Adding it Up: How the Florida College and Career Readiness Initiative Impacted Developmental Education

Christine Mokher, Ph.D.
Associate Professor
Florida State University

Daniel Leeds, Ph.D.
Research Analyst
CNA Education

Julie Harris, Ph.D.
Research Analyst
CNA Education

1

Abstract

The Florida College and Career Readiness Initiative (FCCRI) was a statewide policy requiring college readiness testing and participation in college readiness courses for high school students. We used regression discontinuity to compare outcomes for students scoring just above and below test score cutoffs for assignment to FCCRI. We also examined impacts for students from a wider range of academic performance by using a before-after regression analysis to compare outcomes for targeted students before and after their schools implemented the FCCRI. The FCCRI increased the likelihood of enrolling in nondevelopmental courses for some targeted students, although results differ by academic performance. However, smaller differences in the likelihood of passing nondevelopmental courses suggest that some students were not prepared for these courses.

(Word count: 120/120)

Adding it Up: How the Florida College and Career Readiness Initiative Impacted Developmental Education

In 2010, approximately 86 percent of entering community college students believed that they were academically prepared for college, yet 67 percent tested into developmental coursework (Center for Community College Student Engagement, 2016). High developmental education participation levels are a problem for a few reasons. First, providing developmental education is costly. Researchers have estimated the annual national cost of developmental education at $7 billion (Scott-Clayton, Crosta, & Belfield, 2014). In addition, the evidence for the effectiveness of developmental education is mixed, with some research showing it is effective and some showing it is ineffective (e.g, Bettinger & Long, 2009; Martorell & McFarlin, 2011).

This study examines the impact of legislation enacted by Florida for a statewide program known as the Florida College and Career Readiness Initiative (FCCRI), which was intended to reduce the need for developmental education. The FCCRI consisted of testing grade 11 students to determine their college readiness and offering math and English college readiness and success (CRS) courses in grade 12 for students who did not test college-ready the year before. We begin by reviewing the literature on student transition from high school to postsecondary education and the impact of similar policies in other states. Next we provide an overview of the state policy context in Florida and changes to the policy over time. Then we describe the data and two different methods used to estimate program impacts. We used a sharp intent-to-treat regression discontinuity design to compare outcomes for students scoring just above and below test score cutoffs for assignment to the FCCRI.[1] We also examined the impact of offering the FCCRI to students from a wider range of academic performance levels by using regression analysis to compare outcomes for targeted students in schools before and after the schools implemented the FCCRI. We find that the FCCRI increased the likelihood of enrolling in nondevelopmental courses for some targeted

students who seamlessly enrolled in college, although the results differ based on student performance. However, smaller differences between the treatment and control groups in the likelihood of passing nondevelopmental courses suggest that some students may not be prepared for these courses. We conclude with implications for researchers and policymakers to consider for similar programs in other states, and directions for future research.

Literature Review

Developmental education is necessary because many high school graduates do not have the requisite skills to complete college-level courses (e.g., Achieve, 2016; Boser & Burd, 2009; Strong American Schools, 2008). These students often do not recognize that they lack the preparation necessary to complete for-credit college courses and that they will be required to enroll in developmental education. The gap in understanding college readiness can be traced to the governance divide between K-12 and postsecondary institutions as well as to complexities in procedures that are meant to alert students to their college readiness. Often high schools and colleges within a state use different college readiness tests—even when states attempt to align college readiness testing at high schools and colleges, the resulting policy is often fragmented or incomplete. When college readiness testing is not consistent between the high school and college levels, students do not receive a clear message about their college readiness. Better alignment would not only help students prepare appropriately for postsecondary education, but might also assist teachers by providing focused and centralized college readiness standards.

High school curriculum also plays an important role in preparing students for college. Students who participate in rigorous high school classes are more likely to persist at a postsecondary institution, remain enrolled at their initial institution, and pursue a bachelor's degree track if they transfer institutions (Attewell & Domina, 2008; Long, Conger, & Iatarola,

4

2012). Furthermore, postsecondary enrollment correlates more closely with high school curriculum than with high school test scores or class rank (Adelman, 2006). These findings suggest that policy initiatives such as the FCCRI that are designed to boost enrollment in more rigorous high school courses may improve postsecondary outcomes.

There is limited evidence on the effectiveness of programs like the FCCRI. Kurlaender, Jackson, Grodsky, and Howell (2016) examined a portion of California's Early Assessment Program (EAP), which was designed to bridge the gap between secondary and postsecondary institutions. The program included a voluntary assessment of college readiness in grade 11 similar to the FCCRI's. Using a difference-in-difference approach, the researchers found that the college readiness assessment slightly reduced the need for developmental education in math and English, with the largest impact for students near the readiness cutoff.

Hurwitz, Smith, Niu, and Howell (2015) assessed a policy in Maine that required students to take the SAT in grade 11. The policy, designed to meet NCLB's accountability requirements by providing an indication of student and school performance, had the additional benefit of helping students meet the admissions requirements of many four-year colleges. Using difference-in-difference methodology, the authors found that mandatory testing raised college-going rates by two to three percentage points. However, the authors did not look at the impact on developmental education participation rates.

To our knowledge, only one other study has examined the impact of transition courses. Phaett, Trimble, and Barnett (2016) studied West Virginia's intervention, which required schools to offer a math transition course to grade 12 students who scored below mastery on the state's standardized exam in grade 11, though course enrollment was voluntary. They found that the math transition course had a negative impact on the likelihood of passing a gateway course. The

researchers hypothesized that the courses may not have been aligned with college readiness testing and that students may have taken the transition course in place of more rigorous courses.

Florida Policy Overview

In 2008/09, Florida introduced its own statewide initiative, the FCCRI, to improve the state's college and career readiness rate. This differs from other college readiness initiatives, such as summer bridge programs, in that it provides students with earlier information about their level of college readiness and focuses solely on improving the academic skills needed for college. The FCCRI is also geared toward a much broader group of students than other readiness interventions, which tend to be more narrowly targeted toward college-bound students from disadvantaged backgrounds.

In grade 10, all students took the Florida Comprehensive Assessment Test (FCAT). The FCAT assesses basic skills learned through grade 10, but not higher-level skills learned later in high school and needed for college success. Students who scored in the midrange of the FCAT were targeted for college readiness testing in grade 11. Students scoring above this range were assumed to be college-ready, while those below were considered at risk of not graduating from high school and were required to take a course to help them pass the FCAT. Students who did not score college-ready on the test in grade 11 were assigned to CRS courses in grade 12. Offered in math and English, CRS courses are intended to inform students of their college-readiness status and help them develop the skills needed to score college-ready and succeed in college courses. The FCCRI differs from other statewide initiatives in that Florida was the only state where community colleges and high schools administered the same test (Achieve, 2012).

Implementation of the FCCRI began during the 2008/09 school year under Senate Bill 1908 (2008). Under the initial policy, herein referred to as the voluntary FCCRI, students could

choose whether or not to take the College Placement Test (CPT) in grade 11 and CRS courses in grade 12. Starting in 2011/12, the FCCRI became mandatory for targeted students under House Bill 1255 (2011). The updated policy, herein referred to as the mandatory FCCRI, made three further changes. First, a new assessment, the Postsecondary Education Readiness Test (PERT), replaced the CPT. Second, high schools were required to administer the PERT directly, rather than partnering with colleges to administer the CPT. Third, CRS coursework became mandatory for targeted students scoring below college-ready. There were three cohorts under the voluntary FCCRI (V1-V3) and two cohorts under the mandatory FCCRI (M1-M2).

New legislation in 2015 eliminated the requirements for common placement testing and CRS courses, and participation in both components became voluntary at both the student and school levels beginning in the 2015/16 school year. These changes occurred after the cohorts for our evaluation completed high school and do not affect the impact analyses.

Description of Treatment and Counterfactual Conditions

The Florida Department of Education (FLDOE) set standards for CRS courses that defined the topics to be covered. However, districts, high schools, and teachers had considerable discretion in how they implemented these courses and in the curricular materials that were used. This means that even though all CRS courses shared the same label, implementation differed across schools and even within schools. However, many math CRS courses took the form of review of Algebra I and II. English CRS courses tended to resemble a traditional English IV course, although many teachers indicated that they placed more emphasis on integrating ACT or PERT preparation and study skills into the CRS course (Mokher et al., 2014). Schools tended to assign teachers with strong credentials to these courses; CRS teachers were more likely to have graduate degrees and more years of experience compared to the statewide population of teachers

(Mokher, Jacobson, Rosenbaum, & LaLonde, 2013).

During the first four years of the FCCRI (the timeframe for this analysis), approximately 84 percent of students were targeted for college readiness testing in math and 57 percent in English; thus, most of the student population is affected by this initiative. The state did not impose sanctions for noncompliance, so not all schools participated. Roughly 60 percent of schools offered college readiness testing during the voluntary FCCRI, rising to 96 percent during the first mandatory year. Fewer than 50 percent of schools offered CRS math courses and fewer than 25 percent offered CRS English courses in the voluntary period; both rates rose to over 90 percent when the policy became mandatory.

Nontargeted students also participated in CRS courses for several reasons. At some schools, students who did not take the PERT were placed in CRS courses. Additionally, some districts eliminated standard-level grade 12 English courses, making the CRS course the default standard-level English course. In some cases, administrators placed nontargeted students in CRS courses when the students were struggling in more advanced classes or needed an additional credit to graduate.

Course-taking by voluntary cohorts presents evidence of the counterfactual. In math, lower- to mid-performing targeted students primarily shifted away from "other regular math" courses (which includes courses such as Financial Applications, Analytic Geometry, and Liberal Arts Math) and Algebra II. Higher-performing targeted students also shifted away from "other regular math," as well as from "other advanced math" courses (which includes courses such as AP Statistics and Honors Probability). In English, targeted students, particularly those who were higher-performing, were more likely to take regular English courses (including CRS), and less likely to take honors-level English courses. This suggests that the FCCRI could have induced

8

students on the margin into less rigorous coursework, which influence the effect of the FCCRI for these students.

Research Questions

The purpose of this study is to examine the impact of the FCCRI on student outcomes at the end of high school and during the first year of college. The research questions are as follows:

1. What is the impact of the FCCRI on outcomes by the first year of college for:

   a. Students just below the upper FCAT threshold for assignment to college-readiness testing (highest-performing targeted students) compared to students just above the threshold?

   b. Students just above the lower FCAT threshold for assignment to college-readiness testing (lowest-performing targeted students)?

   c. Students just below the PERT college readiness threshold who are assigned to CRS courses?

2. What is the impact of the FCCRI once schools switch from low to high compliance for all students with eligible FCAT scores (not just those near the thresholds)?

3. How does the impact of the FCCRI differ among all eligible students by baseline achievement and the level of initial college course taken?

Data and sample

Data Sources

Our primary data source for both analyses is student-level records from the Florida K–20 Education Data Warehouse. These data follow all Florida public school students from grade 10 for as long as they remain in Florida's public education system, and are supplemented with school-level variables from the National Center for Education Statistics' Elementary/Secondary

Information System and reports produced by FLDOE. We omitted students who transferred to an out-of-state, private, or home school or withdrew from school for medical reasons in grade 11 or who did not have a school enrollment record in grade 12.

Regression Discontinuity Sample

The RD analysis includes only cohort M1 because the voluntary cohorts had very low participation rates, and tended to be similar on both sides of the test score cutoffs in both the college readiness testing and the CRS courses. We combined test score data with information on race/ethnicity, gender, free or reduced-price lunch (FRPL) status, English language learner (ELL) status, grade 10 GPA, and college outcomes. Casewise deletion of students with missing values gives N=145,580 in math and N=145,754 in English among all students who took the FCAT in grade 10.[2]

We examine two different FCAT cutoffs using two different quasi-experiments and samples. For example, few students near the low FCAT margin score college ready on the PERT, while nearly all of those near the high FCAT margin do. The former group is therefore less likely to succeed in college but more likely to enroll in CRS coursework. As our grade 11 PERT sample consists of students targeted based on FCAT scores, it represents a third quasi-experiment and sample with achievement levels bounded by those of the two FCAT samples. However, students near this cutoff have above-average FCAT scores among targeted students. Our grade 12 PERT sample (a fourth quasi-experiment using a fourth sample) was targeted based on FCAT scores, scored below college ready on the PERT, and then took a CRS course, and therefore tends to be lower-achieving than the overall PERT sample. Additionally, since PERT retesting is not required under the FCCRI, there may be both observed and unobserved differences among students who choose to retest compared to those who do not retest. While

these differences are not huge, they do suggest that retesters are more likely to come from lower in the grade distribution than single-testers, perhaps because those at the top of the grade distribution have concordance scores on another assessment (see table A1).[3] These differences in composition at each cutoff do not invalidate our estimates at any given cutoff; however, we cannot separately identify the effect of different policies at each cutoff from composition effects at each cutoff.

Analyses using grade 11 PERT values as the running variable use only students targeted for PERT testing (N = 89,225 for math, 51,383 in English under casewise deletion).[4] The modal student in these samples comes from the middle of the FCAT targeting range. Students near the bottom of the targeted FCAT range may be less likely to have postsecondary plans, and may therefore have little incentive to comply with assignment to take the PERT, while students near the top of the range may be more likely to have concordance scores on the SAT or ACT that exempt them from PERT testing. Analyses using grade 12 PERT values as the running variable use only students targeted for PERT testing who did not score college-ready on the grade 11 PERT, complied with assignment to a CRS course, and retook the PERT in grade 12 (N = 29,264 for math, N = 10,043 in English under casewise deletion). Figure 1 shows how students progress through the steps of the FCCRI, with sample sizes (without casewise deletion) at each stage.

For the college coursetaking outcomes, we limit the analyses to the subsample of targeted students who are seamless college enrollees so the results are not diluted by students who do not attend college. This should not bias our estimates because the FCCRI did not have an impact on high school graduation or seamless college enrollment. Tables A2 and A3 provide descriptive statistics of the characteristics of students in the math and English subsamples.

Before-After Regression Sample

11

Our sample for the regression analysis started with all targeted students from cohorts V2 through M1. We then dropped students who were not part of the treatment or comparison groups. The analytical sample was reduced further because of missing data. Models were estimated separately for students targeted in math and English. We refer to this as the full sample of students, and it is used for the high school diploma and seamless college enrollment outcomes. For other outcomes, we restricted the sample to students who seamlessly enrolled in college. The full sample was 147,302 in math and 157,646 in English, and the subsample was 69,718 in math and 76,772 in English.

To determine whether the sample was representative of the population it was drawn from, we compared student- and school-level characteristics based on whether the student attended a school that was in or out of B-A analysis. First, we compared student-level characteristics of all targeted students statewide to those who attended schools that were included in the before-after regression analysis (Tables A2 and A3). The students attending schools in the before-after regression sample have slightly higher achievement than students in schools that are out of the analytic sample. There are also small differences in demographic characteristics, but most of the differences are small in magnitude. Second, we compared the characteristics of schools that switched from low- to high-compliance to those of school that did not, and found that most of the mean differences were also small in magnitude (Table A4). The largest differences were in school locale, as the in-sample group had more schools located in suburbs and towns and the out-of-sample group had more schools located in the city and rural areas. These differences in urbanicity may also contribute to the small differences in student achievement, as suburban schools (which are overrepresented in our sample) tend to have higher achievement levels (e.g. Lleras, 2008).

We also compared student and school characteristics in the treatment group versus the comparison group (Tables A5 and A6, respectively). Almost all of the baseline student-level characteristics have a standardized mean difference that is less than 0.05. The largest standardized mean difference was -0.088, for FRPL status. Even though differences between the treatment and comparison group are small, we included all baseline student characteristics as control variables in our regression analysis to ensure that we controlled for any small differences between the treatment and comparison groups. The differences in school characteristics are also small in magnitude. The largest difference is in the percentage of FRPL students, where students in the treatment group attend schools with roughly 5 percent more FRPL students as compared to students in the comparison group.

Methodology

This study uses two methodologies to examine the FCCRI's effects on student outcomes. First, we use regression discontinuity (RD) analysis to estimate the FCCRI's impact on students scoring just above or below thresholds for participation in the FCCRI (RQ 1). The RD analysis uses data for the first cohort of students under the mandatory FCCRI (M1). Second, we use regression analysis with a before-after design and school-level fixed effects to analyze the impact for students from a broader range of baseline achievement (RQs 2-3). The before-after analysis explores changes over time using data from M1 as well as two cohorts of students under the voluntary FCCRI (V2 and V3). The first voluntary cohort, V1, is excluded because it was subject to different math requirements for high school graduation from those for the other cohorts.

Outcome Variables

At the high school level, we examined the probability that students graduated with any type of diploma. At the college level, outcomes included enrollment during the fall semester

immediately following the cohort's on-time graduation (seamless college enrollment) and taking or passing nondevelopmental math or English courses in the first year of college. Math is unique in offering a transition course (intermediate algebra) that counts for elective credit but does not count toward math graduation requirements. Transition courses may be the most relevant outcome in math because the college readiness cutoff on the PERT is the score at which students are recommended for the transition course rather than developmental courses. The regression analysis also includes a categorical outcome for the level of the first course a student takes in college (no course, lower-level developmental education, upper-level developmental education, transitional (math only), or degree credit). Table 1 has a list of outcomes and definitions.

Regression Discontinuity

RD analysis is used when assignment to a policy treatment is determined by whether a continuously valued variable (the "running variable") has crossed a predetermined cutoff; Lee and Lemieux (2010) contains an extremely comprehensive study of its features and requirements. RD analysis has a strong causal interpretation, provided the data meet strict validity requirements; Lee and Lemieux (2010) provide a list of its applications to that point by education economists, showing how it has been applied to a variety of contexts, outcomes, and treatments. Here, RD analysis is used to determine the impact of being assigned to take the PERT in grade 11, enroll in a CRS course in grade 12, and place in developmental coursework prior to college enrollment (since students scoring college-ready on the PERT in grade 12 were exempt from enrolling in developmental education). The state policy did not require students to retake the PERT in grade 12; three districts (Flagler, Gulf, and Hamilton) used the grade 12 PERT as a component of CRS course grades, but incentives and/or requirements for retesting were often determined at the school or teacher levels. Students who retook the PERT were

slightly more likely to be female (in math only), minority, non-native English speakers and/or economically disadvantaged, but these differences all had standardized mean differences of 0.11 or less, far below the level of 0.25 at which the What Works Clearinghouse maintains that baseline equivalence is violated (WWC, 2015).

RD analysis is useful because it isolates the impact of the policy being analyzed without capturing extraneous factors. If assignment to the treatment group is the only thing that changes noticeably at the cutoff for treatment, any difference in student outcomes should be attributable to that treatment. The main drawback of RD is that its results apply only around the cutoff for treatment and are not generalizable to the full sample of students. RD analysis cannot separately analyze individual components of treatment—for example, it cannot differentiate between a true null result and offsetting positive and negative component effects. One case in which this latter results might occur if the benefits from a well-designed CRS course are canceled out by discouragement from being labeled "not college ready." Dougherty (2015) finds in an RD design that African American students may be particularly susceptible to discouragement effects and/or stereotype threat from labels; future work will explore heterogeneous effects of the FCCRI among student subgroups.

Table 2 shows how FCAT performance levels are used to assign students to college readiness testing (in the subject they were targeted in) and how the PERT is used in grade 11 to assign students to CRS courses in grade 12 (and onward) to assign students to developmental education coursework in college. Although the FCAT and PERT assessments group students into broad categories, students also receive scaled scores—between 100 and 500 on the FCAT and between 50 and 150 on the PERT—that function as nearly continuous measures of student

achievement. Using these scores, it is possible to compare students on either side of a cutoff who have extremely similar profiles and differ primarily in their assignment to treatment.

As a score of 300 or higher on each section of the FCAT is required to graduate from high school, the FCAT is the higher-stakes exam of the two; students have an incentive to perform well on the PERT only if they plan to attend a postsecondary institution or if they care strongly about the courses they take in high school. However, as the samples of students near FCAT graduation requirements and near PERT college readiness benchmarks are likely to be quite different, different students might view either exam as higher-stakes in their particular case.

Our estimates use a sharp RD design, which is modeled in a local linear framework as:

$$Y_i = \alpha_1 \widetilde{R}_\iota + \gamma_i * \mathbf{1}\{\widetilde{R}_\iota \geq 0\} + \alpha_2 * \mathbf{1}\{\widetilde{R}_\iota \geq 0\} * \widetilde{R}_\iota + X_i\beta + \varepsilon_i \qquad (1)$$

where Y_i is an outcome of interest for individual i, $\widetilde{R}_\iota = R_i - c$ is the running variable R_i recentered around cutoff c, X_i is a vector of individual characteristics, and ε_i is a mean-zero error term.[5] The parameter of interest in this equation is γ_i, the impact on individual i of being in the treatment group; $E[\hat{\gamma}]$ is the average treatment effect at the cutoff for treatment when compliance is perfect and the intent-to-treat (ITT) effect at the cutoff otherwise. The expectation operator here is crucial, as there is little reason to believe that the FCCRI will have a uniform effect on students, even when accounting for prior achievement, given heterogeneity of both students and implementation. We present an ITT estimator because it most clearly reflects the overall impact a state could expect from implementing a similar program, even under mandatory participation. In our context, the running variable, R_i, is a student's scale score on either the FCAT or the PERT; depending on the regression being run, c may be the scale score cutoff between FCAT proficiency levels or the PERT score required for college readiness in a particular subject area.

The grade 12 PERT is not technically part of the FCCRI, as students on both sides of the college readiness cutoff have been targeted for college readiness testing and taken a CRS course; however including this assessment in the analysis has two advantages. First, it allows us to contrast the effects of targeting students for college readiness testing and coursework against the effects of a known policy with a predictable effect on student outcomes. Second, it allows us to assess the difference between college readiness and college success—to see how many of the students whom CRS courses prepare for college (in the sense of scoring "college-ready") are capable of passing college-level coursework. It does not, however, provide any information about the impact of the FCCRI, as students in both the treatment and control groups participated in the intervention.

Before-After Regression Analysis

To supplement the RD analysis, we used a before-after regression analysis with school-level fixed effects to assess the FCCRI's effects for students from a broader range of achievement levels than the RD design. This is important because the likelihood that students will be successful varies by students' academic achievement, and the FCCRI's ability to help students achieve college readiness likely varies with students' pretreatment achievement levels. For this analysis, variation over time in school-level FCCRI compliance rates was used as an exogenous predictor of treatment takeup. This yielded an analytical sample where treatment takeup was conditionally independent from the outcomes of interest.

We calculated the school-level compliance rate as the proportion of grade 12 students who enrolled in a CRS course after being targeted by the FCAT and not scoring college-ready on the PERT or CPT. We defined the treatment group as students who were targeted by the FCAT and attended a high school with at least a 50 percent compliance rate (a high-compliance school).

17

The comparison group consisted of students who were targeted by the FCAT and attended a school with less than a 5 percent compliance rate (a low-compliance school). We considered high-compliance schools to have implemented the FCCRI, and the low-compliance schools to have not; from this, we obtained a treated group and a comparison group. We limited the analytical sample to students in the same schools that were categorized as both low- and high-compliance at some point between V2 and M1 (low- to high-compliance schools are in the in-sample group), because we were concerned that low-compliance and high-compliance schools might have differences related to the outcomes. We believe that for almost all of the schools that switch from low- to high-compliance, the switch to being in the treatment group reflects an exogenous change – the FCCRI becoming mandatory – as opposed to some other change in the school. Additionally, the inclusion of school level dummy variables addresses any time invariant, unobservable differences in schools. Essentially, this is a before-after analysis, in which we compared student outcomes before and after schools implemented the FCCRI.

One limitation of this approach is that the results are not generalizable to schools that do not switch from low- to high-compliance; however, a large proportion of schools is included in the sample (46 percent in math and 68 percent in reading).Another limitation is that during the first year of college for cohort M1, Senate Bill 1720 (2011) changed the laws on developmental course-taking. Under this bill, beginning in the spring semester of 2013 (the second semester for cohort M1's seamless enrollees), developmental courses were no longer required for recent high school graduates. This change could induce students in cohort M1 to hold off on required developmental courses in the fall because these courses would not be required in the spring. This could incorrectly make the FCCRI appear to reduce developmental course-taking. However, instances of students deferring developmental education in this way are likely uncommon, as the new policy was not well known prior to its implementation. Additionally, other researchers

found little change since fall 2011 in developmental course-taking rates in Florida until the fall

semester of 2014 (Hu et al., 2016); this semester is not included in our evaluation. Thus, this

policy change is not a substantial concern for our analysis, but might become an issue for later

cohorts or for cohort M1's long-term outcomes. We are not aware of any other changes over

time that would threaten the validity of the results for the cohorts examined in our analyses.

We used multiple regression analysis to estimate the ITT effect of the FCCRI for all

targeted students and for the subsample of students who seamlessly enrolled in college. For

binary variables, we used a logit model with maximum likelihood estimation (MLE). For the

categorical outcome, we used a multinomial logit model with MLE. The treatment effect was

obtained by estimating $\hat{\beta}$ as shown in Equation 2 for individual i in school j.

$$Y_{ij} = \hat{\alpha}_i + \hat{\beta} T_i + \boldsymbol{X_i}\widehat{\boldsymbol{\theta}} + \hat{\delta}_j + \mathrm{e}_{ij} \tag{2}$$

The outcome (Y_{ij}) is a function of the treatment status (T_i), a vector of control variables shown

in Table 1 ($\boldsymbol{X_i}$), a school-level fixed effect ($\hat{\delta}_j$), and an error term clustered by the student's

grade 12 school (e_{ij}). Some nonlinear terms are included (see Table 1). Functional form was

chosen based on scatter plots and the significance of such terms when they were included in the

models. All regressors were captured prior to treatment and were therefore unaffected by

treatment. The ITT effect was derived from the estimated coefficient for the treatment

indicator ($\hat{\beta}$). Results are presented as predicted probabilities for ease of interpretation.

Regression Discontinuity Validity

The What Works Clearinghouse (WWC, 2015) has three criteria necessary for a valid RD

design; sharp RD studies must meet four further sets of criteria to meet WWC evidence

standards. To qualify as an RD study, policy treatment must be based on a running variable; this

accurately describes the FCCRI. The running variable must be ordinal, with a sufficient number

of unique values; the FCAT and PERT are ordinally scored, with a large range of possible values above and below each cutoff point. Finally, no other policies may be implemented at the same cutoff value. This is certainly true on a statewide level for the FCAT, as there are no statewide policies uniformly affecting students at the cutoff for (avoiding) targeting. It may not be true at more finely grained levels, as some schools might institute interventions for students in FCAT level 1 (deemed at risk for dropping out of high school); however, there is no clear indication of this in our estimates.[6] The high school graduation cutoff is located sufficiently far from proficiency level cutoffs that bandwidth selectors should avoid any confounding effects. The grade 11 PERT is a more challenging case, as the college readiness cutoff is used to simultaneously inform students whether they are required to take CRS coursework and whether they will be exempt from developmental coursework in college (if they attend). When taken in grade 12, the PERT affects placement into college developmental coursework but should not affect high school course selection.

The first validity requirement for RD is that the running variable must be immune to manipulation. Manipulation requires that students know the cutoffs for treatment, have incentives to change their running variable values, and have the means to do so. As the FCAT and PERT are administered and scored by independent contractors with no known incentive to modify scores or treatment statuses, manipulation within a single exam sitting is unlikely.

Non-random retesting may be a greater threat to validity, however. Students may retake both the FCAT (if they score below a scaled score of 300, which is required for graduation) and the PERT; this could lead to selective retesting on the two assessments to avoid FCCRI requirements. If so, our estimates might capture the effects of student motivation as much as of the FCCRI. Furthermore, some students may benefits more than others from retesting due to

additional home or school supports (see, for example, Papay, Murnane, and Willett, 2010). To avoid this, we used students' initial grade 11 PERT scores on each section; while students' highest scores are more likely to affect their placement into or out of CRS courses, initial scores cannot be manipulated through selective retesting.[7] Since our data do not include FCAT dates, we instead use students' lowest FCAT scores as proxies for their initial scores.

Further evidence on manipulation via retesting is presented in McCrary density tests in Figure 2, which show the number of students at each FCAT or PERT score. Bunching just above or below any of the policy cutoffs would indicate that students are systematically working to avoid a particular policy outcome. The clearest signs of bunching are at FCAT scores of 300, which represent the cutoff for high school graduation. However, this cutoff is not associated with targeting for the FCCRI, and is sufficiently far from the FCAT level 2 cutoff that students' efforts to graduate should not be mistaken for attempts to clear the FCAT level 2 threshold

A more troubling case of bunching is at the cutoff for college readiness in the grade 11 PERT, where there appears to be substantial bunching in the number of students who score college-ready in reading. Although some of this discontinuity may be due to wide variations in pass rates just to the left of the college readiness cutoff, estimates using the grade 11 reading PERT should be taken with caution.

The second validity requirement is that there cannot be excessive attrition overall or by treatment status. Within a narrow bandwidth of all cutoffs, both average and differential attrition are sufficiently low to avoid introducing substantial bias. Overall attrition in both subjects is less than 10 percent at all cutoffs (under five percent when the low FCAT cutoff is omitted) and the difference between treatment and control groups is less than 5 percentage points at all cutoffs (four percent for the low FCAT margin in reading, and one percent or less for all other cutoffs).

21

Given an overall attrition rate of 10%, WWC standards permit up to 6.3 percentage points of differential attrition under the most stringent set of criteria. As a result, our data are well within acceptable boundaries for attrition.

The third RD validity requirement is that outcome variables be continuous everywhere but at the policy cutoff and that the outcome variable does not have any unexplainable discontinuities. We check for baseline equivalence of FCAT scores, cumulative high school GPA, and FRPL status and find effect sizes ranged from 0.000 to 0.0939 standard deviations, with the majority less than 0.05 (Table A7). Since no effect sizes were greater than 0.25, these tests do not invalidate our estimates. However, because some effect sizes were greater than 0.05, our estimates control for a full set of covariates in Table 1. Removing covariates does not significantly change our estimates; while some magnitudes change, only one specification has a different significance level when covariates are removed (results without covariates are available upon request).

The second approach for examining the continuity of the outcome-running variable relationship is to demonstrate that outcome variables are continuous or explainably discontinuous away from any cutoffs. Explainable discontinuities arise at the FCAT cutoff for high school graduation and the PERT cutoff for degree credit coursework in math. Discontinuities caused by either of these cases are well explained and sufficiently far from relevant cutoffs that bandwidth selectors will avoid them if needed.

The final validity requirement is that RD estimation must use an appropriate functional form and/or bandwidth. We use local linear estimation, with bandwidths selected using a cross-validation method initially presented by Imbens and Lemieux (2008) and further explicated by Lee and Lemieux (2010). This strategy process uses a similar principle as the RD analysis itself,

but applied to a broader range of values. RD analysis, at its crux, examines the difference between the predicted value of an outcome variable at the cutoff value and the actual (average) value at that cutoff. Cross-validation applies this logic to each point within a particular range of the cutoff in order to determine the bandwidth that contains the least amount of statistical noise.

To ensure that our bandwidth selector is not confounded by the high school graduation cutoff (near the low FCAT margin) or the cutoff for degree credit eligibility (near the PERT college readiness cutoff), we use two strategies as robustness checks. The first strategy is to cap bandwidth selectors so that bandwidths cannot contain the second cutoff value. The second strategy is to use an estimator designed by Calonico, Cattaneo, and Titiunik (2014) without providing any boundaries to the bandwidth (our main strategy considers bandwidths between 5 and 20 points on the assessment in question). This estimator converged only for the FCAT analyses and only without covariates, but selected quite large bandwidths, going far beyond the other cutoff values. Neither set of estimators produced large, statistically significant results; (available upon request).[8] Our bandwidth selector therefore does not appear to be influenced by known policy cutoffs near those that we are analyzing. Furthermore, to be certain that variation in optimally selected bandwidths does not have an impact on our results, we run all specifications using bandwidths of five, ten, and twenty points and using cross-validated bandwidths under a local quadratic regression and include the results in an appendix (tables A8-A11); these results are broadly similar to those presented in the results section.

Results

Regression Discontinuity Results

We present several sets of estimates for the regression discontinuity analysis that cover the following sets of effects:

- The intent to treat with PERT testing, based on FCAT score (using both upper- and lower-level cutoffs).

- The intent to treat with CRS courses, based on grade 11 PERT score, conditional on being targeted based on FCAT scores.

- The intent to treat with placement into college developmental college coursework, based on grade 12 PERT score, conditional on being targeted based on FCAT scores, scoring below college-ready on the grade 11 PERT, and enrolling in a CRS course.

Conditional statements for both grade 11 and grade 12 PERT apply to all results, which are shown in Table 3. Each row in the table shows an assessment (and cutoff, when the assessment in question is the FCAT), while each column refers to a different outcome. Readers should recall each cutoff represents a distinct quasi-experiment, and that differences in samples across experiments may help in interpreting some results.

Columns (1) and (2) show that no stage of the FCCRI had a statistically significant effect on high school graduation or on seamless college enrollment. Since the FCCRI did not impact either of these factors, we focus the remainder of both the RD and before-after analysis on the subset of students who seamlessly enrolled in college. Graphical analysis of the remaining outcomes is shown in Figure 3. On each graph that makes up Figure 3, the x axis represents the range of scores on the exam in question (with solid lines representing the cutoffs between FCAT proficiency levels and the PERT college readiness level, and dashed lines representing the FCAT high school graduation requirement), and the y axis represents the percent of students meeting the outcome in question. Markers on each graph represent the average outcome value at each FCAT or PERT score, and the lines on either side of each cutoff being evaluated are local linearizations based on average outcome values. The difference in local linearizations at the

24

cutoff being evaluated is equivalent to the point estimate of a local linear RD regression with no coefficients.

In math, FCCRI targeting at the low FCAT cutoff had no impact on whether students enrolled in (column (3)) or passed (column (4)) nondevelopmental courses at either FCAT cutoff. The bottom half of Table 3, on the other hand, shows a coefficient of 0.0425 for students targeted at the low FCAT cutoff in English, meaning that these students were 4.3 percentage points (from a baseline of 59.2 percent) more likely to enroll in a nondevelopmental English course but no more likely to pass. This suggests that those students who were pushed to take nondevelopmental English may not have been prepared to do so. At the high FCAT cutoff, students who were targeted for the FCCRI were no more likely to enroll in or pass a nondevelopmental credit English course.

Students targeted for CRS courses in math at the college readiness cutoff on the grade 11 PERT were no more likely than nontargeted students to enroll in or pass a nondevelopmental math course. Students at the college readiness cutoff in English were 2.3 percentage points (from a baseline of 79.1 percent) less likely to enroll in a nondevelopmental English course but no less likely to pass.

For comparison, college readiness on the grade 12 PERT had a large impact on student course-taking. Students at the college readiness cutoff in math who were exempted from developmental coursework were 30.6 percentage points (from a baseline of 45.9 percent) more likely to take a nondevelopmental course than students who were not exempted. However, these students were only 18.1 percentage points (from a baseline of 27.9 percent) more likely to pass, meaning that over 40 percent of students induced to take nondevelopmental coursework were unable to pass. In English, students at the college readiness cutoff were 16.8 percentage points

(from a baseline of 54.4 percent) more likely to enroll in nondevelopmental coursework and 13.5 percentage points (from a baseline of 40.6 percent) more likely to pass; nearly 20 percent of students induced to enroll in nondevelopmental coursework were therefore unable to pass.

One important finding is that students targeted for CRS courses performed comparably to students who were already college-ready in grade 11. Based on performance in grade 11, all students who were below college-ready had scores corresponding to developmental courses, while all students who were above college-ready were exempt from developmental education. Thus, if the students' performance remained the same between grade 11 and college enrollment, we would expect those scoring just below college-ready to be much more likely to be placed into developmental education courses. Yet in both subjects, there was little to no difference in the likelihood of passing nondevelopmental courses. This stands in contrast to the results for the grade 12 PERT, where students just below college-ready are significantly less like to enroll in and pass non-developmental courses than students just above college-ready in grade 12. However, it is not possible to know how many students in CRS courses would have enrolled in and passed nondevelopmental courses if they had taken another course in grade 12 instead.

Results from the Grade 12 PERT help put the null results elsewhere into context. We find that seamless college enrollees who were targeted for college readiness testing, did not score college ready, took a college readiness course, and passed the PERT in grade 12 enrolled in for-credit courses at a rate very similar to that of seamless college enrollees who passed the PERT on their first attempt. While we cannot separately identify the effects of discouragement from being labeled "not college ready," any psychological effects from taking a course that advertises college readiness, and any human capital gained from the course, it is nonetheless clear that the grade 12 PERT is the main factor affecting many students' decisions to enroll in DE or in a for-

credit course. Since the RD construction implies that students narrowly on either side of the college readiness cutoff in grade 12 are of nearly identical ability, we also conclude that the grade 12 PERT is likely underplacing approximately 18 percent of students near the math cutoff and 14 percent near the reading cutoff whom we might expect to pass. (Due to the RD construction, we of course cannot say whether this rate would apply to students far from the cutoff. Future work will investigate this topic more thoroughly.) However, we also find that students who score college ready in grade 12 pass for-credit courses in both subjects at slightly lower rates than those who score college ready in grade 11. This is to be expected, as students in our grade 12 sample are likely to be somewhat less proficient than those in our grade 11 sample.

These results for students assigned to CRS courses are particularly striking when considering the counterfactual conditions. On the grade 11 PERT, there is evidence that students who barely scored college-ready were much more likely than those who did not to enroll in advanced coursework such as honors and college-credit level courses during grade 12, while those who scored just below this cutoff were much more likely to enroll in standard-level coursework, including CRS courses. This suggests that students assigned to CRS courses were being compared with a very high standard and actually fared quite well by demonstrating similar performance in college.

<center>*Before-After Regression Results*</center>

We were concerned the RD analysis might not pick up on the full treatment effect because of its limited analytical sample. Thus, we supplemented the RD results with the before-after analysis in three ways. First, the before-after analysis is capable of including students in the middle of the targeted range of students, as opposed to including only students near the cutoffs. Thus, we estimated the same set of outcomes as the RD analysis to see how the results change

<center>27</center>

when we include a wider range of students (Table 4). As a sensitivity analysis, we also restricted the treatment group to include students from M1 only since the RD analysis is limited to M1. The estimates from this sensitivity analysis were all within 0.001, so the differences between the RD and before-after estimates are likely from differences in the range of baseline achievement included in the sample as opposed to differences in which cohorts are included in the treatment group.

Other than the RD results that relied on the Grade 12 PERT cutoff, the results from the before-after analysis point towards a larger treatment effect for the enrollment in and passing nondevelopmental course outcomes. All of the results for the enrollment and passing nondevelopmental course outcomes are positive and significant in both math and English, whereas most of these estimates were negative and/or insignificant in the RD analysis.

Second, because the before-after analysis includes a range of students in the treatment group, we were able to include a breakdown of the results by baseline achievement, as measured by the continuous variable for grade 10 FCAT score, to determine if the average marginal effect is hiding variation in the treatment effect across baseline achievement. The results are presented in Figure 4, which contains predicted probabilities of enrolling in each course level by treatment status and baseline achievement. The difference between the two lines is the marginal effect of the intent to treat. There does appear to be variation in the impact across baseline achievement. Figure 4 shows that the FCCRI had a larger impact for students in the lower and middle range of baseline achievement. The largest difference across treatment status in math was for students near the FCAT level 3 cutoff. In English, the largest difference was for students at the bottom of the targeted range.

Third, the before-after analysis allowed us to consider how enrollment in nondevelopmental courses changed across the full spectrum of course levels through the use of a multinomial outcome, as opposed to simply comparing nondevelopmental and developmental coursetaking. We present results in Figures 5 and 6 by plotting predicted probabilities of enrolling in each course level by treatment status and baseline achievement, where the difference between the two lines is the marginal effect of the intent to treat. The multinomial results indicate that the FCCRI reduced enrollment in both lower- and upper-level developmental education in math. The impact was especially large for students at the lower end of the targeted range (FCAT level 2), where the treatment group was up to 12.6 percentage points less likely to enroll in lower-level developmental education courses. Among treatment students in the middle of the targeted range (FCAT level 3), there was a 10.7 percentage point increase in the likelihood of enrolling in a nondevelopmental course, with most of the increase occurring in the transitional course. These students moved away from both lower- and upper-level developmental education courses at similar rates. Developmental course enrollment was also lower for the treatment group in English, though differences were smaller than in math. The main change seems to be in moving lower-performing students from upper-level developmental education to degree credit courses, as treatment students had up to a 5.8 percentage point decrease in the likelihood of enrolling in upper-level developmental courses. There was little to no difference for higher-performing targeted students.

Overall, the results from the before-after analysis provided stronger evidence of a positive impact of the FCCRI than the regression discontinuity analysis. The results of the two lines of analysis likely differ because 1) the RD analysis reduced the analytical sample to students near cutoff points and 2) the treatment rates differ across the RD and before-after

samples, with the before-after sample having larger differences in treatment rates between the treatment and comparison groups (refer to Table A12 of the appendix for a summary of the treatment rates by sample).

Discussion

This study sought to examine the impact of the FCCRI on student outcomes by the first year of college and found mixed results. The FCCRI increased the likelihood of enrolling in nondevelopmental courses for some targeted students, although results differed by academic performance. However, smaller differences in the likelihood of passing nondevelopmental courses suggest that some students were not prepared for these courses. Additionally, impacts were estimated after only the first year of implementation of the mandatory FCCRI, and program impacts may continue to grow as schools learn how to better deliver the CRS courses over time.

In 2015, the legislature ended the requirement that high schools offer the college readiness testing and CRS courses. There is no simple answer as to whether this was a good decision. Some higher-performing students may be better off without the requirement to take CRS courses if they would otherwise be taking more rigorous courses. Yet other mid-performing students may be harmed by the lack of access to the college readiness testing and the CRS courses. Separate legislation in 2014 exempts recent high school graduates from taking the PERT and enrolling in developmental education courses at state colleges. This means students would have no test scores to indicate their level of college readiness in high school or college, and many students may enroll in nondevelopmental courses without knowing they are not college-ready.

Implications

This evaluation has important implications for similar programs in other states, as initiatives like the FCCRI have gained popularity over time. A national scan conducted in 2012 found that 29 states offer transitional math and English courses during the senior year of high

school for students who have not previously met college readiness benchmarks, although some of these are local policies rather than statewide policies like the FCCRI (Barnett et al., 2013).

Our findings suggest that initiatives like the FCCRI may not be effective for improving high school graduation rates or college enrollment. The FCCRI had no effect on these outcomes for students at any level of performance on the FCAT or PERT, rejecting our hypothesis that the FCCRI might encourage students to complete high school and continue to postsecondary education by showing them that they can obtain the skills needed for college. These findings corroborate feedback from educators that the FCCRI seems to be most effective for students who want to attend college but are not quite college-ready and least effective for students who are disengaged from school and lack realistic postsecondary goals (Lansing, Ahearn, Rosenbaum, Mokher, & Jacobson, 2017; Mokher et al., 2014). While it may not be practical to limit CRS courses to students who indicate they are college-bound in grade 11, particularly since high school students' college intent is subject to change, the findings suggest that states looking to improve high school graduation and college enrollment rates should consider other policies.

Yet there is some evidence that the FCCRI may reduce enrollment in developmental education among students who do attend college, particularly lower- to mid-performing students. This is similar to findings in California where students who participated in the statewide EAP program had a lower probability of needing developmental education courses at California State University, particularly those at the margins of remediation risk (Kurlaender et al. 2016). Other states may want to consider this type of initiative if they have similar goals.

Our study also has implications about the types of students who should be targeted by such policies. The FCCRI had different impacts based on students' prior academic achievement. The initiative appears to have most helped students who were neither too far behind to catch up

31

in a single year nor so advanced that they were already college-ready. Other studies have found that targeting higher performing students may even be harmful –an evaluation of a similar policy in West Virginia found negative effects which may be attributed to students taking transition courses at the expense of more rigorous high school courses (Pheatt, Trimble, & Barnett, 2016). States should collect feedback from educators and look at their own data to identify the students who are most likely to benefit and use this information when devising eligibility criteria.

Finally, given state policymaker' interest in improving college-readiness, more effort should be focused on findings ways for high schools and colleges to work together. While FLDOE advised high schools to work with local community colleges to develop the curriculum for the CRS courses, there is very little evidence of secondary-postsecondary collaboration around the CRS courses or college readiness more broadly (Mokher & Jacobson, under review). To support students who are not yet college-ready, states should work to ensure that students have early information about their level of college readiness, appropriate courses in high school to better prepare them for college-level work, and effective remediation options and supports in college.

<center>*Directions for Future Research*</center>

This study also provides several directions for future research, which we are currently investigating. First, what are the impacts of the FCCRI on longer-term outcomes such as college persistence and degree completion? If more students enroll directly into nondevelopmental courses but pass rates are declining, this could have negative implications if students become discouraged and drop out of college, or if their time to degree increases. Second, how do the impacts differ by student and school characteristics? We may expect to find variation in impacts across schools, and the FCCRI may also change achievement gaps in postsecondary outcomes

<center>32</center>

across student subgroups such as FRPL status. Future research could also use additional methods to extrapolate the regression discontinuity impacts away from the cutoffs to further explore variation by student achievement (e.g. Angrist & Rokkanen, 2015; Tang, Cook, Kisbu-Sakarya, Hock, & Chiang, 2017). Third, what impact did the FCCRI have on student coursetaking in grade 12? We are examining the extent to which students took CRS courses at the expense of more rigorous high school courses, and how these results differ by student achievement. Lastly, given that recent high school graduates are no longer required to take the PERT or enroll in developmental education courses in college, is information from students' high school records a sufficient substitute for PERT? We will examine the extent to which students are misplaced into college-level courses, and compare how high school data perform relative to the PERT in predicting student success in college.

Notes

1. While compliance with treatment or control status is imperfect, we use a sharp model to estimate the intent-to-treat effect rather than a fuzzy model to estimate a complier average treatment effect. This allows us to focus on expected compliance and outcomes from a policymaker's perspective.

2. Our initial math sample consists of 158,158 students, after dropping duplicate observations, those with invalid FCAT scores, those who attrited, and those with postsecondary award dates prior to high school graduation or after the data collection window. Of the rest, 947 (0.6%) were dropped from our math analyses due to missing demographic data and another 11,631 (7.4%) due to missing GPA. Corresponding numbers for English were 158,478 in the initial sample, with 953 (0.6%) dropped due to missing demographic data and another 11,780 (7.4%) due to missing GPA data. While students who are missing GPA data have lower FCAT scores on average, GPA is an important measure of ability and predictor of success, and therefore should be included in our regression results. Fewer than one percent of students are dropped due to missing demographic data (the vast majority of whom are missing ELL data); this small number of students can be omitted without major impacts to our results.

3. Some highly motivated students might forgo retesting to take another assessment (such as the SAT or ACT) that could provide a concordance score. Current data do not allow us to determine which students these are or how this might impact our estimates.

4. We do not consider narrowing the sample in this way to be attrition for several reasons. First, the omitted students (PERT-takers not targeted for the PERT) are not policy-relevant, as scoring below college ready would not obligate them to take CRS courses. Second, they are likely to be systematically different from targeted PERT-takers—for example, students below the targeting range might be particularly motivated. Finally,

when the running variable, cutoff, and policy implication all change, narrowing the sample to reflect the individuals for whom these are relevant is appropriate and may be necessary to properly address the impact of the policy in question.

5. If students below the cutoff are assigned to treatment, the equation should be modified either by changing $\widetilde{R}_i \geq 0$ to $\widetilde{R}_i \leq 0$ or by changing $\widetilde{R}_i = R_i - c$ to $\widetilde{R}_i = c - R_i$ (these are functionally equivalent). This is used for the high FCAT cutoff and the grade 11 PERT cutoff; as being below a target score means assignment to elements of the FCCRI. Graphs of our RD results do not contain additional covariates, in order to better illustrate local linearization based on FCAT scores; in practice, effects with and without covariates are extremely similar.

6. It is impossible to rule out that some teachers or guidance counselors might use FCAT or PERT scores in advising students to take particular courses, but preliminary analyses of senior year course-taking patterns do not suggest that the FCAT or PERT are being used on a large scale to sort students into particular courses except through the FCCRI.

7. If disproportionately many individuals appear just above or below an RD cutoff, it may be because individuals are manipulating their running variable levels to participate in what they believe to be a particularly valuable treatment (or to avoid a particularly unpleasant treatment). A discontinuity in the density of the running variable (due to individuals bunching on one side of the cutoff) would therefore indicate that the treatment is not being applied at random and invalidate the RD design.

8. According to both estimators, the ITT effect on taking non-DE English at the low FCAT margin is statistically insignificant. Using the narrow estimator, the ITT effect on both taking and passing non-DE math at the low FCAT margin is statistically significant, but

40-50 percent smaller than the (already small) statistically insignificant effects from our

initial bandwidth.

References

Achieve. (2012). *Closing the expectations gap: 50-state progress report on the alignment of K-12 policies and practice with the demands of college and careers.* Retrieved from http://www.achieve.org/files/Achieve201250StateReport.pdf

Achieve. (2016). *The college and career readiness of U.S. high school graduates.* Retrieved from http://www.achieve.org/files/CCRHSGrads-March2016.pdf

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college.* Retrieved from U.S. Department of Education website: https://www2.ed.gov/rschstat/ research/pubs/toolboxrevisit/index.html

Angrist, J. D., & Rokkanen, M. (2015). Wanna get away? Regression discontinuity estimation of exam school effects away from the cutoff. *Journal of the American Statistical Association*, *110*(512), 1331-1344.

Attewell, P., & Domina, T. (2008) Raising the bar: Curricular intensity and academic performance. *Educational Evaluation and Policy Analysis, 30*(1), 51–71. doi: 10.3102/0162373707313409

Barnett, E., Fay, M., Bork, R., & Weiss, M. (2013). *Reshaping the college transition: States that offer early college readiness assessments and transition curricula.* Retrieved from Columbia University, Community College Research Center website: http://ccrc.tc.columbia.edu/media/k2/attachments/reshaping-the-college-transition-state-scan.pdf

Bettinger, E. P., & Long, B.T. (2009). Addressing the needs of underprepared students in higher education: Does college remediation work? *The Journal of Human Resources, 44*(3), 736–771. doi:10.3368/jhr.44.3.736

Boser, U., & Burd, S. (2009). *Bridging the gap: How to strengthen the PK-16 pipeline to improve college readiness.* Retrieved from New America website: https://static.newamerica.org/attachments/2778-bridging-the-gap/NAF_Bridging_the_Gap.adcd880132044bb480238492c1871b61.pdf

Calonico, S., Cattaneo, M., & Titiunik, R. (2014) Robust data-driven inference in the regression-discontinuity design. *The Stata Journal.* 14(4), 909-946.

Center for Community College Student Engagement. (2016). *Expectations meet reality: The underprepared student and community colleges.* Retrieved from http://www.ccsse.org/docs/Underprepared_Student.pdf

Dougherty, S. M. (2015). Bridging the discontinuity in adolescent literacy? Mixed evidence from a middle grades intervention. *Education Finance and Policy, 10*(2), 157-192. doi:10.1162/ EDFP_a_00157

House Bill 1255, Fla. Stat. § 1008.30 (2011)

House Bill 7069, Fla. Stat. § 1008.30 (2015)

Hu, S., Park, T. J., Woods, C. S., Tandberg, D. A., Richard, K., & Hankerson, D. (2016). *Investigating developmental and college-level course enrollment and passing before and after Florida's developmental education reform* (REL 2017–203). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee/edlabs

Hurwitz, M., Smith, J., Niu, S., & Howell, J. (2015) The Maine question: How is 4-year college enrollment affected by mandatory college entrance exams? *Educational Evaluation and Policy Analysis 37*(1), 138–159. doi: 10.3102/0162373714521866

Imbens, G., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics 142*(2), 615–635. doi: 10.3386/w13039

Kurlaender, M., Jackson, J., Grodsky, E., & Howell, J. (2016). *Ready or Not? California's Early Assessment Program and the Transition to College.* Paper session presented at the meeting of American Educational Research Association, Washington, DC.

Lansing, J., Ahearn, C., Rosenbaum, J., Mokher, C.G., & Jacobson, L. (2017). Improving high school-college alignment: A sociological reform and new challenges. *The Educational Forum, 81*(3):267-280.

Lee, D., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature 48*(2), 281–335. doi: 10.1237/jel.48.2.281

Lleras, C. (2008). Race, racial concentration, and the dynamics of educational inequality across urban and suburban schools. *American Educational Research Journal, 45*(4), 886-912.

Long, M. C., Conger, D., & Iatarola, P. (2012). Effects of high school course-taking on secondary and postsecondary success. *American Educational Research Journal*, *49*(2), 285–322. doi: 10.3102/0002831211431952

Martorell, P., & McFarlin, I. (2011). Help or hindrance? The effects of college remediation on academic and labor market outcomes. *The Review of Economics and Statistics, 93*(2), 436–454. doi: 10.1162/REST_a_00098

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics, 142*(2), 698–714. doi:10.1016/j.jeconom.2007.05.005

Mokher, C.G. & Jacobson, L. (under review). A Partnership Model Approach to Understanding Challenges in Collaboration around College Readiness. *American Educational Research Journal.*

Mokher, C.G., Jacobson, Ahearn, C., Atkin, J., Flory, M., Lansing, J., Pearson, J., Rosenbaum, J. & Sun, C. (2014). *Assessment of the Florida College and Career Readiness Initiative: Year 2 report.* Arlington, VA: CNA Corporation. Retrieved from https://www.cna.org/CNA_files/PDF/IRM-2014-U-009236-Final.pdf

Mokher, C.G., Jacobson, L., Rosenbaum, J., & Lalonde, R. (2013). *Assessment of the Florida College and Career Readiness Initiative: Year 1 Report.* Report prepared for the Florida Department of Education by CNA. http://www.cna.org/research/2013/assessment-florida-college-career-readiness

No Child Left Behind (NCLB) Act of 2001, 20 U.S.C.A. § 6301 et seq. (West 2003).

Papay, J. P., Murnane, R. J., and Willett, J. B. (2010). The Consequences of High School Exit Examinations for Low-Performing Urban Students: Evidence from Massachusetts. *Educational Evaluation and Policy Analysis, 32*(1), 5-23. doi: 10.3102/0162373709352530

Phaett, L., Trimble, M., & Barnett, E. (2016). *Improving the transition to college: Estimating the impact of high school transition courses on short-term college outcomes.* New York, NY: Community College Research Center. Working Paper No. 86. Retrieved from http://ccrc.tc.columbia.edu/media/k2/attachments/improving-transition-estimating-impact-transition-courses_R2.pdf

Senate Bill 1908, Fla. Stat. § 1008.30 (2008).

Scott-Clayton, J., Crosta, P. M., & Belfield, C. R. (2014). Improving the targeting of treatment: Evidence from college remediation. *Educational evaluation and policy analysis, 36*(3) 371–393. doi: 10.3102/0162373713517935

Strong American Schools. (2008). *Diploma to nowhere.* Retrieved from https://www.scribd.com/document/8534051/Diploma-To-Nowhere-Strong-American-Schools-2008

Tang, Y., Cook, T. D., Kisbu-Sakarya, Y., Hock, H., & Chiang, H. (2017). The Comparative Regression Discontinuity (CRD) Design: An Overview and Demonstration of its Performance Relative to Basic RD and the Randomized Experiment. In *Regression Discontinuity Designs: Theory and Applications* (pp. 237-279). Emerald Publishing Limited.

What Works Clearinghouse. (2015). *Procedures and Standards Handbook Version 3.0.* Retrieved from: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf

Table 1

Variable Overview

Variables	Variable description
	Outcome variables
High school diploma	Binary variable equal to one if the student earned a high school diploma on time with their grade 11 cohort.
Seamless college enrollment	Binary variable equal to one if the student enrolled in college the fall semester following their cohort's high school graduation.
First math/English course level	Categorical outcome that captures the level of the first course a student takes in college: no course, lower-level DE, upper-level DE, transitional (math only), or degree credit (regression analysis only).
Transitional/degree credit enrollment	Binary variable equal to one if the student enrolled first in a transitional or degree credit course (math only).
Transitional/degree credit pass	Binary variable equal to one if the student enrolled first in a transitional or degree credit course and passed (math only).
Degree credit enrollment	Binary variable equal to one if the student enrolled in a degree credit math/English course first.
Degree credit pass	Binary variable equal to one if the student enrolled in a degree credit math/English course first and passed.
	Control variables
Student background characteristics	Race indicators Free and reduced-price lunch status English language learner status Native language Special education status Gifted and talented status Gender Cumulative high school grade point average (through grade 10)
Pretreatment achievement[a]	Grade 10 FCAT score in math Grade 10 FCAT score in reading

Note. DE = developmental education; FCAT = Florida Comprehensive Assessment Test. For college course outcomes, if no course was taken during the first year, dual enrollment and Advanced Placement courses are considered.

[a] Pretreatment achievement variables are used only as control variables in the before-after regression analysis. We also included squared terms for both FCAT subjects and an interaction term between the intent to treat and the FCAT score for the subject of interest (e.g., for the math estimates we interact the intent to treat with the math FCAT score).

Table 2

Assignment to PERT, CRS Courses, and Remedial Coursework

Score category	Test subject	
	Math	Reading
FCAT		
Level 1	Scores 100–286 At risk for failing to graduate Not assigned to take PERT	Scores 100–286 At risk for failing to graduate Not assigned to take PERT
Level 2	Scores 287–314 Eligible for graduation at 300 Assigned to take PERT	Scores 287–326 Eligible for graduation at 300 Assigned to take PERT
Level 3	Scores 315–339 Assigned to take PERT	Scores 327–354 Assigned to take PERT
Level 4	Scores 340–374 Assigned to take PERT	Scores 355–371 Assumed to be college-ready Not assigned to take PERT
Level 5	Scores 375–500 Assumed to be college-ready Not assigned to take PERT	Scores 372–500 Assumed to be college-ready Not assigned to take PERT
Grade 11 PERT		
Not college-ready	Scores 50–112 Assigned to CRS course	Scores 50–103 Assigned to CRS course
College-ready	Scores 113–150 Not assigned to CRS course	Scores 104–150 Not assigned to CRS course
Grade 12 PERT		
Not college-ready	Scores 50–112 Assigned to developmental coursework	Scores 50–103 Not assigned to developmental coursework
College-ready	Scores 113–150 Not assigned to developmental coursework	Scores 104–150 Not assigned to developmental coursework

Note. FCAT = Florida Comprehensive Assessment Test; PERT = Postsecondary Education Readiness Test; CRS= college readiness and success.

Table 3

RD Estimated Intent to Treat Effect, Local Linear Regression with Optimal Bandwidths

Model Version	Outcome			
	(1)	(2)	(3)	(4)
	Any High School Diploma	Seamless College Enrollment	Nondevelopmental Enrollment	Nondevelopmental Pass
	Math			
FCAT, Low Margin	-0.0008	0.0076	-0.0290	-0.0212
	(0.0097)	(0.0126)	(0.0194)	(0.0155)
	[bw = 20]	[bw = 20]	[bw = 19]	[bw = 16]
FCAT, High Margin	-0.0011	0.0053	-0.0252	-0.0239
	(0.0035)	(0.0122)	(0.0112)	(0.0164)
	[bw = 10]	[bw = 10]	[bw = 13]	[bw = 14]
Grade 11 PERT	-0.0024	-0.0010	0.0100	-0.0042
	(0.0033)	(0.0059)	(0.0098)	(0.0140)
	[bw = 8]	[bw = 19]	[bw = 12]	[bw = 13]
Grade 12 PERT	0.0001	0.0011	0.3057 ***	0.1812 ***
	(0.0024)	(0.0114)	(0.0233)	(0.0202)
	[bw = 18]	[bw = 19]	[bw = 14]	[bw = 16]
	English			
FCAT, Low Margin	-0.0027	0.0008	0.0425 **	0.0218
	(0.0052)	(0.093)	(0.0191)	(0.0169)
	[bw = 15]	[bw = 15]	[bw = 20]	[bw = 20]
FCAT, High Margin	-0.0004	0.0077	0.0048	-0.0036
	(0.0038)	(0.0102)	(0.0081)	(0.0096)
	[bw = 19]	[bw = 19]	[bw = 18]	[bw = 19]
Grade 11 PERT	-0.0011	-0.0029	-0.0225 *	-0.0025
	(0.0025)	(0.0087)	(0.0125)	(0.0098)
	[bw = 20]	[bw = 18]	[bw = 18]	[bw = 20]
Grade 12 PERT	-0.0014	0.0088	0.1676 ***	0.1352 ***
	(0.0042)	(0.0218)	(0.0418)	(0.0268)
	[bw = 18]	[bw = 18]	[bw = 7]	[bw = 7]
Sample	Students within bw	Students within bw	Seamless college enrollees within bw	Seamless college enrollees within bw

Note. FCAT = Florida Comprehensive Assessment Test; bw = bandwidth; PERT = Postsecondary Education Readiness Test. Results are for cohort M1. Results shown are for local linear regression with bandwidths optimally determined on a regression-by-regression basis via cross-validation. Numbers reported are the difference in predicted probabilities across treatment status. Standard errors are shown in parentheses, while bandwidths are listed in brackets. Regressions control for gender, free or reduced-price lunch status, race/ethnicity, English language learner status, native English speaker status, disability status, gifted/talented status, cumulative GPA as of grade 10, and district indicators. N varies by specification; for FCAT N ranges from 4,607 to 32,810, for grade 11 PERT N ranges from 22,835 to 67,630, for grade 12 PERT N ranges from 2,246 to 25,552.
*p<.10. **p<.05. ***p<.01.

Table 4
Before-After Regression Estimated Marginal Effect of Treatment

Targeted Subject	Outcome			
	(1)	(2)	(3)	(4)
	High school diploma	Seamless college enrollment	Nondevelopmental enrollment	Nondevelopmental pass
Math	0.001	0.002	0.057***	0.035***
	(0.002)	(0.003)	(0.005)	(0.005)
English	0.001	0.002	0.043***	0.035***
	(0.002)	(0.003)	(0.004)	(0.004)
Sample[a]	All targeted students	All targeted students	Targeted students who seamlessly enroll in college	Targeted students who seamlessly enroll in college

Note. Numbers reported are the difference in predicted probabilities across intent-to-treat status. Standard errors are shown in parentheses below. Results are for cohorts V2, V3, and M1. Models followed a logit specification and included student background characteristics, pre-treatment achievement, and a school-level fixed effect as regressors.

[a]All students in the sample must meet the definition of either the treatment or comparison group as discussed in the methodology section. Targeting is determined by the grade 10 FCAT. The sample size for all targeted students is 147,302 in math and157,646 in English, and for the subsample of students who seamlessly enroll in college the sample size is 69,718 in math and 76,772 in English.

*p<.10. **p<.05. ***p<.01.

Figure 1

*Number and percent of students at each stage of the FCCRI progression*

Full Math Sample | Targeted for PERT? | Took PERT? | Targeted for CRS? | Took CRS?

All FCAT-takers [N=166,677]

No; too low (FCAT 100 - 286) [N=18,503; 11.1%]

Yes; targeted (FCAT 287 - 374) [N=131,963; 79.2%]

No; too high (FCAT 375 - 500) [N=16,211; 9.7%]

No [N=36,656; 22.0%]

Yes [N=95,307; 57.2%]

No; college-ready (PERT 113 - 150) [N=24,313; 14.6%]

Yes; below ready (PERT 50 - 112) [N=70,994; 42.6%]

Yes; took CRS [N=43,272; 26.0%]

No; other course [N=19,701; 11.8%]

No; no math [N=8,021; 4.8%]

Full English Sample | Targeted for PERT? | Took PERT? | Targeted for CRS? | Took CRS?

All FCAT-takers [N=167,070]

No; too low (FCAT 100 - 286) [N=53,176; 31.8%]

Yes; targeted (FCAT 287 - 354) [N=76,453; 45.8%]

No; too high (FCAT 355 - 500) [N=37,441; 22.4%]

No [N=22,032; 13.2%]

Yes [N=54,421; 32.6%]

No; college-ready (PERT 104 - 150) [N=26,125; 15.6%]

Yes; below ready (PERT 50 - 103) [N=28,296; 16.9%]

Yes; took CRS [N=16,831; 10.1%]

No; other course [N=9,469; 5.7%]

No; no English [N=1,996; 1.2%]

Note. Results are for cohort M1.

Figure 2
McCrary Density Tests
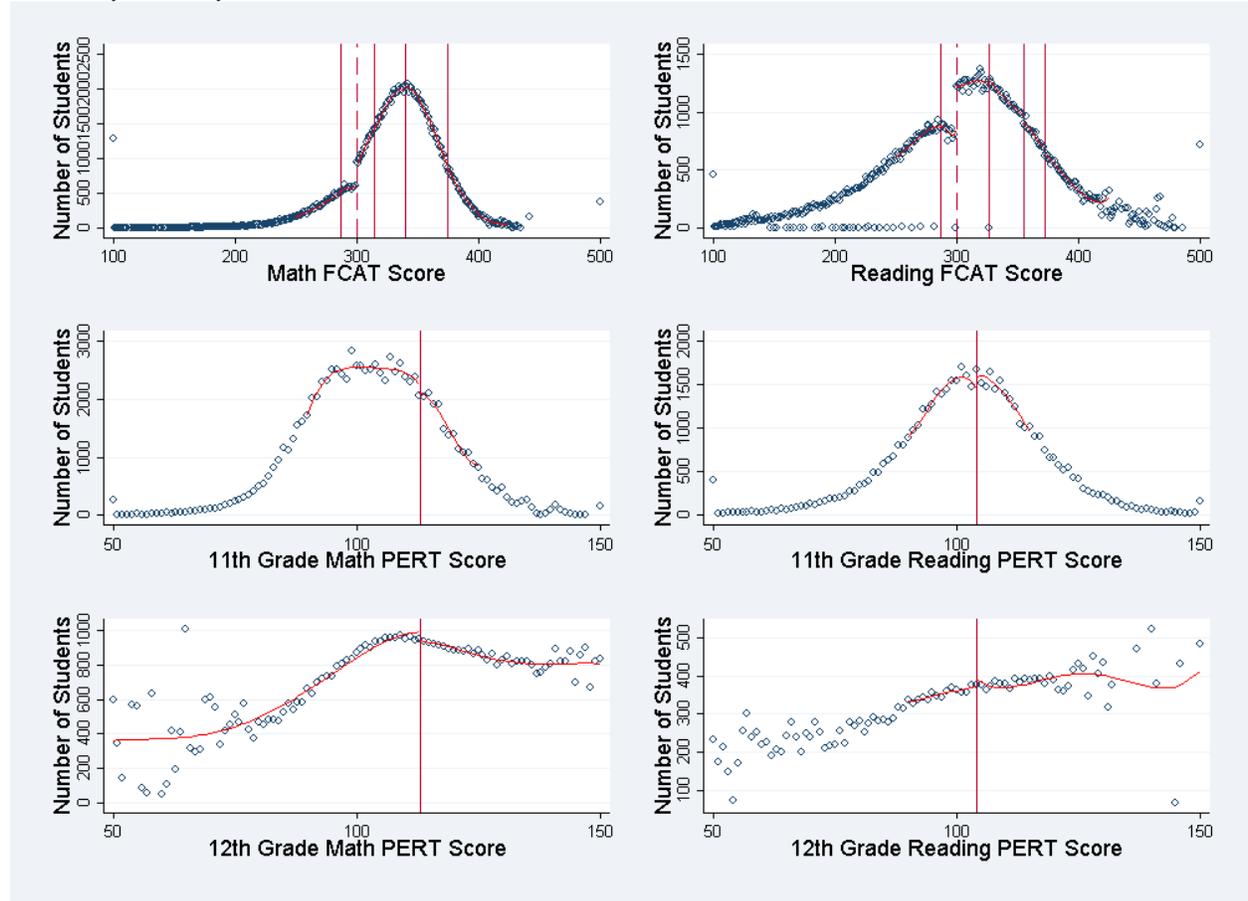
Note. FCAT = Florida Comprehensive Assessment Test; PERT = Postsecondary Education Readiness Test. Results are for cohort M1.Circles on the graphs show the number of test takers at each given score; each circle represents one possible score. The reading FCAT has a large number of scores obtained by 10 or fewer students, very likely representing scaled scores that raw scores do not easily map to. Red lines represent best-fit quartic polynomials, with bandwidths selected to maximize R^2. Solid vertical lines on FCAT graphs represent cutoffs for proficiency levels, while dashed vertical lines represent the cutoff for high school graduation. Vertical solid lines on PERT graphs represent the college readiness cutoff.

Figure 3
RD Estimated Intent to Treat Effect of Treatment for Enrolling in and Passing
Nondevelopmental Courses
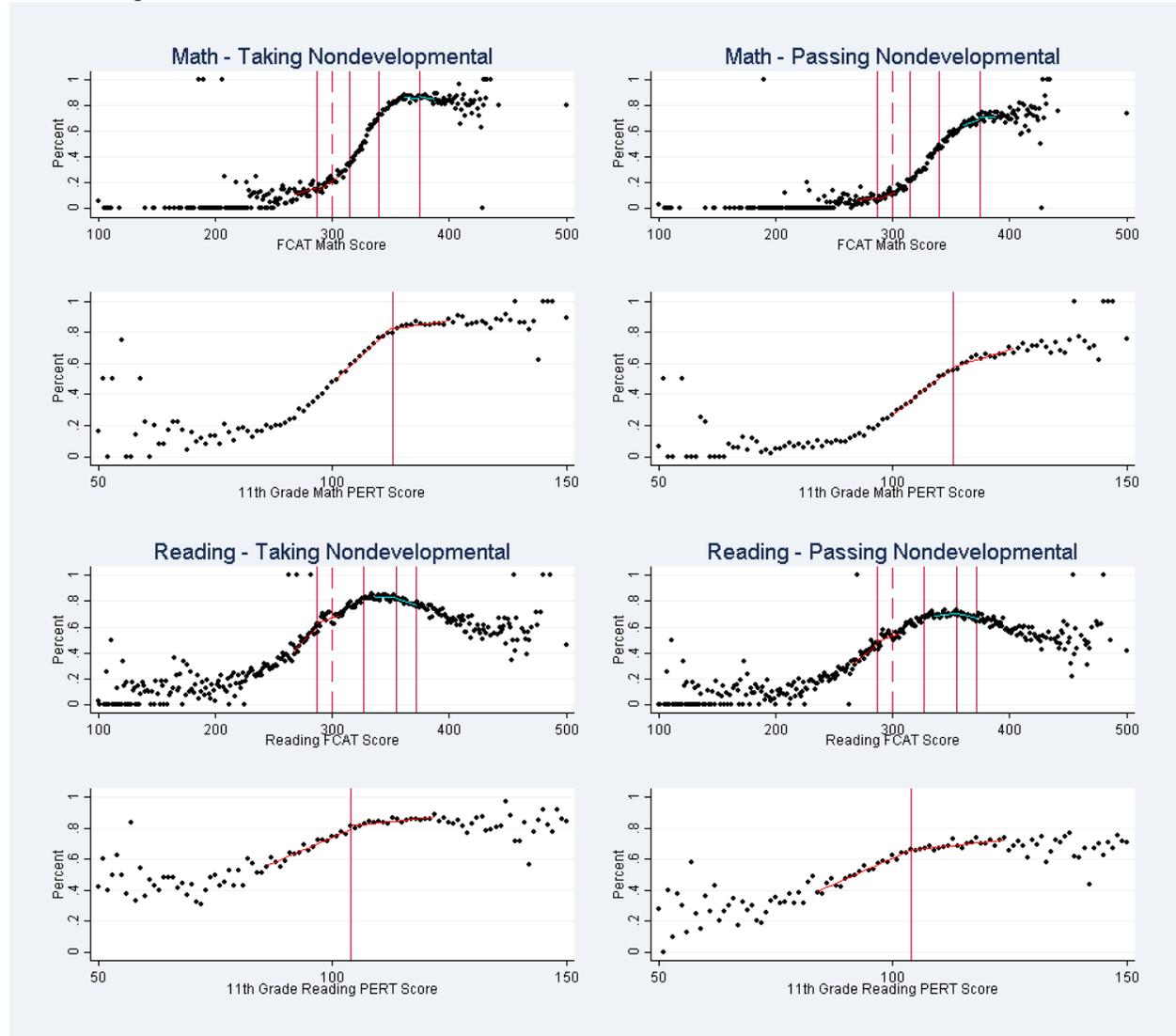
Note. FCAT = Florida Comprehensive Assessment Test; PERT = Postsecondary Education Readiness Test. Markers on the graphs show the average outcome level at each score; each marker represents one possible score. Red and blue best-fit lines represent local linear approximations, with bandwidths selected to maximize R^2. Solid vertical lines on FCAT graphs represent cutoffs for proficiency levels, while dashed vertical lines represent the cutoff for high school graduation. Vertical solid lines on PERT graphs represent the college readiness cutoff. Results are for cohort M1.

Figure 4

Predicted Probability of Enrolling in and Passing Nondevelopmental Courses by Baseline Achievement
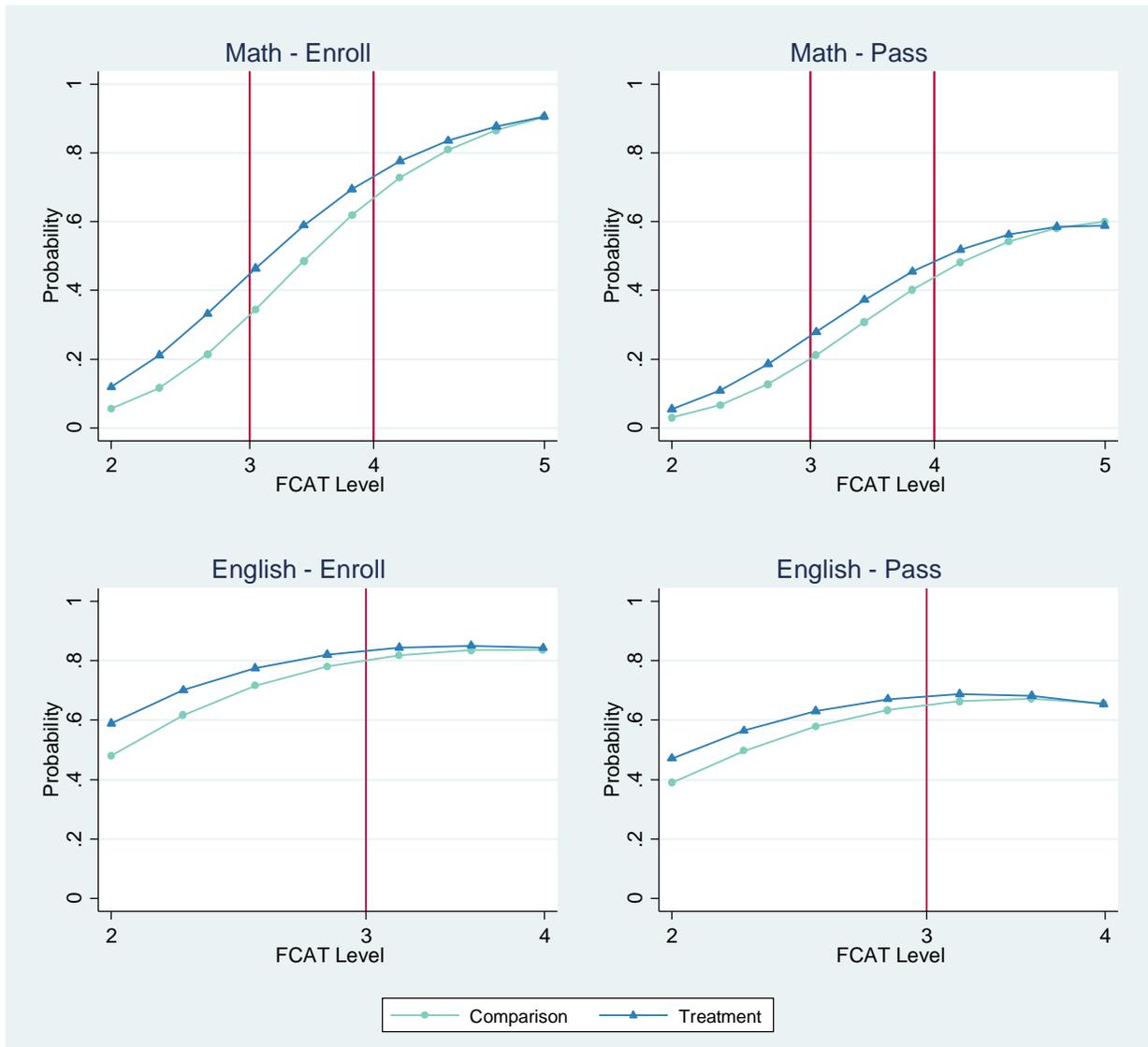
Note. FCAT = Florida Comprehensive Assessment Test. Red lines are drawn between FCAT levels. Results are for the seamless college enrollee subsample in cohorts V2, V3, and M1, with N = 69,718 in math and N = 76,772 in English. Models followed a logit specification and included student background characteristics, pre-treatment achievement, and a school-level fixed effect as regressors.

Figure 5
Predicted Probability of Enrolling in Each Course Level, Math
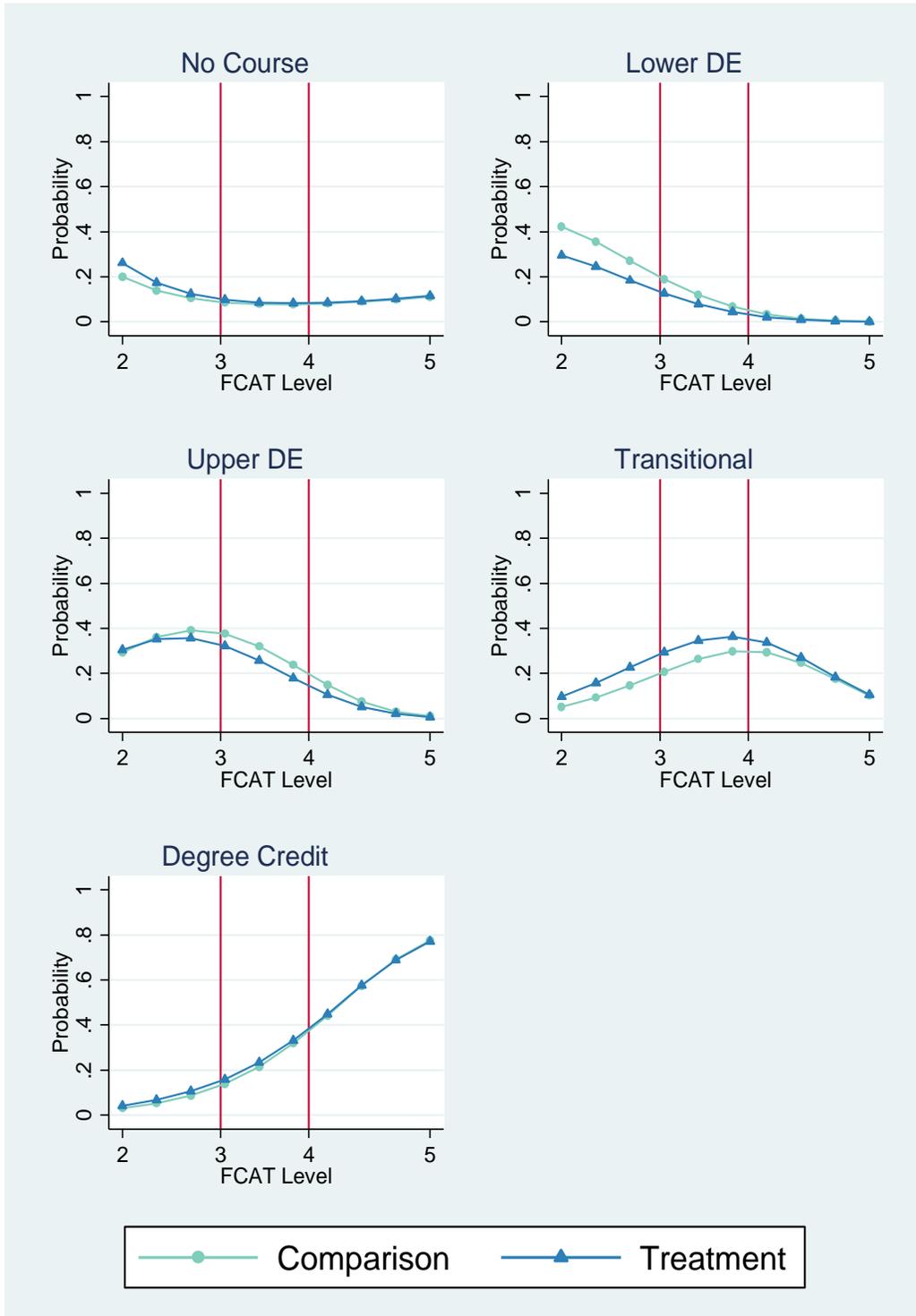
Note. FCAT = Florida Comprehensive Assessment Test; DE = developmental education. Red lines are drawn between FCAT levels. Results are for the seamless college enrollee subsample in cohorts V2, V3, and M1, with N = 69,718 in math and N = 76,772 in English. Models followed a multinomial logit specification and included student background characteristics, pre-treatment achievement, and a school-level fixed effect as regressors.

Figure 6
Predicted Probability of Enrolling in Each Course Level, English
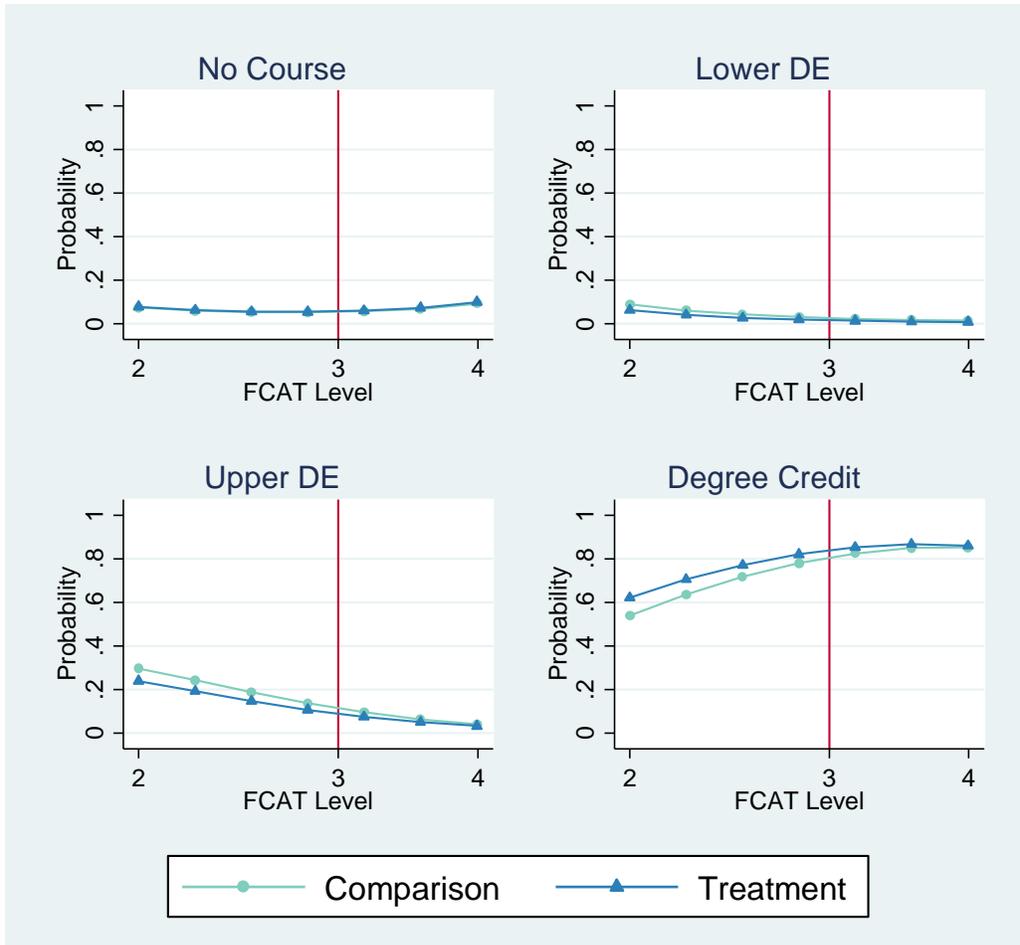
Note. FCAT = Florida Comprehensive Assessment Test; DE = developmental education. Red lines are drawn between FCAT levels. Results are for the seamless college enrollee subsample in cohorts V2, V3, and M1, with N = 69,718 in math and N = 76,772 in English. Models followed a multinomial logit specification and included student background characteristics, pre-treatment achievement, and a school-level fixed effect as regressors.

Online-Only Appendix

Table A1
PERT retesting rates by subject and CRS grade

Grade	Math	English
A	64.4	57.9
B	70.6	63.6
C	70.9	65.2
D	67.9	59.6

Table A2
Descriptive Statistics for Students Targeted in Math

Variable	Sample				
	All targeted students	RD: lower FCAT cutscore	RD: upper FCAT cutscore	RD: PERT cutscore	Before-after
10th grade GPA	2.92	2.29	3.24	2.89	2.90
	(0.56)	(0.54)	(0.54)	(0.52)	(0.57)
Math FCAT	340.00	290.32	370.79	342.91	338.82
	(20.28)	(8.87)	(9.11)	(16.30)	(20.32)
Reading FCAT	331.38	265.66	366.81	327.46	329.74
	(37.41)	(46.09)	(39.83)	(40.68)	(37.27)
FRPL	0.79	0.80	0.39	0.58	0.54
ELL	0.04	0.27	0.12	0.20	0.04
Special Education	0.06	0.22	0.04	0.06	0.06
Female	0.58	0.56	0.45	0.54	0.58
Asian	0.03	0.01	0.05	0.03	0.03
Black	0.19	0.38	0.08	0.18	0.20
Hispanic	0.28	0.30	0.22	0.28	0.31
White	0.47	0.29	0.62	0.48	0.43
Native English	0.77	0.71	0.82	0.76	0.74
N	170,758	4,602	21,791	19,973	69,718
Bandwidth	N/A	16	17	8	N/A

Note. RD = regression discontinuity; GPA = grade point average; FCAT = Florida Comprehensive Assessment Test; PERT = Postsecondary Education Readiness Test; FRPL = free and reduced-price lunch ELL = English language learner. Means are shown for each variable, and for continuous variables, the standard deviation is shown below in parentheses. RD results use the minimum bandwidth of those presented in the Results section for each cutoff. Results are for the seamless college enrollee subsample for cohort M1 for RD and cohorts V2, V3, and M1 for before-after regression.

Table A3

Descriptive Statistics for Students Targeted in English

Variable	All targeted students	RD: lower FCAT cutscore	RD: upper FCAT cutscore	RD: PERT cutscore	Before-after analysis
			Sample		
10th grade GPA	2.88	2.55	3.02	2.74	2.88
	(0.54)	(0.57)	(0.57)	(0.56)	(0.54)
Math FCAT	339.80	326.58	355.39	339.17	339.17
	(21.73)	(22.96)	(21.82)	(21.74)	(21.67)
Reading FCAT	323.07	286.98	353.41	321.12	322.66
	(17.50)	(6.33)	(10.31)	(17.88)	(17.53)
FRPL	0.50	0.67	0.47	0.60	0.52
ELL	0.04	0.21	0.14	0.21	0.04
Special Education	0.06	0.11	0.05	0.08	0.06
Female	0.56	0.51	0.50	0.50	0.56
Asian	0.03	0.02	0.04	0.03	0.03
Black	0.19	0.26	0.12	0.19	0.20
Hispanic	0.28	0.28	0.25	0.30	0.28
White	0.46	0.41	0.57	0.46	0.46
Native English	0.77	0.76	0.81	0.75	0.76
N	121,977	7,637	20,305	22,802	76,772
Bandwidth	N/A	11	18	18	N/A

Note. RD = regression discontinuity; GPA = grade point average; FCAT = Florida Comprehensive Assessment Test; PERT = Postsecondary Education Readiness Test; FRPL = free and reduced-price lunch ELL = English language learner. Means are shown for each variable, and for continuous variables, the standard deviation is shown below in parentheses. RD results use the minimum bandwidth of those presented in the Results section for each cutoff. Results are for the seamless college enrollee subsample for cohort M1 for RD and cohorts V2, V3, and M1 for before-after regression.

Table A4

Mean Comparison of Schools in and out of the Sample for the Before-After Regression

	Math			English		
Variable	In sample	Out of sample	Difference	In sample	Out of sample	Difference
Targeted in math (%)	86.3	84.4	1.9	86.0	83.6	2.4
Targeted in English (%)	55.1	53.9	1.2	54.9	53.5	1.4
ELL (%)	6.9	6.1	0.8	6.3	6.7	-0.4
Special education (%)	12.2	11.0	1.2	12.2	10.1	2.1
Total Enrollment	1654	1620	34	1653	1601	52
Student/teacher ratio	17	16	1	16	17	-1
FRPL (%)	42.8	38.2	4.6	41.5	37.8	3.7
Female (%)	48.5	48.9	-0.4	48.5	49.1	-0.6
Asian (%)	1.9	2.4	-0.5	1.9	2.6	-0.7
Hispanic (%)	23.6	18.5	5.1	20.6	21.3	-0.7
Black (%)	22.7	21.7	1.0	21.8	22.8	-1.0
White (%)	49.1	54.4	-5.3	52.8	50.1	2.7
Rural (%)	25.2	28.1	-2.9	28.0	24.2	3.8
Suburb/Town (%)	55.0	44.9	10.1	50.5	47.7	2.8
City (%)	19.7	27	-7.3	21.5	28.1	-6.6
Attendance (daily average %)	93.2	93.8	-0.6	93.3	94.0	-0.7
Dropout rate (%)	1.7	1.4	0.3	1.7	1.2	0.5
Attending 4-year college (%)	15.7	18.5	-2.8	15.7	20.4	-4.7
Attending 2-year college (%)	35.2	32.9	2.3	34.9	32.0	2.9
N	218	256		321	153	

Note. GPA = grade point average; FCAT = Florida Comprehensive Assessment Test; FRPL = free and reduced-price lunch ELL = English language learner. Results are for cohorts V2, V3, and M1. Schools are included in the sample based on whether or not they switched from low- to high-compliance.

Table A5

Mean Comparison of Baseline Student Characteristics across Treatment Status for the Before-After Regression

	Math			English		
			Std mean			Std mean
Variable	C	T	difference	C	T	difference
10th grade GPA	2.90	2.89	0.019	2.88	2.87	0.012
Math FCAT	338.78	338.88	-0.005	339.17	339.17	0.000
Reading FCAT	329.57	330.00	-0.012	322.86	322.31	0.031
FRPL	0.52	0.56	-0.088	0.50	0.55	-0.086
ELL	0.04	0.04	-0.002	0.04	0.03	0.023
Special Education	0.06	0.07	-0.062	0.06	0.07	-0.066
Female	0.58	0.58	-0.009	0.56	0.57	-0.016
Asian	0.03	0.03	-0.012	0.03	0.03	-0.006
Black	0.20	0.20	-0.008	0.20	0.19	0.013
Hispanic	0.30	0.31	-0.002	0.29	0.28	0.022
White	0.44	0.43	0.004	0.45	0.47	-0.033
Native English	0.74	0.73	0.027	0.76	0.76	-0.006
N	42,184	27,534		48,976	27,796	

Note. C = comparison; T = treatment; GPA = grade point average; FCAT = Florida Comprehensive Assessment Test; FRPL = free and reduced-price lunch ELL = English language learner. Results are for the seamless college enrollee subsample in cohorts V2, V3, and M1.

Table A6

Mean Comparison of School Characteristics across Treatment Status for the Before-After Regression

Variable	Math			English		
	C	T	Difference	C	T	Difference
Targeted in math (%)	85.2	84.2	1.0	84.8	84.1	0.7
Targeted in English (%)	59.5	56.1	3.4	59.5	56.3	3.2
ELL (%)	7.4	7.1	0.2	7.3	6.8	0.5
Special education (%)	10.4	11.0	-0.7	10.7	11.3	-0.6
Total Enrollment	2139.0	2058.0	81.0	2117.0	2026.0	91.0
Student/teacher ratio	20.0	20.0	0.0	19.5	19.7	-0.2
FRPL (%)	45.4	50.7	-5.3	44.2	49.3	-5.1
Female (%)	49.1	49.7	-0.6	49.0	49.6	-0.6
Asian (%)	2.6	2.5	0.1	2.6	2.6	0.0
Hispanic (%)	29.7	31.8	-2.1	27.3	28.5	-1.2
Black (%)	20.9	21.0	-0.1	20.9	20.1	0.8
White (%)	44.0	41.9	2.1	46.4	45.9	0.5
Rural (%)	21.1	21.2	-0.1	17.4	19.8	-2.4
Suburb/Town (%)	59.4	58.9	0.5	60.3	58.8	1.5
City (%)	19.5	19.8	-0.3	22.3	21.4	0.9
Present (daily average %)	93.1	93.6	-0.5	93.2	93.6	-0.4
Dropout rate (%)	1.3	1.0	0.3	1.3	1.0	0.3
Attending 4-year college (%)	18.5	18.2	0.3	18.5	18.0	0.5
Attending 2-year college (%)	36.8	36.7	0.1	36.4	36.3	0.1
N	42,184	27,534		48,976	27,796	

Note. GPA = grade point average; FCAT = Florida Comprehensive Assessment Test; FRPL = free and reduced-price lunch ELL = English language learner. Results are for cohorts V2, V3, and M1. School means include all students and are weighted at the student level.

Table A7

Baseline Equivalence of Student Characteristics for the Regression Discontinuity Analyses

	Math				English			
Variable	bw	N	beta	Effect size	bw	N	beta	Effect size
			RD FCAT, low margin					
10th grade GPA	20	6,662	0.05	0.0870	19	14,156	0.02	0.0398
Math FCAT	-				20	14,971	0.75	0.0231
Reading FCAT	19	6,183	-4.22	0.0787	-			
FRPL	19	6,190	0.0228	0.0455	20	15,015	0.0120	0.0241
ELL	19	6,190	0.0062	0.0162	20	15,015	0.0044	0.0115
Special Education	5	1,523	0.0000 **	0.0000	5	3,770	0.0000	0.0000
Female	20	6,662	-0.0151	0.0304	20	15,015	-0.0068	0.0136
Asian	20	6,662	0.0072	0.0375	19	14,156	-0.0010	0.0055
Black	20	6,662	0.0034	0.0087	20	15,015	0.0166	0.0432
Hispanic	20	6,662	-0.0108	0.0247	20	15,015	-0.0255 **	0.0582
White	20	6,662	-0.0100	0.0200	17	12,422	0.0088	0.0176
Native English	19	6,190	-0.0084	0.0200	19	14,156	-0.0045	0.0107
			RD FCAT, high margin					
10th grade GPA	18	23,401	0.00	0.0060	20	22,898	-0.01	0.0096
Math FCAT	-				19	21,817	-0.29	0.0089
Reading FCAT	19	24,590	-0.55	0.0102	-			
FRPL	20	25,909	0.0135	0.0270	20	22,898	0.0032	0.0064
ELL	16	20,784	-0.0023	0.0059	20	22,898	-0.0076	0.0199
Special Education	5	6,818	0.0000 ***	0.0000	5	6,553	0.0000 **	0.0000
Female	20	25,909	-0.0096	0.0194	20	22,898	0.0091	0.0184
Asian	17	22,103	0.0025	0.0133	19	21,838	-0.0010	0.0050
Black	12	15,725	-0.0131 **	0.0340	20	22,898	0.0049	0.0128
Hispanic	18	23,401	-0.0031	0.0072	20	22,898	-0.0075	0.0172
White	20	25,909	0.0088	0.0176	20	22,898	0.0050	0.0099
Native English	18	23,401	0.0013	0.0032	20	22,898	-0.0183 ***	0.0438
			RD grade 11 PERT					
10th grade GPA	13	30,076	-0.02 **	0.0396	18	23,135	-0.02	0.0344
Math FCAT	11	26,639	0.71 *	0.0345	19	23,678	-0.73 *	0.0319
Reading FCAT	15	33,204	1.79 **	0.0388	15	21,156	0.12	0.0068
FRPL	20	39,295	-0.0158 **	0.0316	18	23,135	0.0111	0.0222
ELL	18	37,173	0.0004	0.0011	16	21,910	-0.0114 **	0.0297
Special Education	5	14,229	0.0000 ***	0.0000	8	13,897	0.0000 **	0.0000
Female	19	38,267	-0.0032	0.0065	19	23,711	-0.0073	0.0147
Asian	18	37,173	0.0034	0.0192	20	24,192	0.0008	0.0045

Variable	Math				English			
	bw	N	beta	Effect size	bw	N	beta	Effect size
Black	17	35,912	0.0083	0.0211	19	23,711	-0.0207 ***	0.0537
Hispanic	14	31,841	-0.0092	0.0207	19	23,711	-0.0048	0.0109
White	17	35,912	-0.0015	0.0030	17	22,562	0.0224 **	0.0448
Native English	18	37,173	0.0008	0.0020	18	23,135	-0.0091 **	0.0216
RD grade 12 PERT								
10th grade GPA	15	11,778	-0.01	0.0218	14	3,686	-0.01	0.0229
Math FCAT	18	12,745	1.22 ***	0.0710	15	3,812	-1.68	0.0749
Reading FCAT	13	10,921	1.78 *	0.0439	9	2,822	0.01	0.0007
FRPL	15	11,778	0.0000	0.0001	14	3,686	0.0026	0.0054
ELL	18	12,745	-0.0118	0.0273	5	1,828	0.0036	0.0083
Special Education	8	8,159	0.0000	0.0000	8	2,591	0.0000	0.0000
Female	15	11,778	0.0301 *	0.0610	14	3,686	-0.0367	0.0734
Asian	13	10,934	-0.0054	0.0395	20	4,293	-0.0125	0.0808
Black	10	9,430	-0.0376 **	0.0874	18	4,128	0.0398 **	0.0939
Hispanic	20	13,227	0.0214 *	0.0458	20	4,293	0.0009	0.0018
White	11	9,947	0.0187	0.0385	20	4,293	-0.0197	0.0405
Native English	13	10,934	-0.0081	0.0180	8	2,591	-0.0190	0.0419

Note. FCAT = Florida Comprehensive Assessment Test; PERT = Postsecondary Education Readiness Test; FRPL = Free or Reduced-Price Lunch; ELL = English Language Learner. These tables show the coefficients of regression discontinuity analyses using the listed characteristics as outcome variables. Columns show the bandwidth used, the number of students within that bandwidth under casewise deletion, the point estimate, and the effect size (in full-sample standard deviations). * = significant at the 10% level, ** = significant at the 5% level, *** = significant at the 1% level. In the FCAT analyses, we were unable to use a single measure as both the running variable and outcome variable, so these rows are left blank.

Table A8
RD Estimated Intent to Treat Effect, Local Linear Bandwidth = 5

Model Version	Outcome			
	Any High School Diploma	Seamless College Enrollment	Nondevelopmental Enrollment	Nondevelopmental Pass
	Math			
FCAT, Low Margin	0.0114	-0.0175	-0.0002	-0.0016
	(0.0189)	(0.0281)	(0.0421)	(0.0224)
FCAT, High Margin	-0.0015	0.0065	0.0039	-0.0127
	(0.0063)	(0.0217)	(0.0165)	(0.0243)
Grade 11 PERT	0.0028	0.0068	0.0133	0.0181
	(0.0035)	(0.0132)	(0.0170)	(0.0233)
Grade 12 PERT	0.0034	0.0151	0.2647 ***	0.0920 ***
	(0.0045)	(0.0148)	(0.0300)	(0.0308)
	English			
FCAT, Low Margin	-0.0019	-0.0025	0.0136	-0.0001
	(0.0076)	(0.0181)	(0.0486)	(0.0425)
FCAT, High Margin	0.0075	-0.0156	0.0143	-0.0044
	(0.0047)	(0.0182)	(0.0247)	(0.0242)
Grade 11 PERT	-0.0020	-0.0058	-0.0245	-0.0071
	(0.0053)	(0.0162)	(0.0229)	(0.0206)
Grade 12 PERT	-0.0149 *	0.0071	0.1487 **	0.1207 ***
	(0.0082)	(0.0323)	(0.0634)	(0.0425)
Sample	Students within bw	Students within bw	Seamless college enrollees within bw	Seamless college enrollees within bw

Note. FCAT = Florida Comprehensive Assessment Test; bw = bandwidth; PERT = Postsecondary Education Readiness Test. Results are for cohort M1. Results shown are for local linear regression with bandwidths of five points on the relevant exam and cutoff. Numbers reported are the difference in predicted probabilities across treatment status. Standard errors are shown in parentheses, while bandwidths are listed in brackets. Regressions control for gender, free or reduced-price lunch status, race/ethnicity, English language learner status, native English speaker status, disability status, gifted/talented status, cumulative GPA as of 10th grade, and district indicators. N varies by specification; for FCAT N ranges from 1,350 to 9,096, for grade 11 PERT N ranges from 8,566 to 21,754, for grade 12 PERT N ranges from 1,658 to 10,446.
*p<.10. **p<.05. ***p<.01.

Table A9

RD Estimated Intent to Treat Effect, Local Linear Bandwidth = 10

	Outcome			
Model Version	Any High School Diploma	Seamless College Enrollment	Nondevelopmental Enrollment	Nondevelopmental Pass
	Math			
FCAT, Low Margin	-0.0017 (0.0137)	0.0030 (0.0193)	-0.0475 ** (0.0210)	-0.0291 * (0.0174)
FCAT, High Margin	-0.0011 (0.0035)	0.0053 (0.0122)	-0.0186 (0.0134)	-0.0193 (0.0163)
Grade 11 PERT	-0.0009 (0.0028)	-0.0029 (0.0083)	0.0121 (0.0111)	0.0027 (0.0177)
Grade 12 PERT	0.0023 (0.0029)	0.0110 (0.0110)	0.2918 *** (0.0222)	0.1298 *** (0.0201)
	English			
FCAT, Low Margin	0.0020 (0.0064)	-0.0062 (0.0117)	0.0147 (0.0264)	-0.0033 (0.0255)
FCAT, High Margin	0.0044 (0.0040)	-0.0023 (0.0149)	0.0001 (0.0132)	-0.0166 (0.0127)
Grade 11 PERT	0.0042 * (0.0024)	0.0036 (0.0130)	-0.0267 (0.0175)	-0.0055 (0.0164)
Grade 12 PERT	0.0039 (0.0048)	0.0108 (0.0245)	0.1988 *** (0.0402)	0.1349 *** (0.0287)
Sample	Students within bw	Students within bw	Seamless college enrollees within bw	Seamless college enrollees within bw

Note. FCAT = Florida Comprehensive Assessment Test; bw = bandwidth; PERT = Postsecondary Education Readiness Test. Results are for cohort M1. Results shown are for local linear regression with bandwidths of 10 points on the relevant exam and cutoff. Numbers reported are the difference in predicted probabilities across treatment status. Standard errors are shown in parentheses, while bandwidths are listed in brackets. Regressions control for gender, free or reduced-price lunch status, race/ethnicity, English language learner status, native English speaker status, disability status, gifted/talented status, cumulative GPA as of 10th grade, and district indicators. N varies by specification; for FCAT N ranges from 2,700 to 18,139, for grade 11 PERT N ranges from 15,730 to 40,364, for grade 12 PERT N ranges from 2,963 to 17,855.
*p<.10. **p<.05. ***p<.01.

Table A10
RD Estimated Intent to Treat Effect, Local Linear Bandwidth = 20

Model Version	Outcome			
	Any High School Diploma	Seamless College Enrollment	Nondevelopmental Enrollment	Nondevelopmental Pass
	Math			
FCAT, Low Margin	-0.0008	0.0076	-0.0322	-0.0272
	(0.0097)	(0.0126)	(0.0201)	(0.0166)
FCAT, High Margin	0.0022	0.0153	-0.0073	-0.0114
	(0.0025)	(0.0095)	(0.0075)	(0.0143)
Grade 11 PERT	0.0075 ***	-0.0020	0.0109	-0.0137
	(0.0018)	(0.0060)	(0.0114)	(0.0123)
Grade 12 PERT	-0.0002	0.0027	0.3404 ***	0.1992 ***
	(0.0027)	(0.0105)	(0.0228)	(0.0196)
	English			
FCAT, Low Margin	-0.0012	-0.0009	0.0425 **	0.0218
	(0.0047)	(0.0099)	(0.0191)	(0.0169)
FCAT, High Margin	0.0005	0.0101	0.0079	-0.0004
	(0.0037)	(0.0097)	(0.0078)	(0.0097)
Grade 11 PERT	-0.0011	-0.0084	-0.0201	-0.0025
	(0.0025)	(0.0092)	(0.0123)	(0.0098)
Grade 12 PERT	-0.0001	0.0057	0.2484 ***	0.1447 ***
	(0.0040)	(0.0201)	(0.0288)	(0.0241)
Sample	Students within bw	Students within bw	Seamless college enrollees within bw	Seamless college enrollees within bw

Note. FCAT = Florida Comprehensive Assessment Test; bw = bandwidth; PERT = Postsecondary Education Readiness Test. Results are for cohort M1. Results shown are for local linear regression with bandwidths of 20 points on the relevant exam and cutoff. Numbers reported are the difference in predicted probabilities across treatment status. Standard errors are shown in parentheses, while bandwidths are listed in brackets. Regressions control for gender, free or reduced-price lunch status, race/ethnicity, English language learner status, native English speaker status, disability status, gifted/talented status, cumulative GPA as of 10th grade, and district indicators. N varies by specification; for FCAT N ranges from 6,247 to 37,387, for grade 11 PERT N ranges from 23,945 to 69,983, for grade 12 PERT N ranges from 4,278 to 26,105.
*p<.10. **p<.05. ***p<.01.

Table A11
RD Estimated Intent to Treat Effect, Local Quadratic Regression with Optimal Bandwidths

Model Version	Any High School Diploma	Seamless College Enrollment	Nondevelopmental Enrollment	Nondevelopmental Pass
		Outcome		
	Math			
FCAT, Low Margin	0.0039	-0.0056	-0.0569 **	-0.0400 **
	(0.0146)	(0.0216)	(0.0240)	(0.0164)
	[bw = 20]	[bw = 20]	[bw = 18]	[bw = 18]
FCAT, High Margin	-0.0007	-0.0004	-0.0117	-0.0189
	(0.0059)	(0.0190)	(0.0172)	(0.0213)
	[bw = 13]	[bw = 13]	[bw = 15]	[bw = 15]
Grade 11 PERT	0.0055	-0.0018	0.0119	0.0093
	(0.0042)	(0.0089)	(0.0161)	(0.0181)
	[bw = 8]	[bw = 19]	[bw = 11]	[bw = 20]
Grade 12 PERT	0.0059 *	0.0120	0.2635 ***	0.0671 **
	(0.0033)	(0.0152)	(0.0254)	(0.0280)
	[bw = 18]	[bw = 20]	[bw = 16]	[bw = 13]
	English			
FCAT, Low Margin	-0.0005	-0.0038	0.0289	-0.0012
	(0.0071)	(0.0144)	(0.0311)	(0.0264)
	[bw = 17]	[bw = 17]	[bw = 19]	[bw = 20]
FCAT, High Margin	0.0060	-0.0079	-0.0027	-0.0171
	(0.0041)	(0.0144)	(0.0165)	(0.0153)
	[bw = 19]	[bw = 19]	[bw = 18]	[bw = 20]
Grade 11 PERT	0.0012	0.0070	-0.0259	-0.0082
	(0.0025)	(0.0145)	(0.0177)	(0.0156)
	[bw = 20]	[bw = 18]	[bw = 18]	[bw = 19]
Grade 12 PERT	-0.0002	0.0169	0.1954 ***	0.1489 ***
	(0.0066)	(0.0283)	(0.0451)	(0.0306)
	[bw = 18]	[bw = 18]	[bw = 17]	[bw = 17]
Sample	Students within bw	Students within bw	Seamless college enrollees within bw	Seamless college enrollees within bw

Note. FCAT = Florida Comprehensive Assessment Test; bw = bandwidth; PERT = Postsecondary Education Readiness Test. Results are for cohort M1. Results shown are for local quadratic regression with bandwidths optimally determined on a regression-by-regression basis via cross-validation. Numbers reported are the difference in predicted probabilities across treatment status. Standard errors are shown in parentheses, while bandwidths are listed in brackets. Regressions control for gender, free or reduced-price lunch status, race/ethnicity, English language learner status, native English speaker status, disability status, gifted/talented status, cumulative GPA as of 10th grade, and district indicators. N varies by specification; for FCAT N ranges from 5,424 to 32,810, for grade 11 PERT N ranges from 22,835 to 67,630, for grade 12 PERT N ranges from 4,018 to 26,105.
*p<.10. **p<.05. ***p<.01.

Table A12
Treatment Rates by Sample

Treatment Type	Math		English	
	Treatment	Comparison	Treatment	Comparison
RD FCAT, low margin				
No treatment (%)	7	35	16	38
College readiness test only (%)	27	14	38	18
CRS course only (%)	7	24	6	22
Both (%)	60	27	40	23
RD FCAT, high margin				
No treatment (%)	45	88	34	82
College readiness test only (%)	47	9	50	11
CRS course only (%)	1	2	3	6
Both (%)	7	0	13	1
RD grade 11 PERT				
No treatment (%)	0	0	0	0
College readiness test only (%)	41	88	45	82
CRS course only (%)	0	0	0	0
Both (%)	59	12	55	19
Before-after regression				
No treatment (%)	15	86	16	82
College readiness test only (%)	30	14	35	18
CRS course only (%)	7	0	11	0
Both (%)	48	0	39	0

Note. RD = regression discontinuity; FCAT = Florida Comprehensive Assessment Test; CRS = college readiness and success; PERT = Postsecondary Education Readiness Test. Results are for cohort M1 for RD and cohorts V2, V3, and M1 for before-after regression. All treatment rates were calculated for the seamless college enrollee subsample. For the grade 12 PERT analysis, all students in the treatment and comparison groups participated in both the college readiness test and CRS course.