We then present a simple theoretical model. This is useful for fixing ideas about the mechanisms likely to give to rise to effects on dropout and graduation rates. We stress that this model makes several assumptions that are not needed for our basic identification strategy and so our results are robust to violations of these assumptions. These assumptions are useful only in so far as they help simplify the theory and thereby illuminate possible mechanisms.

[10] Special education students that have met the other graduation requirements but not passed the FCAT can receive a “Special Diploma for Students with Disabilities”.

3.1. Empirical strategy

Theoretical models such as those developed by Costrell (1994) and Betts (1998) suggest that a higher standard could affect students at all points of the ability distribution. That is because these are signaling models, and because a higher standard could change the returns associated with each observable type of “student” (e.g., students that obtain a diploma and students that do not). But if a higher standard affects all students, it is difficult to construct a counterfactual for the effect of higher standards on a particular student and, by extension, the aggregate impact of higher standards. One possible strategy would be to estimate the aggregate impact using an interrupted time series design. That is, compare the cohort first subject to the new higher standards with the last cohort subject to the old lower standards, adjusting for cohort trends. This approach is not, however, ideal: cohort trends are difficult to control for and there may be other reasons why outcomes differ across adjacent cohorts.

With this difficulty in mind, we use a difference-in-difference (DD) approach to identify the impact of the higher standard. This compares outcomes among two groups. The first group are students with initial scores that would have passed under the new higher standard irrespective of whether they took the exam before or after the standard was raised. We label these the “pass both” control group. The second group is students with initial scores that would have passed under the old standard but not the new standard. We label these the “pass old, fail new” treatment group. We argue that this approach will be valid provided two conditions are met. First, the higher standard must have no impact on the outcomes of the pass-both treatment students. This ensures that the evolution of their outcomes through the standards increase serves as a good counterfactual for the evolution of outcomes among the pass-old fail-new control group in the event that standards had not increased. Second, because these groups are defined in terms of initial exam scores, the higher standard must not have changed behavior in advance of the initial exit exam. This ensures that the effects of the higher standard can be measured using outcomes that occur after the initial exam is taken.

We think both of these assumptions are plausible. While the return to passing will not be independent of the standard in a formal model of signaling, it might be plausible to suppose that this condition holds in practice. Even if it is not, an increase in the returns to passing caused by the higher standard will cause us to *over-estimate* the impact on dropout.[11] Since we find at worst small dropout effects, these are still interesting if interpreted as worst-case scenarios. Since the higher standard was announced only a few months before the initial exam, and since students have many several opportunities to retake the exam if they fail

[11] For example, if the higher standards have no impact on the pass-old fail-new group, but decrease drop-out rates among the pass-both group (because the higher standards increase the return to continuing in education), we will conclude that the higher standards caused drop-out rates to increase among the pass-old fail-new group.

first time, it seems reasonable to suppose that the higher standards did not change behavior in advance of the initial exit exam. This is consistent with at least two aspects of the data. First, the distribution of initial scores is similar among the pre- and post-change cohorts (not reported). Second, we estimate difference-in-difference models in which the dependent variables include background characteristics such as eligibility for a free school meal and the probability that grade 10 students are "on time". Consistent with no anticipation effects, we see no impacts on background characteristics.

We also extend this model by considering a second treatment group: students with initial scores that would have failed under the old lower standards and the new higher standards (we call this the "fail both" group). This is an interesting group to consider because the higher standards meant they had to make larger improvements in order to pass. This could, for example, have made them more likely to drop out. This group presents no new conceptual problems. It does however present a practical problem: how to define it when it could, in theory, include all students that failed the exam. We experiment with several lower bounds. That is, we make different assumptions about how low the initial score must be for students not to have been affected by the higher standards. We show that this decision has no real impact on our main results.

3.2. Simple model

To help fix ideas about the possible effects of higher standards, we now sketch a simple model. In doing so, we make a number of assumptions. These include the two assumptions on which the DD strategy rests and others that help illuminate how the effects of higher standards might operate.

To begin, we assume that students take a single test at the end of grade 10. We denote their score on this test by t_1, where $t_1 \in [\underline{t}, \bar{t}]$. We assume that the score necessary to pass (i.e., the standard) is announced at the same time as the results of this test, such that pre-test behavior is the same across cohorts hence can be ignored. This means that we can think of the initial score as a measure of ability. We assume the standards for the pre- and post-change cohorts are L and H, where $\underline{t} < L < H < \bar{t}$.

Upon observing their scores and the passing score, we assume that students make two decisions. First, they decide whether to dropout of school at the end of grade 10 or stay in school until the end of grade 12. Second, if they decide to stay in school, they choose how much effort to exert. Effort is denoted e and the cost of effort (i.e., the cost of raising test scores) is assumed to be $(ce^2)/2$. If they stay in school, students retake the test at the end of grade 12. We assume their score on this test, denoted by t_2, is determined as follows:

$$t_2 = t_1 + e + \eta \tag{1}$$

where η is a normally distributed mean-zero random variable and $t_1 + e = E(t_2 \mid t_1)$. It follows that the probability of passing (and thereby graduating) for a student with first score t_1 that exerts effort e is:

$$P(grd|t_1, e) = P(t_1 + e + \eta > L) = F_\eta[e - (L - t_1)] \tag{2}$$

We assume that the wage return to a diploma is constant across students and independent of the passing standard. It follows that the utility of staying in school and the utility of dropping out of school can be written as:

$$\begin{aligned} U(stay|t_1, e) &= U_0^S + RP(grd|t_1, e) - \frac{ce^2}{2} \\ U(dropout) &= U_0^D + \varepsilon \end{aligned} \tag{3}$$

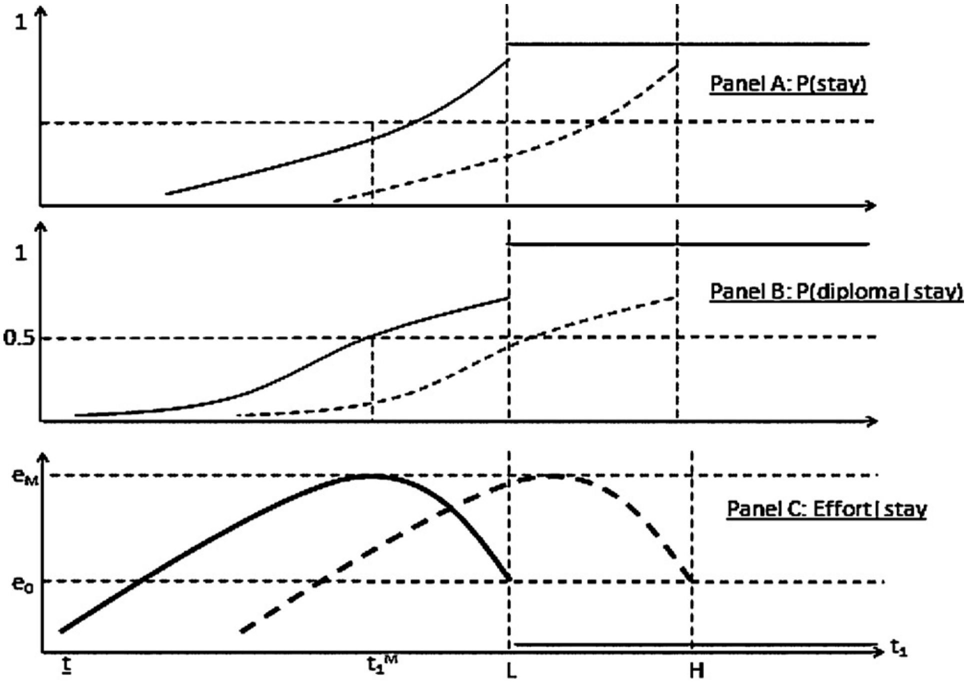
where U_0^S is the return to completing school without graduating and U_0^D is the return to dropping out of school at the end of grade 10. Both are assumed independent of ability and effort. The term R is the return to graduating conditional on completing grade 12 (i.e., the return to passing the exam). The term ε is a normally distributed error term that can be thought of as a person-specific utility term that is independent of ability. We assume this is also normally distributed with mean zero.

3.2.1. Outcomes under the low standard

The main results can be seen in Fig. 1 (Appendix A contains formal derivations). Starting with the bottom panel, the solid line characterizes optimal effort as a function of the initial test score and when facing a standard L. Note that optimal effort is highest when the pass probability is 0.5. That is because effort has the greatest marginal return at this point. It follows that optimal effort for high- and low-scoring students is lower than it is for medium-scoring students (i.e., for the medium-scoring group, effort moves the pass probability over a range closer to 0.5). For all students that failed the first test (i.e., scores less than L), optimal effort is positive (since the marginal return to a unit of effort exceeds the marginal cost). For all students that passed the first test (i.e., scores greater than L), optimal effort is zero: these students have already passed hence can obtain a diploma with no additional effort.

Moving up to the middle panel, the solid line characterizes the implied pass probability as a function of the initial scores given optimal effort under standard L. There are two reasons why, compared to low-scoring students, medium-scoring students are more likely to pass. First, they have a higher probability of passing conditional on effort. Second, they exert more effort. In contrast, high-scoring students exert less effort than medium-scoring students (see the bottom panel), but, as scores increase, effort decreases by less than one-for-one, hence the implied pass probability rises. Moving up to the top panel, the solid line characterizes the probability of staying in school under standard L. This is increasing in initial scores because higher initial scores imply a higher equilibrium pass probability.

Optimal effort depends, among other things, on R/c, the diploma return relative to the marginal costs of effort (see Appendix A). The larger is this ratio, the more effort will be exerted. It follows that this parameter will also help to determine the equilibrium probability of passing conditional on staying. It will also help to determine the probability of staying in school beyond grade ten: the larger



Notes: see text and Appendix for a discussion of these figures.

Fig. 1. The impact of tougher standards.

is this ratio, the more likely students are to stay. For example, if this ratio is zero, students will exert zero effort, will pass with only small probability and will view staying in school as a less attractive proposition.

3.2.2. *The impacts of a higher standard*

It is easy to analyze the impacts of a higher standard $H > L$. That is because, under the assumptions of the model, a student with first score t_1 facing standard $H > t_1$ will make the same decisions as a student with first score $t_1' = L - (H - t_1)$ facing standard L. Behavior under the higher standard is depicted by the dashed lines in Fig. 1. Note that there are two types of students that fail to meet the standard H: those that would have passed under the old standard (the pass old, fail new group with scores $L \leq t_1 < H$) and those that would not (i.e., the fail both group with scores $t_1 < L$).

For the "pass old, fail new" group, the effects of the higher standard are obvious. Under the low standard, they passed first time hence received a diploma with zero additional effort. Under the high standard, optimal effort is positive and both the equilibrium pass probability and the equilibrium probability of staying in school are increasing in the score. The extent to which the higher standard affects outcomes for this group will be increasing in c and R. Intuitively, for any given gain to passing, a larger c implies that it is more "expensive" to pass hence less attractive to stay on. For any given cost of passing, a larger R implies that it is more attractive to stay on under both standards, but less attractive in the high standards case because the probability that the standard is not passed imposes an implicit tax on this extra gain.

For the "fail both" group, the effort effects of a high standard are ambiguous, although the equilibrium pass probability is lower, as is the equilibrium staying-on probability. The intuition for the ambiguous effort effect is as follows. If the standard is only slightly higher (i.e., such that $H - L > L - t_1^M$), these students will be incentivized to work hard and try to meet it. If it is much higher, they will "give up" trying to meet it. The ambiguous effort effect means it is not clear how the cost of effort affects the impact of higher standards on the probability of staying on. Again though, the impact on staying on will be larger the larger is R. The intuition is as before: larger R increases the return to staying on, but higher standards impose a higher implicit tax on this gain.

We can summarize these predicted impacts as follows. First, for the pass old, fail new group, higher standards will increase effort, decrease the probability that students pass if they stay in school and decrease the probability that they stay in school. These effects on staying in school will be increasing in both R and c. Second, for the fail both group, higher standards will have ambiguous effects on effort, will decrease the probability that students pass if they stay in school and decrease the probability that they stay in school. These effects on staying in school will be increasing in R.

3.3. *Discussion*

As already noted, this model makes many assumptions. Most of these we do not need for our identification strategy to hold; they merely simplify the theoretical analysis. Three sets of assumptions stand out as being especially unrealistic. First, the model treats the labor market in a very superficial way. In particular, it assumes that there is a return to obtaining a diploma but otherwise no additional returns to effort. A more general model would allow for returns to effort in addition to possible sheepskin effects associated with graduating high school. Second, the model ignores postsecondary education. Again, this is done to simplify the analysis. Third, only one component of effort is modeled – that required to pass the high school exit exam. Other relevant components of effort include those required to stay in school (i.e., persistence, as opposed to learning) and that which improves productivity in the labor market. Again, the goal is to draw attention to a parameter that we expect will have a large impact on the effects of higher standards: the cost of improving test scores.

4. Data

The data used in this paper were provided by the Florida Education and Training Placement Information Program (FETPIP), a unit within the Florida Department of Education. FETPIP follows students from school into post-secondary education and into the labor market. The FETPIP data combine several data files that are linked at the student level using identifying information such as the student's name, date of birth and social security number. Since the data were linked before they were provided to us, match rates are unknown.

4.1. Data types

The Florida data consists of a base enrollment record matched to several other types of data. The base enrollment record refers to the academic year in which a student was first enrolled in grade ten in a Florida public school (defined to include charter schools but not private schools). Every individual in the Florida data is therefore associated with a unique base enrollment record and we use this to define the cohort to which the individual belongs. In particular, we define the "pre" and "post" treatment cohorts to include individuals in grade ten in 2000–01 and 2001–02, respectively.

Several types of data are matched to this base enrollment record:

1. Subsequent enrollments in Florida public schools: These enrollment records (and the base enrollment record) include school and grade details and time-varying student details such as free lunch status and special education classification.
2. Demographic data: These include information on sex, month and year of birth and race. Race is classified as White, Black, Hispanic, Asian, and other.
3. Grade 10 FCAT data: In principle, these include details of all grade ten FCAT attempts, including dates and scores obtained on both the math and reading sections. For the first and second cohorts, grade ten FCAT data are only available for spring exams and retakes (i.e., not for summer and fall retakes). Since both the first and last-chance exams are administered in spring, this is not an important constraint.
4. Awards data: These include details of all certificates and diplomas awarded to students. These data include the type of diploma awarded (e.g., high school diploma) and further details of the route by which it was obtained (e.g., met standard requirements, exempted from FCAT requirement).
5. Postsecondary enrollment and awards data: These are available for all students that attend state community colleges (CCs) and state universities (SUs) in Florida. The data include enrollment and awards files similar to those available at the high school level. We use the postsecondary information to define variables including "semesters in CCs" and "semesters in SUS".
6. Earnings data: These come from the Unemployment Insurance (UI) tax reports submitted to the Florida Department of Revenue by employers covered by the state's UI law. Covered employers are required to report, on a quarterly basis, the wages paid to each employee in order to determine the firm's tax liability. Wages will be reported for nearly all individuals working for earnings in Florida. The major categories not covered will be those working in the informal sector, the self-employed and those working in the military. For each individual in our data we have, for each quarter, earnings information as provided by each employer. We sum this to obtain "total earnings in the quarter" and deflate it to $2000 using the CPI-U series. Note that earnings could be observed to be zero because the samples have zero earnings, because they work in the uncovered sector or because they have left the state. Hence while we do not have any attrition from our data, we may have some observations with "false zero" earnings. Martorell and Clark (2010) consider this possibility and argue that it will not have a first-order impact on the estimates reported there. Similar conclusions likely imply here.

4.2. Analysis sample

The analysis sample includes students in grade 10 in 2000–01 and 2001–02. We define three groups of students based on their initial test scores. The groups are depicted in Fig. 2. The first group ("failed both") scored at levels that would have failed under both standards (students fail when at least one score is below the passing cutoff). The second group ("passed old, failed new") scored at levels that would have passed under the old standard but failed under the new. The third group ("passed both") scored at levels that would have passed under both. We could have included more students in the fail both and pass both groups but chose not to. The common trends assumption underlying the difference-in-difference strategy seem more plausible when the groups are defined using tighter test score ranges.

Table 1 presents some descriptive statistics for these groups. A majority of these students are girls, a large fraction (around one quarter) are eligible for a free or reduced price lunch (FRPL) and almost one quarter did not reach grade 10 on time.[12] Not surprisingly, students with initial scores that fail under both standards ("fail both") are more

[12] We generate this variable using data on year and month of birth.

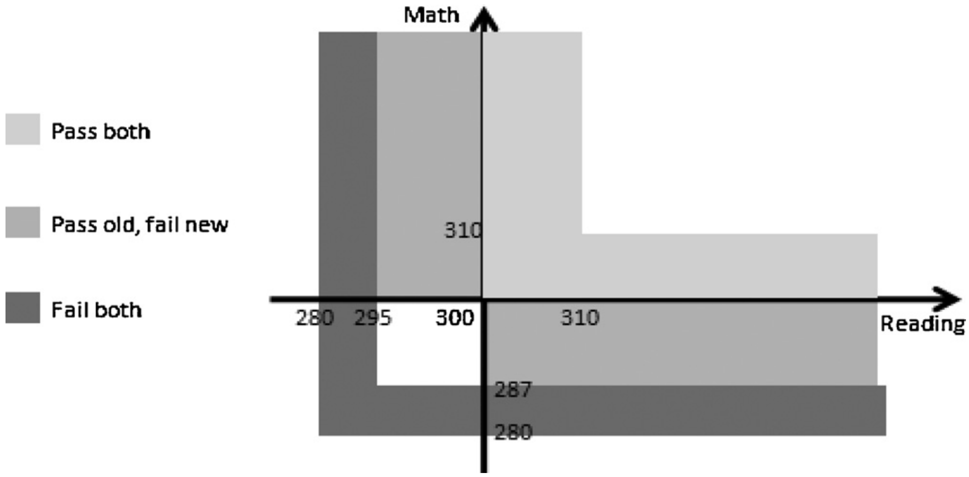


Fig. 2. Group definitions.

disadvantaged than students that pass under the old lower standard (“pass old, fail new”) or pass under both the old lower standard and the new higher standard (“pass both”).

5. Results

In this section we report difference-in-differences estimates of the impacts of the tougher standards. As discussed in the empirical strategy section, these estimates are the coefficients on the interaction of the two group dummies (“pass the old exam, fail the new exam” and “fail both exams”) and the “post” cohort dummy in regressions of outcomes on various control variables, the two group dummies, a post dummy that indicates the cohort faces the new higher standard and the interactions of the group dummies and the post dummy. The excluded group is the “pass both” group - students whose initial scores were high enough to pass the exam under both the old lower and new higher standards.

We begin by estimating these models on the subsample of students with math and reading scores in the 280–310 range. The “common trends” assumption underlying the difference-in-difference analysis is more plausible over this narrower range. The first column presents estimates from models without controls. In the second column we add baseline controls (reduced/free lunch, white, and 10th grade on time). In the third column we add school fixed effects. In the fourth through sixth columns, we report estimates generated from the same models but estimated over an even narrower range of data (285–305). These provide a check on the common trends assumption, since violations of this assumption should be revealed by differences between the first three and next three estimates. In the seventh through ninth columns we report placebo treatment estimates generated using a set of students that would all have passed under both the old lower and new higher standard (i.e., scores 300–310), some of which are incorrectly coded as having failed to meet a fictitious higher standard of 305. Since these placebo treatment effect estimates should be zero, they provide a useful specification test. In other words, non-zero placebo treatment effect estimates suggest violations of the common trends assumption and suggest that we should be skeptical of the estimates reported in columns one through six.

5.1. High school outcomes

We focus first on high school outcomes and we begin with highest grade completed (i.e., the inverse of high school dropout). Recall that we argued that tougher

Table 1
Descriptive statistics.

	Full sample		Group 1 “Fail both”		Group 2 “Pass old, fail new”		Group 3 “Pass both”	
	2000	2001	2000	2001	2000	2001	2000	2001
Male	45.82	45.66	42.90	44.13	47.81	46.40	46.18	46.02
White	54.02	52.90	48.48	45.90	53.41	52.25	58.43	58.02
Black	22.21	21.59	27.57	27.58	22.33	21.97	18.35	17.36
Hispanic	19.50	20.97	19.99	22.24	19.95	21.30	18.77	19.85
Free or reduced price lunch	27.91	26.48	30.39	30.03	28.38	26.98	25.76	23.73
Limited English proficient	17.21	18.74	18.68	21.25	17.83	19.46	15.64	16.49
Special education	5.99	6.45	7.85	8.87	5.85	6.44	4.80	4.87
Gifted	0.66	0.80	0.25	0.31	0.52	0.73	1.07	1.18
Grade 10 on time	77.49	77.08	74.12	72.89	77.66	77.00	79.71	79.89
N	31,759	35,259	8729	9161	10,602	12,063	12,428	14,035

Notes: First two columns present statistics for students in all three groups. Other columns are for each group (see text).

Table 2
The impacts of tougher standards on high school outcomes.

		(1) Using scores 280–310	(2)	(3)	(4) Using scores 285–305	(5)	(6)	(7) Using scores 300–310	(8)	(9)
			X	School FE		*X*	School FE		*X*	School FE
Highest grade completed [Mean = 11.2]	Fail both	0.002	0.003	0.005	−0.009	−0.005	0			
		(0.013)	(0.012)	(0.012)	(0.018)	(0.018)	(0.017)			
	Pass old, fail new	−0.008	−0.007	−0.006	−0.014	−0.011	−0.009			
		(0.011)	(0.011)	(0.011)	(0.013)	(0.013)	(0.013)			
	Score ≥ 305							−0.027	−0.01	−0.006
								(0.013)*	(0.013)	(0.013)
Last-chance sample [Mean = 3.8%]	Fail both	4.043	3.997	4.01	2.982	2.934	2.945			
		(0.380)**	(0.379)**	(0.382)**	(0.512)**	(0.506)**	(0.512)**			
	Pass old, fail new	4.203	4.208	4.211	4.203	4.209	4.214			
		(0.183)**	(0.183)**	(0.187)**	(0.183)**	(0.183)**	(0.189)**			
	Score ≥ 305							N/A	N/A	N/A
High school diploma [Mean = 82%]	Fail both	−1.209	−1.138	−1.224	−0.613	−0.286	−0.082			
		(0.790)	(0.777)	(0.773)	(1.119)	(1.100)	(1.095)			
	Pass old, fail new	−2.101	−2.066	−2.093	−1.658	−1.539	−1.408			
		(0.690)**	(0.682)**	(0.678)**	(0.819)*	(0.678)	(0.678)			
	Score ≥ 305							−0.763	0.267	0.422
								(0.777)	(0.814)	(0.812)
Certificate of completion [Mean = 0.98%]	Fail both	0.785	0.776	0.768	0.279	0.2790.2	0.262			
		(0.202)**	(0.203)**	(0.204)**	(0.269)	(0.269)	(0.272)			
	Pass old, fail new	0.794	0.794	0.804	0.708	0.707	0.742			
		(0.107)**	(0.107)**	(0.109)**	(0.114)**	(0.114)**	(0.115)**			
	Score ≥ 305							−0.029	−0.024	−0.007
								(0.083)	(0.091)	(0.092)
		67,018	67,018	67,018	45,530	45,530	45,530	26,463	26,463	26,463

Notes: Columns (l)–(6) report difference-in-difference estimates of the impact of tougher standards (i.e., the post period dummy) on the groups "fail both" and "pass old, fail new". Columns (l)–(3) use the main sample discussed in the text. Columns (4)–(6) use a subset of this sample. Columns (7)–(9) estimate placebo tests using a subsample of the data and a hypothetical treatment (score less than 305 on both tests). Robust standard errors in parentheses. Means in first column generated using subsample that are "post" and "pass both". * and ** refer to significance levels at the 5% and 1% levels respectively.

standards could reduce highest grade completed if the higher standards discourage students. We assume that the higher standards do not deter the younger of these cohorts from taking the initial exam and we argued that this was consistent with the empirical distributions of initial scores. Instead, we focus on highest grade completed conditional on taking the initial exam.

The estimates in columns (1) through (6) of Table 2 suggest that the higher standards had little impact on this outcome. The signs switch across specifications and all estimates are small and statistically indistinguishable from zero. The placebo estimates are also small and statistically insignificant, at least once the baseline controls have been added. Even in the worst case scenario for tougher standards (column (4)), the estimates suggest effects on highest grade completed of −0.04.

These estimates suggest that the tougher standards had no impact on dropout. Since the tougher standards increased the fraction of students that failed the initial attempt, we might expect to find a larger fraction of students taking the "last chance exam" – the retake which takes place at the end of grade twelve, a student's last chance to pass before scheduled graduation.[13] As seen in the second row of Table 2, this is exactly what we find. In particular, these estimates suggest that the tougher standards increased the probability of taking the last-chance exam by around four percentage points for the "pass old, fail new" group and by around three percentage points for the "fail both" group. These estimates are reasonably robust to the narrow observation window. We cannot implement the placebo specification because students that passed under both standards are never observed to retake at the end of grade 12.

Because we find no dropout effects and because the tougher standard increases the fraction of students retaking at the end of grade 12, we might expect the higher standard to have decreased the graduation rate (i.e., the fraction of students that passed the exam hence met all of the graduation requirements and obtained a high school diploma). Specifically, because the higher standard increased the probability of taking the last-chance exam by four percentage points, and because the pass rate on this exam is around one third, we might expect the tougher standards to decrease the graduation rate by one or two percentage points. This is roughly what we see in the third panel of Table 2. For the "pass old, fail new" group, there is a roughly 1.5 percentage point decrease in the graduation rate. Among the "fail both" group, the decrease is smaller and not statistically significant. The placebo tests are comfortably satisfied.

In Florida, but not necessarily in other states, students that complete twelve grades but do not receive a diploma are eligible to receive a certificate of completion. We might expect the graduation rate reduction caused by the tougher standards to be mirrored by an increase the fraction of

[13] If we found no effects on this outcome it would imply that students that failed at the new higher standard passed one of the retakes administered before the last-chance exam.

Table 3
The impacts of tougher standards on post-secondary outcomes.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Using scores 280–310			Using scores 285–305			Using scores 300–310		
			X	School FE		X	School FE		X	School FE
Semesters in 2-year college [Mean = 2.4]	Fail both	0.016	0.038	0.037	−0.003	0.028	0.029			
		(0.060)	(0.059)	(0.059)	(0.084)	(0.083)	(0.082)			
	Pass old, fail new	0.046	0.045	0.04	0.035	0.036	0.038			
		(0.057)	(0.056)	(0.056)	(0.066)	(0.067)	(0.056)			
	Score ≥ 305							−0.039	−0.052	−0.04
								(0.061)	(0.065)	(0.065)
Semesters in 4-year college [Mean = 0.63]	Fail both	0.137	0.141	0.138	0.108	0.129	0.13			
		(0.043)**	(0.042)**	(0.042)**	(0.057)	(0.056)*	(0.056)*			
	Pass old, fail new	0.075	0.071	0.063	0.035	0.038	0.034			
		(0.046)	(0.045)	(0.045)	(0.045)	(0.053)	(0.053)			
	Score ≥ 305							−0.118	−0.102	−0.093
								(0.048)*	(0.052)*	(0.052)
Earnings in 1st year post HS [Mean = 5227]	Fail both	64.21	84.10	74.93	−16.33	−1.39	−13.00			
		(109.42)	(108.71)	(108.82)	(131.40)	(130.56)	(132.13)			
	Pass old, fail new	−81.49	−59.02	−60.61	−128.62	−136.53	−146.92			
		(109.94)	(109.27)	(108.98)	(158.74)	(157.83)	(158.55)			
	Score ≥ 305							−243.675	−70.21	−86.144
								(123.7)*	(128.89)	(133.80)
Earnings within 5 years [Mean = 37,056 9]	Fail both	276.16	408.47	407.29	−266.97	−131.78	−102.71			
		(662.30)	(660.14)	(659.36)	(784.56)	(781.43)	(787.41)			
	Pass old, fail new	−147.59	−44.97	−123.09	−308.59	−340.69	−394.00			
		(683.75)	(681.82)	(680.56)	(964.14)	(961.08)	(966.44)			
	Score>=305							−1459.16	−138.30	−236.78
								(753.48)	(788.42)	(820.03)
	N	67,018	67,018	67,018	45,530	45,530	45,530	26,463	26,463	26,463

Notes: see notes to Table 2.

students receiving such a certificate. As seen in the bottom panel of Table 2, this is roughly what we find. The Certificate rate increase is lower than the graduation rate decrease, perhaps because students choose not to obtain a Certificate because they plan to retake the exam after the twelfth grade and try to earn a diploma.

To summarize, we find that the tougher standards did not cause students to dropout earlier than they would have done otherwise. This is an important result, which we return to below. Instead, we find that by the end of twelfth grade, a higher fraction of students have still not passed the exam. As a result, the tougher standards reduce the graduation rate, one consequence of which is to increase the fraction of students that obtain a Certificate of Completion.

5.2. Postsecondary education

We noted above that students do not need a high school diploma to enroll in college. As such, it is not clear why we would expect to find any effects of tougher standards on postsecondary outcomes. Consistent with this line of reasoning, the top two panels of Table 3 do not reveal any clear impacts on the number of semesters that students were enrolled in two-year college. There appear to be significant and positive (i.e., wrong-signed) impacts on the number of semesters that students were enrolled in four-year college, but only for the "fail-both" group, only a tiny fraction of whom enroll in a four-year college. The placebo tests underline the fragility of this result: these show a negative and statistically significant impact of the false placebo treatment on the number of semesters enrolled in four-year college.[14]

5.3. Earnings

Our estimates of the effects of tougher standards on high school outcomes have ambiguous implications for their impacts on earnings. On the one hand, since tougher standards reduce graduation rates, they could reduce earnings. This assumes there is a signaling value to a high school diploma, such that workers that do not graduate suffer an earnings loss relative to workers that do. On the other hand, by forcing students to retake the exams that they failed, the tougher standards could increase student effort and increase the amount of time spent on basic skills and the amount of attention received by teachers. All of these effects could serve to increase earnings (Tyler, 2004).[15]

Even in the best- and worst-case scenarios associated with these hypotheses, it is difficult to imagine the tougher standards having large impacts on earnings. For example, assuming the tougher standards caused a three percentage point drop in the graduation rate (at the top end of our estimates) and assuming a signaling value of 30 percent (at the very top end of the range of estimates found in the

[14] These outcomes exclude enrollment in private colleges or in colleges outside of Florida. Since students can enroll in two-year colleges without a high school diploma, it is not obvious why the tougher standards should impact the fraction of students that enroll in these types of colleges.

[15] If students are demotivated by failing exams (e.g., suffer a loss of confidence), then this could counteract these effort effects.

Table 4
Falsification tests.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Using scores 280–310			Using scores 285–305			Using scores 300–310		
			X	School FE		*X*	School FE		*X*	School FE
Male [Mean=46.5%]	Fail both	0.014 (0.010)		0.029 (0.013)*						
	Pass old, fail new	−0.013 (0.009)		−0.015 (0.011)						
	Score ≥ 305							−0.058 (0.010)**		
Eligible for free or reduced price lunch [Mean=25.9%]	Fail both	1.675 (0.869)			1.378 (1.220)					
	Pass old, fail new	0.64 (0.799)			1.214 (0.950)					
	Score ≥ 305							−1.671 (0.874)		
"On time" in grade 10 [Mean = 77.1%]	Fail both	−1.415 (0.825)			−2.044 (1.170)					
	Pass old, fail new	−0.839 (0.745)			−1.453 (0.888)					
	Score ≥ 305							−1.098 (0.830)		
	N	67,018	67,018	67,018	45,530	45,530	45,530	26,463	26,463	26,463

Notes: see notes to Table 2.

literature – see Martorell and Clark (2010)), we would expect to see negative earnings effects of around one percentage point – around $400 when the outcome is cumulative earnings over the first five years (average $37,000 – see Table 3). Assuming the tougher standards increased effort (broadly construed) by 10%, and assuming a rate of return to a year of education of 10 percent (in line with the estimates reported by Card (1999)), we would expect to see positive earnings effects of around one percentage point. If there some of both effects operate, we would expect net earnings effects to be somewhere within $400 of zero.

We consider two earnings outcomes: earnings in the first year after high school (i.e., the first four quarters after the scheduled high school graduation) and earnings in the first five years after high school (i.e., the first 20 quarters). The estimates are reported in the bottom two panels of Table 3. Thee are nearly all in the predicted range. They are not, however, precise enough to reveal which types of effects dominate. Indeed, in columns (3) and (6), we see a mixture of positive and negative estimates. The only reasonable conclusion is that any earnings effects are too small to be detected in these data. Note that adding data on earnings at older ages is unlikely to solve this problem. That is because we might expect the effects of tougher standards to be weaker at older ages. We would certainly expect signaling effects to weaken over time (as firms acquired more productivity information). Effort effects might also weaken, although the dynamic effects of increased effort are not obvious. Either way, it is plausible to suppose that adding earnings at later ages would dilute any earnings effects of tougher standards.

5.4. Robustness checks

Finally, we check robustness by estimating "effects" on pre-determined characteristics. Because these characteristics are pre-determined, the higher standard cannot possibly affect them. As a result, these estimates give us a sense of the likely robustness of our findings. In particular, significant effects on these characteristics would suggest that the treatment is correlated with observables (even in the difference-in-difference framework), hence may also be correlated with unobservables. In fact most of these estimates are statistically indistinguishable from zero (as are others not reported). The only outcome for which one might suspect that there are effects is "grade 10 on time". Since the effects are negative across all specifications, including the placebo specification, it may be that the ability profile of retention in earlier grades was different in the pre and the post cohort (i.e., higher-ability students were more likely to be retained in the post cohort). For two reasons, we do not think this is major concern. First, nearly all of our estimates are robust to including controls for this variable (compare columns (1) and (2) and (4) and (5) in Tables 2 and 3). Second, nearly all of our placebo estimates suggest no effects, even though there is a placebo effect on this variable (albeit not statistically significant) (Table 4).

6. Discussion and interpretation

Our analysis suggests that the tougher standards had no impact on high school dropout rates, but led to small reductions in high school graduation rates. These are associated with increases in the fraction of students taking the last-chance exam, increases in the fraction of students that obtain a Certificate of Completion, but no impact on post-secondary enrollment or earnings.

To put these graduation effects into perspective, we used our estimates to derive an estimate of the impact of the tougher standards on the overall graduation rate (i.e., not just the impact on the fail both and fail old, pass new groups. To do so, we calculated the pre-reform graduation rate for every score in the initial math and reading distributions. We then used these pre-reform graduation rates and our estimates of the graduation rate impacts of the reform to calculate post-reform graduation rates at each score level. From there it is easy to aggregate up and calculate the difference in graduation rates in the pre and post reform periods. In calculating the post-reform graduation rate, we assumed no effect on the pass both group, we imposed the estimated effect on the pass old, fail new group and we imposed the estimated effect on the fail both group.

Since it is difficult to know how to define the fail both group (i.e., at what point does this become so big that the difference-in-difference assumptions no longer hold?) we repeated this procedure for many possible definitions of this group. Using the definition employed until now (i.e., only including students with scores greater than 280 – see Fig. 2) yields an overall effect of 0.2 percentage points. Changing the 280 threshold to 200 increases this to one percentage point. From there it is robust to thresholds based on even lower scores. Although this range of estimates is a wide one, this exercise does at least rule out an overall graduation rate effect of bigger than one percentage point. Interestingly, a Florida report (Office of Program Policy Analysis and Government Accountability (2007) shows that the fraction of students receiving a standard diploma decreased by 0.4 percentage points between 2003 and 2004, when most of the tenth grade students in the pre and post cohorts would have graduated.

It is interesting to consider whether these results are consistent with estimates found in the previous literature. Although we are unaware of previous analyses of the impact of toughening standards, there are at least two strands of related research. One strand looks at the impacts of introducing these types of exams. Dee and Jacob (2007) is the latest and perhaps the most comprehensive paper in this line of research. They use an across-state difference-in-difference strategy that exploits the fact that different states adopt these policies in different years. Like the previous literature, they find no clear evidence that high school exit exams reduced high school completion rates (i.e., completion of grade 12). They find small statistically significant effects when they split by sex and race, although the race- and gender-specific effects are unrelated to academic performance, at least as proxied by the baseline completion rate. They find no effects on college enrollment and no clear evidence on labor market outcomes.[16] A second strand of literature uses regression discontinuity designs to look at the impacts of failing the initial exam on outcomes including high school dropout and graduation rates. Using data from Texas, Martorell (2010) finds that failing an initial exam does not discourage students. He finds however, that students that fail an initial exam are less likely to graduate, because they cannot pass the exam by the end of twelfth grade. Using data from Massachusetts, Papay, Murnane, and Willett (2008) and Ou (2010) find similar results.[17]

Both sets of findings are broadly consistent with those presented here. This is interesting for two reasons. First, it suggests that our findings of small effects is not driven by our focus on a well-established program or by our decision to condition on grade ten scores. Second, it suggests that the regression discontinuity estimates, which identify effects only for students at the pass-fail margin, generalize to students further away from the cutoff, such as those in the "fail old, pass new" and "fail both" groups that we consider.

It is also interesting to consider why tougher standards appear to have such small effects. One possibility is that students were not aware of the policy. That is, the students in the "pass old, fail new" group did not realize they would have to pass the exam at a later attempt in order to graduate; the students in the "fail both" group did not realize they would have to meet a higher standard in order to graduate. This does not seem plausible, especially when applied to the first of these two groups: those students must have been told they would have to retake the exam. A second possibility is that the return to graduation (R in the theoretical model) is low. In the extreme case, if R was zero, such that there was no signaling value to the high school diploma, rational students would act as if there was no threshold at which the exam was passed and would instead choose effort levels in a way that optimally traded off the current costs of effort against the future labor market returns to effort. Evidence presented by Martorell and Clark (2010) suggests that the signaling value of a diploma is, indeed, small, although this does not imply that students are aware of this information and act on it.

A third possibility is that students are overconfident, and under-estimate the probability that they will be unable to pass the exam. Combined with a tendency to follow the default route of staying in school unless forced to do otherwise, this would be a powerful force serving to blunt any effects of the tougher standards. With the data at hand, it is difficult to determine the importance of over-confidence. Nevertheless, there is ample evidence to suggest it is likely a factor. For example, Fischoff et al. (2000) use data from NLSY 1997 to show that only 7% of teens expect to have not completed high school by age 20, compared with the 16% that they report from then-current data. Other data suggest over-confidence is not confined to lower-ability or younger teenagers. In a study of the impact of financial incentives on the completion of the first year of an economics course at the University of Amsterdam, Leuven, Oosterbeek, and van der Klaauw (2010) report evidence from a baseline survey which suggests that almost twice as many students expect to complete the first year as actually complete the first year.

7. Conclusion

Our analysis suggests that a toughening of Florida high school graduation standards led to a small decrease in

[16] In a separate analysis of the impact on dropout rates using data from Minnesota, they find that the introduction of an exit exam led to a slight increase in graduation rates. When they focus on low-income districts however, they find that the exam decreased graduation rates by two or three percentage points.

[17] Papay et al. (2008) and Ou (2010) do however find a large effect of failing the mathematics exam for low-income urban students.

high school graduation rates with no adverse impacts on other outcomes. While some might conclude that standards should be made even tougher, it is not clear that the tougher standards yielded any benefits. We can be reasonably sure that they resulted in some students spending more time studying basic math and reading (since they had to retake the exams and since we know that schools typically enroll retakers in remedial math and reading classes), yet this implies that less time is spent on other courses or activities. It is not clear whether this improves earnings. Unfortunately, even the large samples that we employ are too small to detect any such earnings effects. Moreover, it is not clear that a more drastic toughening of standards would have had the same effect. It may be that standards can rise without consequence until a tipping point is reached, at which point the higher standard causes large numbers of students to drop out of school. Again, given the policy being studied and the data to hand, we cannot shed light on this possibility.

Others might draw the opposite conclusion, that weaker standards do no harm and hence that these exams could profitably be abolished. Again, such a conclusion cannot reasonably be drawn from the evidence presented here. For example, the abolition of these exams could result in a less intense focus on math and reading skills, and this may harm students' labor market prospects.

In our view, there are likely diminishing returns to additional evaluation of exit exam policies – whether based on cross-state difference-in-difference methods, within-state regression discontinuity methods or the within-state difference-in-difference method used here. The research based on each of these approaches points to broadly the same conclusion, and it is hard to imagine further evaluations coming to very different conclusions.[18] A more fruitful approach may be to focus on the mechanisms that will ultimately determine the effects of these policies. For example, while the theory behind exit exams assumes that diplomas carry a signaling value (e.g., Costrell (1994) and Betts (1998)), Martorell and Clark (2010) find the signaling value of a diploma to be small. Further research on signaling can inform the discussion of exit exams. Research into student expectations and motivation in the face of high-stakes exams can also inform the discussion and help policy-makers make more informed decisions in relation to the existence and design of exit exams and related policies designed to incentivize students in high school.

Acknowledgments

For useful discussions, we thank Paco Martorell and seminar participants at Bristol University, Clark thanks the Florida Department of Education for generously supplying the Florida data used in this paper and for providing invaluable information on the relevant institutions. Much of this work was completed while Clark was visiting the Industrial Relations Section at Princeton University. Financial support was generously provided by the U.S. Department of Education (Under Grant R305R060096) and by the National Institutes of Child Health and Human Development (Under Grant R01HD054637-01A1). The opinions expressed in this paper are ours, and do not necessarily reflect the views of the Florida Department of Education, the U.S. Department of Education or the National Institutes of Child Health and Human Development.

[18] Where the results differ across these studies (e.g., the effects on particular subgroups or in particular schools), these differences may reflect specific features of the systems being studied. Since these systems differ so much across states however, one doubts that these differences can be used to identify the effects of specific features of these systems.

Appendix A.

We make a series of arguments to establish the properties of the graphs presented in Fig. 1.

Bottom panel

From Eqs. (2) and (3), effort is chosen to maximize the value of staying in school (which we denote $V(e, t_1)$):

$$V(e,t_1) = RF_\eta[e-(L-t_1)] - \frac{ce^2}{2}$$

the first-order condition that the optimal effort choices must satisfy is:

$$\frac{R}{c} f[e^* - (L-t_1)] = e^*$$

For any t_1, the left-hand-side of this equation is a positive function that takes a maximum at $e = L - t_1$ and is symmetric around this. This follows from the properties of f. The right-hand side of this equation is a function $g(e)$ for which $g(0) = 0$, $g' > 0$ and $g'' = 0$. We now establish several facts about $e^*(t_1)$:

1. For some value of t_1 which we call t_1^M, we can show that $e_M = L - t_1^M$ is a unique solution to this equation. For the existence of such a solution, we only need to show that there is a t_1^M such that $f(0) = (c/R)(L - t_1^M)$. Such a value is $t_1^M = L - (R/c)f(0)$. For uniqueness, we need to show that for $t_1 = t_1^M$, there can be no other e such that $f(e, t_1^M) = g(e)$. There can be no other $e > e_M$, since $f(e, t_1^M) < g(e)$ for all $e > e_M$. This follows because $f' < 0$ and $g' > 0$ for $e > e_M$. There can be no other $e < e_M$, since $f(e, a_M) > g(e)$ for all $e > e_M$. This follows because $f(0, a_M) > 0$ and $f'' < 0$ and $g(0) = 0$ and $g'' = 0$ and f and g are continuous.
2. For $t_1^M < t_1 \leq L$, we can show that there is a unique value of e^* that solves this first-order condition, where $L - t_1 < e^*(t_1) < e_M$. We must have $e^* < e_M$ because $f(e_M, t_1) < f(e_M, t_1^M) = f(0) = g(e_M)$. Hence $f(e_M, t_1) < g(e_M)$ and we know that $f' < 0$ and $g' > 0$ for $e > e_M$ so that no $e > e_M$ can solve this equation. We must have $e^* > L - t_1$, since $f(L - t_1, t_1) = f(0) = g(e_M)$. Since $t_1 > t_1^M$, $g(L - t_1) < g(e_M)$, hence $f(L - t_1, t_1) > g(L - t_1)$ and we know that $f'' < 0$ and $g(0) = 0$ and $g'' = 0$ and f and g are continuous so no value of e lower than $L - t_1$ can solve this equation. By the continuity of f and g, it follows that a unique e^* solves this equation, where $L - t_1 < e^*(t_1) < e_M$.
3. For $t_1^M < t_1 \leq L$, we can show that the unique value of e^* that solves this first-order condition is decreasing in t_1. This follows from differentiation of the first-order condition. This yields

$e^{*'}(t_1) = f'/((c/R) - f') < 0$ since $f' < 0$ for $t_1^M < t_1 \leq L$ and $L - t_1 < e^*(t_1) < e_M$.

4. For $\underline{t} < t_1 < t_1^M$, we can show that there is a unique value of e^* that solves this first-order condition, where $0 < e^*(a) < e_M$. We must have $e^* < e_M$ because $f(e_M, t_1) < f(e_M, t_1^M) = f(0) = g(e_M)$ as in (2a). We must have $0 < e^*(t_1)$ because $f(0) > g(0) = 0$. Again, by continuity and the properties of f and g, it follows that $0 < e^*(t_1) < e_M$.
5. For $\underline{t} < t_1 < t_1^M$, we can show that the unique value of e^* that solves this first-order condition is increasing in t_1. This follows from differentiation of the first-order condition. This yields $e^{*'}(t_1) = f'/((c/R) - f') > 0$ since $f' > 0$ and $(c/R) - f' > 0$ for $\underline{t} < t_1 < t_1^M$ and $0 < e^*(t_1) < e_M$.
6. For $t_1 = L$, denote optimal effort e_L. For $t_1 = \underline{t}$, denote optimal effort e_0. It follows that $e_0 < e_L$ provided $f(e_L, L - t_1) < f(e_L, 0)$. We assume this is true.
7. From (1)–(6) and the continuity of f and g, the optimal effort function can be depicted as in the bottom panel of Fig. 1.

Middle panel

From Eq. (2), the equilibrium pass probability (i.e., the pass probability given t_1 and optimal effort) is:

$$P^*(t_1) = F_\eta[e^*(t_1) - (L - t_1)]$$

where $e^*(t_1)$ is defined in the bottom panel of Fig. 1. Note that:

1. It is simple to show that $P^*(t_1^M) = 0.5$.
2. Since $e^{*'}(t_1) > 0$ for $\underline{t} < t_1 < t_1^M$ then $P^{*'}(t_1) > 0$ for $\underline{t} < t_1 < t_1^M$.
3. For $t_1^M < t_1 \leq L, f' < 0$ hence $|e^{*'}(a)| = |f'/((c/R) - f')| < 1$. It follows that $P^{*'}(t_1) > 0$.
4. It must be the case that $P^*(L) < 1$ since $F_\eta(e_L) < 1$.

Top panel

The maximized value of staying in school can be written:

$$V^*(t_1) = RP^*(t_1) - \frac{ce^*(t_1)^2}{2}$$

From the envelope theorem, this maximized value is increasing in t_1, decreasing in L, increasing in R and decreasing in c. That follows from the differentiation of $V^*(t_1)$ with respect to those parameters (treating effort as fixed). Intuitively, the value of staying in school increases with ability relative to the standard and with the return to graduation relative to the cost of graduation.

References

Betts, J. (1998). The impact of educational standards on the level and distribution of earnings. *American Economic Review*, *88*, 66–76.

Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter, & D. Card (Eds.), *Handbook of labour economics volume 3A*. Amsterdam: Elsevier.

Costrell, R. (1994). A simple model of educational standards. *American Economic Review*, *84*, 856–971.

Dee, T. S., & Jacob, B. (2007). Do high school exit exams influence educational attainment or labor market performance. In A. Gamoran (Ed.), *Standards-based reform and children in poverty: Evidence from no child left behind*. Washington, DC: Brookings Institution.

Fischoff, B., Parker, A., de Bruine, W. B., Downs, J., Palmgren, C., Dawes, R. M., et al. (2000). Teen expectations for significant life events. Public Opinion Quarterly, 64, 189–205.

Florida Department of Education (2005a). Assistance to 11th and 12th grade students that have not earned passing scores on grade 10 FCAT for 2004–2005 (available online).

Florida Department of Education (2005b). FCAT handbook – A resource for educators (available online).

Greene, J. P., & Winters, M. (2009). Identification and estimation of treatments with a regression discontinuity design. *Economics of Education Review*, *28*, 135–142.

Leuven, E., Oosterbeek, H., & van der Klaauw, B. (2010). The effect of financial rewards on students achievement: Evidence from a randomized experiment. *Journal of the European Economic Association*, *8*, 1243–1265.

Martorell, P. (2010). The impact of high school exit exams on high school graduation and persistence. Mimeo, RAND.

Martorell, P., & Clark, D. (2010). The signaling value of a high school diploma. Industrial Relations Section Working Paper 557, Princeton University.

Office of Program Policy Analysis and Government Accountability (2007). Report Number 07–27.

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., Murnane, R. J., & Willett, J. B. (2008). The consequences of high school exit examinations for struggling low-income urban students: Evidence from massachusetts. NBER Working Paper 14186.

Tyler, J. (2004). Basic skills and the earnings of dropouts. *Economics of Education Review*, *23*, 221–235.