An NCPR Working Paper

The Impact of Postsecondary Remediation Using a Regression Discontinuity Approach: Addressing Endogenous Sorting and Noncompliance

Juan Carlos Calcagno
Mathematica Policy Research and
Community College Research Center, Teachers College, Columbia University
jcalcagno@mathematica-mpr.com

Bridget Terry Long
Harvard Graduate School of Education, National Bureau of Economic Research,
and National Center for Postsecondary Research
longbr@gse.harvard.edu

April 2008

National Center for Postsecondary Research
www.PostsecondaryResearch.org

The National Center for Postsecondary Education is a partnership of the
Community College Research Center, Teachers College, Columbia University;
MDRC; the Curry School of Education at the University of Virginia;
and professors at Harvard University and Princeton University.

This research was generously supported by the Spencer Dissertation Fellowship, Lumina Foundation for Education through the Achieving the Dream: Community Colleges Count initiative, and the National Center for Postsecondary Research (NCPR), which was established by a grant from the Institute of Education Sciences of the U.S. Department of Education.

For information about NCPR and NCPR publications, visit www.postsecondaryresearch.org.

Abstract

Remedial or developmental courses are the most common policy instruments used to assist underprepared postsecondary students who are not ready for college-level coursework. However, despite its important role in higher education and its substantial costs, there is little rigorous evidence on the effectiveness of college remediation on the outcomes of students. This study uses a detailed dataset to identify the causal effect of remediation on the educational outcomes of nearly 100,000 college students in Florida, an important state that reflects broader national trends in remediation policy and student diversity. Moreover, using a Regression Discontinuity design, we discuss concerns about endogenous sorting around the policy cutoff, which poses a threat to the assumptions of the model in multiple research contexts. To address this concern, we implement methods proposed by McCrary (2008) and discuss the strengths of this approach. The results suggest math and reading remedial courses have mixed benefits. Being assigned to remediation appears to increase persistence to the second year and the total number of credits completed for students on the margin of passing out of the requirement, but it does not increase the completion of college-level credits or eventual degree completion. Taken together, the results suggest that remediation might promote early persistence in college, but it does not necessarily help students on the margin of passing the placement cutoff make long-term progress toward earning a degree.

Contents

Acknowledgments

We would like to thank Josh Angrist, Tom Bailey, Eric Bettinger, Melissa Clark, John Deke, Kevin Dougherty, Tom Kane, Hank Levin, and Miguel Urquiola for detailed comments and suggestions that have improved the paper as well as participants at the Teachers College Society of Economics and Education Seminar and the Spencer Foundation Fall Fellows Workshop. We are also grateful to Justin McCrary for providing the Stata codes; to Pat Windham, Judith Thompson, and Sandra Burkholder for sharing the data and for their suggestions; and to Peter Crosta and Matthew Jacobus for excellent research assistance. All errors, omissions, and conclusions are our own.

1. Introduction

Remedial or developmental education, defined as coursework below college-level offered at a postsecondary institution, is a topic of considerable debate in higher education.[1] The conceptual foundation for remedial coursework is straightforward — students are tested to determine whether they meet a given level of academic proficiency in order to enroll in college-level coursework. Deficiencies in tested skills are addressed through some form of supplementary instruction, most often remedial courses. Many are concerned, however, about the significant costs of remediation. Colleges and states devote substantial resources to remediation. One conservative estimate suggests that public colleges spend one to two billion dollars annually on remedial education programs (Breneman & Haarlow, 1998). More recently, a report found that remediation at Florida community colleges cost $118.3 million during school year 2004-2005 with 53 percent of this being paid by the state (Office of Program Policy and Government Accountability [OPPAGA], 2006). Not surprisingly, many policymakers have begun to question the need to pay for academic preparation that they believe should have occurred in secondary school, and many states have recently introduced plans to reduce the availability of postsecondary remedial courses or limit its cost (Merisotis & Phipps, 2000; Bettinger & Long, 2007). Remediation is also costly to students. While the courses often do not qualify for college credit, students must nonetheless pay tuition for them and bear the opportunity cost of foregone earnings. In 2003-04, Florida community college students who required remediation took an average nine credit hours of remedial coursework and paid an additional $504 for college prep coursework during their first year of college (OPPAGA, 2006, p. 4).

Meanwhile, student demand for remediation has increased in recent decades. Nationally, it is estimated that only one-third of students leave high school at least minimally prepared for college (Greene & Foster, 2003). Of those who enter higher education, over one-third are required to take remedial courses in reading, writing, or mathematics (National Center for Education Statistics [NCES], 2003). Remediation rates are particularly high at two-year community colleges, which open their doors to all students regardless of their level of academic preparedness (Dougherty, 1994). Based on longitudinal data from the high school class of 1992, nearly 60 percent of first-time community college students took at least one remedial course (Attewell, Lavin, Domina, & Levey, 2006), and similar numbers were found among community college students in Ohio (Bettinger & Long, 2007). In fact, partly due to the belief that remedial courses can be offered for a lower cost at community colleges, at least ten states have elected to

[1]The literature sometimes defines remediation as coursework that is retaken while developmental courses are classes that focus on new material. Here, however, the terms “remediation,” “college prep,” and “developmental education” are used interchangeably.

focus their remediation efforts at the two-year colleges and more are considering doing so (Bettinger & Long, 2007). This study focuses on remedial courses at two-year colleges, and so reflects this larger national trend.

Unfortunately, the ongoing debates about whether and where to offer remediation lack a large knowledge base about the effectiveness of the courses. The lack of research knowledge is due primarily to the unavailability of data but also to the failure of most research to account for the non-random assignment into remedial courses. By definition, less-prepared students are more likely to be placed in remedial education, and hence, straightforward OLS regressions on the impact of remediation on academic outcomes are biased due to selection (Bettinger & Long, forthcoming; Grubb, 2001). However, several recent efforts have attempted to address the selection problem using quasi-experimental approaches. Bettinger and Long (forthcoming) make use of differences in remedial policies across public institutions in Ohio to compare similar students who have had varying experiences with remediation based on the college they attend. This study instead uses a regression discontinuity (RD) design, which exploits the fact that remedial placement in Florida is largely based on a test score. This quasi-experimental approach assumes that in the absence of the treatment, a sample of students close to the cutoff will be academically equivalent due to some randomness in test outcomes around the discontinuity; thus, students who barely pass the remedial testing cutoff are good counterfactuals for their treated peers. Although this approach has been widely used in other contexts to obtain causal inferences when selection bias exists (Trochim, 1984; Angrist & Lavy, 1999; Van der Klaauw, 2002; Jacob & Lefgren, 2004; Lee, 2008), it has rarely been applied to the study of remediation programs in higher education. With the availability of new data sources, however, this may be changing.

This study uses the RD approach to compare students just above and below the placement test cutoff to examine the impact of postsecondary remediation on student outcomes. The few other studies on remediation using an RD design have generally used only very small samples and are thus difficult to interpret.[2] One exception is Martorell and McFarlin (2007), which also applies the RD approach to compare students in Texas. However, this paper extends the literature in several important ways. It uses a unique, large administrative dataset of college students in Florida to explore more contexts and issues concerning postsecondary remediation than earlier work as well as to comment on the challenges inherent in determining the causal impact of remediation. We detail the application of the RD approach to this research context, including ways to address noncompliance with the placement rule. Additionally, we address a major methodological threat to the RD research approach in much research: there may be

[2]See Aiken, West, Schwalm, Carroll, and Hsiung (1998); Lesik (2006); and Moss and Yeaton (2006). The difficulty in interpreting the results develops because the RD approach generally requires large samples in order to allow for the comparison of students around a narrow band of the remedial placement cutoff.

endogenous sorting around the policy cutoff. In the context of remediation, some students may be permitted to take the remedial placement exam multiple times in order to pass out of the courses, and this invalidates the fundamental underlying assumption of the RD design. We address this concern by applying the test of manipulation proposed by McCrary (2008) to provide a solution for the retesting problem and a sensitivity test for the more straightforward estimates. This should help inform future research on remedial programs as well as analyses using the RD approach to evaluate interventions in other contexts.

The other major contribution of the paper is to provide additional information on the impact of math and reading remediation on student outcomes. Using an incredibly rich data source, we provide detailed information on a number of outcomes for nearly 100,000 students. Moreover, because we focus on Florida, our estimates provide information about remediation that is relevant nationwide. Florida is one of the ten states that discourage the offering of remedial education at four-year institutions, which is a growing policy trend (Bettinger & Long, 2007; Jenkins & Boswell, 2002). The Florida community college system is also the third largest in the nation and enrolls nearly six percent of community college students nationwide.[3]

The results suggest remediation has limited or mixed benefits. After controlling for noncompliance and endogenous sorting around the placement test cutoff score, students on the margin of requiring math remediation were slightly more likely to persist to their second year than their non-remedial peers, but there was no detectable impact for reading. Meanwhile, the likelihood of passing subsequent college-level English composition was slightly lower for remedial students while no difference was found in future course performance for math remedial students. Finally, the impacts for both math and reading remediation are found to be positive in terms of total credits earned, but no statistically significant difference was found in terms of total college-level (non-remedial) credits earned. Taken together, the results suggest that remediation might promote early persistence in college, but it does not necessarily help community college students on the margin of passing the cutoff make long-term progress toward a degree.

The remainder of the paper is organized as follows. Section 2 provides a literature review and explains the methodological challenges associated with the evaluation of remedial education. It also provides background information on remediation in Florida and on the data used in the analysis. Section 3 details the research design and empirical strategy, including the RD research approach and application of McCrary (2008) to deal with nonrandom sorting. Section 4 discusses the results, and Section 5 presents our conclusions.

---

[3]Source: Authors' computations based on the *Digest of Education Statistics* (NCES, 2004).

2. Literature Review and Background on the Florida Context

Is Remediation Effective? Methodological Challenges and Past Causal Estimates

While postsecondary remediation plays an important role in higher education, little is known about its effectiveness in improving the outcomes of underprepared students. There are reasons to believe that the effects of remedial courses could be positive or negative. Advocates claim that remediation is an important, necessary, and effective component of higher education. On the other side, critics argue that remediation is a barrier that increases the requirements that are needed before taking college-level courses, thereby lowering completion and transfer probabilities. Moreover, the literature suggests that placement into remediation may lower self-esteem and educational expectations, possibly due to a student being stigmatized by peers and faculty, and hence negatively impact student outcomes.[4]

Even though 35 to 40 percent of first-time college students are placed into remediation each year, the topic remains an understudied component of higher education. Early research on remediation has been mainly descriptive, simply comparing the outcomes of students in remediation to those not in remedial courses. However, selection issues preclude such a straightforward analysis because there are inherent differences between students placed in remediation and those who pass out of the courses. Unfortunately, until recently, few studies have been able to overcome these research concerns. Two reviews of the literature on remedial and developmental education found the bulk of studies to be "methodologically weak" with almost two-thirds reflecting "serious methodological flaws" (O'Hear & MacDonald, 1995; Boylan & Saxon, 1999). Another concern of the past research is that most studies often do not track students across time, which prevents analysis of longer-term outcomes such as degree completion.

With the availability of new data sources, several major studies on the impact of remediation have been completed in recent years. The first set of large scale studies, by Bettinger and Long (2004, forthcoming), use an instrumental variable strategy that combines between-college variation in remediation placement policies and the importance of distance in college choice to estimate the causal effect of remedial courses on higher education outcomes. This sort of comparison is possible in Ohio, the target state of the analysis, because institutional policies regarding remediation differ across the public colleges and universities. Therefore, two students with the same characteristics face dissimilar probabilities of remediation if they attend

---

[4]For a comprehensive discussion of advocates' arguments, see McCabe (2000). Deil-Amen and Rosenbaum (2002) provide the critics' perspective.

different schools. The analysis focuses on degree-seeking, traditional-age (18 to 20 years old), full-time undergraduates who entered a public college in fall 1998. Their results suggest that remedial students at Ohio colleges are more likely to persist in college and to complete a bachelor's degree in comparison to students with similar test scores and backgrounds who were not required to take the courses. Moreover, Bettinger and Long (2005) found that community college students placed in math remediation were 15 percent more likely to transfer to a four-year college and to take ten more credit hours than students with similar test scores and high school preparation. Overall, these results suggest that remedial classes have beneficial effects for students in Ohio.

Martorell and McFarlin (2007) instead examine the impact of remediation in Texas, a state with a single placement exam and cutoff score, similar to Florida. Using an RD design similar to the basic model of this paper, the study exploits information on college students' remedial placement exam scores to compare students just above and below the placement cutoff. Martorell and McFarlin find that remediation has little effect on a wide range of educational and labor market outcomes. The estimates are small and statistically insignificant but suggest that students are neither harmed nor greatly benefited by remediation.

Even with the recent research developments on the effectiveness of remediation, little is known about the causal impact of remedial courses on underprepared students beyond Ohio and Texas. Moreover, past causal results provide conflicting evidence with positive effects found in Ohio and no effect found in Texas. This paper provides additional estimates using an incredibly rich data source of nearly 100,000 students in Florida, a large, important state that reflects broader national trends in remediation policy and student diversity. Moreover, we discuss the application of the RD approach in this context, address the issue of noncompliance, and note concerns about endogenous sorting around the policy cutoff. To address the issue of endogenous sorting, we implement a solution using methods proposed by McCrary (2008) and discuss the strengths of this approach. The section below gives details on postsecondary remediation in Florida and describes the dataset we use to examine the impact of remediation in that context.

Postsecondary Remediation in Florida: Background and the Dataset

All first-time degree-seeking applicants for admission to community colleges and universities in Florida must be tested before registration to demonstrate certain basic skills before beginning college-level courses. Basic skills are measured using standardized test scores on the Florida College Entry Level Placement Test (CPT).[5] The CPT is a computer adaptive college placement testing program and is part of the ACCUPLACER system, developed by the College Board at the request of the Florida Department of Education.[6] Students must meet certain statewide cutoff scores set by the State Board of Education to be considered "college ready." Incoming students who do not achieve minimum scores on the Elementary Algebra, Reading Comprehension, and Sentence Skills sections of the college placement test must take remedial classes before they begin college-level work in each subject. In other words, students are assigned to either remedial or college-level courses, depending on their scores on the standardized tests. Colleges may exempt students from taking the CPT if the students meet the appropriate college-ready scores on the College Board's SAT or the American College Testing Program's Enhanced ACT.

To examine the impact of remediation in this context, our study uses a unique dataset obtained from the Florida Department of Education K-20 Education Data Warehouse (EDW). EDW integrates existing and transformed data extracted from multiple sources into a single data repository focusing on students served in Florida's public education system as well as educational facilities, curriculum, and staff involved in instructional activities. Our data include information on test scores and demographic characteristics, including age, gender, race/ethnicity, citizenship, previous education (high school diploma, other diploma, or GED), and English language proficiency. For this study, the dataset focuses on the universe of first-time community college students who enrolled at any of the 28 Florida community colleges from fall 1997 to fall 2000 and sought at least an associate (two-year) degree.[7] Additionally, we focus on the sample who reported CPT scores. Among the 130,862 first-time degree-seeking students during the time period of this study, 75 percent (98,146 students) reported the CPT scores while 13 and 12 percent reported the SAT or ACT scores, respectively. Students for whom we have only SAT or ACT test scores are excluded due to artificial "stacking" at different discrete points when these scores are converted to CPT equivalents.[8]

---

[5]High school students in dual enrollment programs are also required to take the CPT before enrolling in college-level courses.

[6]ACCUPLACER is designed to facilitate the evaluation and placement of college students in three basic skills areas: reading, writing and arithmetic. The purpose of ACCUPLACER tests is to determine which course placements are appropriate for students and whether or not remedial work is needed (College Board, 2003).

[7]Student are considered associate degree-seeking if the college classifies them as being in a two-year degree program based on voiced intent and/or first term course selection. Two-year degree programs include: Associate in Arts degree, Associate in Science degree, General Freshman, and Associate in Applied Science degree. Note that only students seeking an associate degree are required to take the CPT placement exam. Because we only include students with these scores in our analysis, we again reinforce our intent to focus on "associate degree-seeking" students.

[8]Although we have SAT or ACT information for students who did not take the CPT, they are excluded for two reasons. First, each test has different score ranges: SAT (200-800), ACT (1-36), and CPT (20-120), and

The main variables of interest in this study, assignment to remediation and participation status, are defined using test scores and longitudinal information on remedial education courses taken by subject (Math and Reading).[9] The dataset tracks term-by-term enrollment for all students in the sample for a total of six years for each cohort. For example, the cohort that began in fall 2000 is tracked until spring 2006, a total of 17 terms or 6 years of outcomes.[10] The term-by-term information includes course-taking patterns. The short-term outcomes investigated include whether a student enrolled and completed the first college-level course in the remediation area (college algebra and freshman English composition) and fall-to-fall persistence. Long-term educational outcomes include completion of a certificate, completion of an associate degree, and transfer to the Florida State University System (SUS). We also use two additional measures of educational attainment: total credits earned (remedial and non-remedial) and total non-remedial or college-level credits earned. All these outcome measures are computed within the six-year window allowed by the dataset.

Summary statistics of the dataset are provided in Table 1. The first column of numbers displays the characteristics of all students who entered a Florida community college for the first time from fall 1997 to fall 2000 while the second column limits the sample to those with CPT test scores, the main sample used in the analysis. Comparisons of columns 1 and 2 show few differences between the two samples. However, there are differences in remedial placement and educational outcomes because the students with only SAT or ACT scores (and no CPT score) were slightly better prepared.[11]

While the CPT is the statewide required tool to assign remediation, the data suggest that all students do not follow the straightforward assignment rules, and this has important implications for the empirical analysis. Such deviation from the assignment rule is common in studies that attempt to use discontinuities in test scores or other criteria to determine the causal impact of an intervention (Angrist & Lavy, 1999; Battistin & Rettore, 2002; Van der Klaauw, 2002; Jacob & Lefgren, 2004). In this context, the first issue of concern is students who, while having CPT scores that dictate they should take remedial courses, do not actually do so. The most likely

---

though there are conversion rules between the tests, conversion leads to additional noise in the CPT distribution due to artificial "stacking" at different discrete points in the CPT score. Second, starting with the fall 2000 semester, the SAT and ACT scores required to be considered "college-ready" were increased in order to align them with the required scores of the CPT exam. Therefore, fall 2000 students with only SAT or ACT scores faced different requirements than earlier cohorts in the data.

[9]For simplicity, remedial writing classes are not analyzed here. Scores on the reading comprehension and sentence skills sections of the CPT are highly correlated (0.8), as are assignment and enrollment rates.

[10]There are three terms per year in Florida: fall, spring, and summer.

[11]Florida colleges accept SAT and ACT as placement scores if they meet a minimum standard. Students who submit such scores often have planned ahead of time to transfer to a four-year college, as these schools require the tests (personal communication with Dr. Patricia Windham, Associate Vice-Chancellor for Evaluation, Division of Community Colleges, Florida Department of Education, May 2006).

explanation for this noncompliance is that Florida rules permit students assigned to one particular remedial subject to take college-level courses concurrently in other curriculum areas for which they are qualified. Almost 52 percent of remedial math students in the non-complier group take advantage of this flexibility, but only a quarter of the students in reading do so. Another possible explanation is that some students might be discouraged by being placed into remediation and leave the institution prior to taking any credits. Analysis suggests that this explains as much as 14 and 19 percent of non-compliers in math and reading remediation, respectively (Calcagno, 2007).[12] Such noncompliance must be addressed in the empirical analysis, and our methods for doing so are detailed below.

A second and more serious concern is that some students may be able to take the CPT multiple times to increase their chances of passing the exam. This could result in nonrandom sorting around the policy cutoff, which is a concern for research using the method more generally (Imbens & Lemieux, 2008; McCrary, 2008; Lee, 2008). Research suggests that this is largely a difference in institutional policies (Windham, 2005; Lesik, 2006, 2007; Perin, 2006). The final two columns of Table 1 begin to examine this issue by calculating the mean characteristics and outcomes of students at institutions with no statistical evidence of endogenous sorting around the cutoff. The methods for identifying these schools are detailed below using methods proposed by McCrary (2008), and the results of these calculations are discussed as well.

[12]Yet another possible explanation is that some institutions might use an additional test for placement beyond the CPT or allow some students to enroll in college-level courses, thereby waiving their remediation requirement (Perin, 2006). However, our analysis suggests that less than two percent of the sample re-tested out of remedial courses using some other criteria.

3. Research Design and Empirical Strategy

The Regression Discontinuity Strategy

This section presents a model to understand the methodological challenges associated with the evaluation of remedial education and the empirical strategy undertaken in this paper. The basic notation follows Rubin's model for causal inference where Y_i^T and Y_i^C are the potential outcomes that a given student *i* would have obtained by taking (superscript T), or not taking (superscript C) remedial education (Rubin, 1974). The individual causal effect of the program could be estimated by the difference in outcomes, $\hat{\beta} = Y_i^T - Y_i^C$, but both outcomes can never be observed for the same student (Holland, 1986).

The ideal solution is to select a sample of *N* students from the population and divide them randomly into a treatment and control group. The latter group serves as a counterfactual to estimate average treatment effects (ATE). Let T be a binary indicator for treatment status (T = 1 for treatment; T = 0 for control), then the causal effect is the difference in the empirical means of Y for each group. However, random assignment is not feasible in remedial education programs (Levin & Calcagno, 2008). Most previous research assumed that by controlling for a set of observable variables, selection to remediation could be ignored (this is known as the selection on observables or conditional independence assumption). For example, it is common to assume a linear relation between remedial education and outcomes and estimate the following regression model:

$$(1) \qquad Y_i = \beta_1 T_i + \gamma X_i + \varepsilon_i$$

where Y is the outcome of interest; X is a set of observable variables (e.g., gender, race/ethnicity, socioeconomic status); and ε is a random error term with $E\,[\varepsilon_i \mid T_i] = 0$. However, controlling for observables is likely insufficient to deal with the selection issue if the assignment of the treatment depends on unobserved variables that are correlated with the outcome. For example, less motivated students are more likely to be placed in remedial courses, but these factors are generally unobservable. Hence, the estimated coefficient $\hat{\beta}_1$ not only captures the program effect, but also the influence of pre-treatment factors (Bettinger & Long, forthcoming; Grubb, 2001).

A regression-discontinuity design (RD) takes advantage of the remedial placement rules and cutoffs to estimate the causal effect of remedial education on educational outcomes.[13] Let Z

[13]For a review of theoretical and practical issues involved in the regression-discontinuity design, see Imbens and Lemieux (2008). Hahn, Todd, and Van der Klaauw (2001) provide a formal analysis of identification issues.

be the continuous score in the standardized test (the variable used for assignment), and $\overline{Z}$ the threshold for assignment to remedial classes. Then the treatment status and the assignment variable are related through a deterministic and discontinuous function $T_i = 1(Z_i \leq \overline{Z})$ that is known to the researcher. Students scoring below $\overline{Z}$ in the test are assigned to remedial courses, while those scoring above are not. Hence, potential outcomes and treatment status are conditionally independent, and a regression within the immediate vicinity of $\overline{Z}$ will yield a causal estimate $\hat{\beta}_1$ at the cutoff, analogous to results from a randomized experiment.[14]

This RD approach assumes that in the absence of the treatment, a sample of students close to the cutoff will be similar. In Table 2, we use student-level covariates to show statistical equivalence in average characteristics for all degree-seeking students in the dataset with scores below and above the cutoff by subject, a test for random assignment around the discontinuity point (Imbens & Lemieux, 2008; Lee, 2008). As expected, the means for observable student factors are statistically different for these two groups in each remedial subject, but the differences vanish when comparing students within a small band around the cutoff. Even for this subsample of similar students, however, there are small differences in the proportion of Hispanic and foreign students (for math and reading) and in age and the proportion of African-American students (for reading). Note that these differences could be purely due to chance; even in a randomized experiment, there will generally be a few differences between groups. Using student-level covariates in the regression analysis allows us to minimize any lack of balance and serves as a test for random assignment around the discontinuity point (Lee, 2008). However, one might still be concerned about unobservable differences between the groups. The fundamental assumption of the RD design may be violated if waivers out of remediation are not distributed at random or if additional unobserved factors that determine the likelihood of retaking and passing the exam are related to educational outcomes. Therefore, we develop techniques to deal with two methodological threats to the sharp RD design: noncompliance and endogenous sorting around the cutoff.

Dealing with Noncompliance: The Fuzzy RD Design

The sharp regression discontinuity method described above assumes full compliance. However, there could be differences between mandated *assignment* and actual *enrollment* (treatment recipient), and as result the average probability of enrollment in remedial courses could be less than one below the cutoff and more than zero above the cutoff. Given a single cutoff policy, two different types of noncompliance can occur: *no-shows,* defined as those

[14]However, the treatment effect can only be identified locally at the point at which the probability of receiving treatment changes discontinuously, unless the impact is constant across different students. The impact of the program on students who are extremely underprepared for college-level courses may be quite different.

treatment group students who do not receive the treatment, and *crossovers,* those control group students who do receive the treatment (Bloom, 1984). Figure 1 shows the probability of enrollment in math and English college prep courses by CPT score in the Florida dataset. Note that on the left side of the graph pertaining to remedial math, enrollment in remedial classes below the cutoff is around 80 percent, generating an average of 20 percent of *no-show* students. As discussed above, this may be the result of students leaving the institution immediately after being placed into remediation or of Florida rules that allow students to take college-level courses in other subjects concurrently with remedial courses. As shown on the right side of the same graph, on average, 8 percent of students scoring above the cutoff in math did enroll in some type of remedial math course (there were almost no crossovers for English). Campbell (1969) terms this situation as fuzzy regression-discontinuity.

Note that it still holds that $\Pr(T_i = 1 | Z_i = z)$ has a discontinuity at $z = \overline{Z}$, and this condition can aid in identifying different parameters of interest. To see this, assume that the basic model with constant treatment effect presented in Equation (1) can be modified to introduce the divergence between assignment (D_i) and actual recipient of the treatment (T_i). Then we can write the regression model for the effect of remedial education on higher education outcomes as follows:

$$(2) \qquad Y_i = \beta_1 D_i + \beta_2 f(Z_i) + \varepsilon_i$$

where D is the indicator for *assignment* to remedial education, $f(Z_i)$ is a smooth function of student's score in the standardized test, and all other variables are as described previously. In this case, after conditioning on the test score, a regression on Equation (2) yields a consistent estimator $\hat{\beta}_1$, often referred to as the intent-to-treat effect (ITT). ITT estimates the gains that a policymaker can realistically expect to observe from implementing the program given the observed levels of noncompliance (Heckman, LaLonde, & Smith, 1999), but it does not represent the effect of the treatment for those who actually receive it.

One approach to address noncompliance is using instrumental variables (Heckman et al., 1999; Gennetian, Morris, Bos, & Bloom, 2005). An instrumental variable (IV) approach combined with the RD design uses the exogenous determination of assignment as an instrument for enrollment in remediation (henceforth, RD-IV). The IV exclusion restrictions are satisfied by design because assignment is strongly correlated with enrollment in remedial classes but is also uncorrelated with the error term in the outcome equation because assignment was exogenously determined by the cutoff policy. In the context of a regression analysis, suppose the first stage regression is:

$$(3) \qquad T_i = \delta_1 D_i + \delta_2 f(Z_i) + v_i$$

and the outcome response is related to the treatment via the equation:

$$(4) \qquad Y_i = \beta_1 \hat{T}_i + \beta_2 f(Z_i) + \varepsilon_i$$

where β_1 is the two-stages estimator of the causal effect of remedial classes on educational outcomes. This IV strategy estimates the local average treatment effect (LATE) that captures the impact of receiving the treatment for the subpopulation of students whose treatment status was induced by the cutoff policy (Imbens & Angrist, 1994; Angrist, Imbens, & Rubin, 1996).[15] In other words, LATE estimates the effect for those students encouraged by the statewide cutoff policy to enroll in college preparatory classes.[16] In the next section, we discuss in detail a potential nonrandom sorting around the cutoff and our proposed solution.

Concerns about Endogenous Sorting: Retesting as an Evaluation Problem

Public knowledge of treatment assignment rules and cutoffs may generate unexpected behavioral responses by students (Imbens & Lemieux, 2008; McCrary, 2008; Lee, 2008). In the context of remedial education and standardized test scores, students might have the option to re-take the placement exam. For example, a student may contact an institution that allows retesting at the beginning of the summer before registration to take the CPT exam. Students scoring below the cutoff who are not interested in remediation might be encouraged to prepare for the exam, retake it at the end of the summer, and use this final CPT score for placement.[17] If additional unobserved factors jointly determine the likelihood of passing the remedial cutoff after retesting and educational outcomes (such as motivation), then re-taking invalidates the key

[15]Besides the IV exclusion restrictions, the LATE estimator also assumes that the treatment causes statistically detectable effects. Moreover, the model assumes a constant treatment effect, although the same framework can be extended to allow heterogeneous treatment effects across observable student characteristics. Another assumption behind the LATE estimator is *local monotonicity*, which holds that any student who would enroll in remedial classes in the absence of assignment would be in the treatment group if assigned to the treatment group. Students who would never comply are called *defiers* and are ruled out by the monotonicity condition.

[16]Students may be exposed to different treatment intensities by enrolling in more than one remedial course while they are in college. The average number of math remedial courses taken in the sample is 1.8 (s.d. 1.03), and the average number of reading courses taken is 1.4 (s.d. 0.73). One would expect the effect to vary by number of courses taken in the same area. LATE estimates in this case are weighted averages of per-unit causal effect (Angrist & Imbens, 1995).

[17]Note that the dataset includes only students who enrolled for the first time at the community college in the fall. Hence, the retesting problem is minimized here by excluding those students who enrolled in credit courses at the community college beforehand. Even so, it is still possible that a student may take the exam at the beginning and end of the summer without formal enrollment at the community college.

identifying assumption behind the RD design (i.e., unobservable characteristics vary smoothly through the cutoff point) and the results will be subject to selection bias.[18]

Unfortunately, our dataset does not contain information on a student's multiple test attempts; only the score used for placement is included. However, if students can take the assessment test repeatedly, then some, especially those who scored below but close to the cutoff score, may do so until they exceed that score. If this happens, one would expect to see a larger number of students who barely exceed the remedial cutoff score than those who barely failed. This situation would lead to a discontinuity of the conditional density of the test score at the threshold that can be detected using graphical analysis (Imbens & Lemieux, 2008; McCrary, 2008).

Figure 2 shows this estimated density by subject (rows) and race/ethnicity groups (columns). The first thing to note across all racial/ethnic groups is that the densities are fairly continuous for math but discontinuous at the cutoff for reading. This suggests that retesting is more likely for reading than for math. There are also differences by race/ethnicity. African-American and Hispanic students appear to be less likely to retest than White students. In principle then, the regression-discontinuity analysis conditional on student-level covariates should reduce the effect of the retesting problem. Nevertheless, if additional unobserved factors (such as motivation) jointly determine the likelihood of passing after retesting and educational outcomes, then retesting will still invalidate the underlying RD identification assumptions, and straightforward RD estimates would be subject to selection bias.

Instead of looking at the retesting problem by observable student-level characteristics, the literature asserts that retesting is an institutional policy (Lesik, 2006, 2007; Perin, 2006). In fact, a recent developmental education survey conducted by the Florida Department of Education shows that retesting was allowed in seventeen institutions (out of 28) under specific conditions such as if scores were near the cutoff, or if the student was not currently enrolled in remediation.[19] A major limitation of this survey for the purposes of this study is that it was conducted in 2005, and the data used here go back to 1997. Institutions may have changed their retesting policies over time, and therefore, results based on the survey's information would be subject to measurement error. Instead, we use the manipulation test recently proposed by McCrary (2008) to identify institutions with no statistical evidence of endogenous sorting around the cutoff. By replicating our analysis on this subsample where the RD identifying assumptions holds, we provide a robustness test for our previous estimates.

---

[18]See Urquiola and Verhoogen (2007) for a similar problem in the context of the class-size reduction debate.

[19]In addition to the information available in Windham (2005, Appendix C, Chart IX), we had access to individual college-level answers.

The test entails estimating the density function of the CPT exam on either side of the cutoff point. As discussed above, a discontinuous density at the cutoff provides evidence of manipulation (retesting), although this is neither necessary nor sufficient for identification except under auxiliary assumptions (McCrary, 2008; p. 5). In practice, McCrary's test is executed in two steps as follows.[20] The first step involves plotting the histogram and creating a frequency table for the CPT exam. The bins for the histogram are defined so that no bin includes points both to the left and right of the cutoff point. McCrary recommends using a bin size equal to $\hat{b} = 2\hat{\sigma}\mathrm{n}^{-1/2}$, where $\hat{b}$ is the estimated bin size, $\hat{\sigma}$ is the sample standard deviation of the CPT exam, and *n* is the number of observations (see McCrary, 2008, p. 10). We use a bin size equal to one, the natural discrete unit of the CPT exam, for large schools (more than 1,500 students and enough observations per bin), and McCrary's recommended bin size for small institutions.

The second step is a local linear regression of the histogram separately on either side of the cutoff to accommodate the discontinuity. The midpoints of the histogram bins are treated as covariates in the regression, and the normalized counts of the number of observations falling into the bins are treated as outcomes. Finally, the discontinuity at the cutoff is then estimated as the log difference in height on the intercept:

$$(5) \qquad \hat{\theta} \equiv \ln \hat{f}^{+} - \ln \hat{f}^{-}$$

where $\hat{f}^{+}$ and $\hat{f}^{-}$ are estimated values for the density just above and below the cutoff respectively.

Once the discontinuity at the cutoff ($\hat{\theta}$) and its standard error ($\hat{\sigma}_{\theta}$) are estimated for each community college, a formal t-test can be constructed for H0: $\hat{\theta} = 0$, or no statistical evidence of discontinuity at the cutoff.[21] Therefore, we define as no-retesting institutions those colleges where there is no statistical evidence of a discontinuity in the density function of the test score at the cutoff; more specifically, they are so defined if McCrary's t-test of the null hypothesis of continuity at the cutoff fails to reject.

Estimation of the Parameters of Interest

A number of important issues are involved in the practical estimation of the parameters of interest. First, for the binary outcomes, we use the maximum likelihood probit method to

[20]We are grateful to Justin McCrary for providing us with the Stata programs for this analysis.
[21]We follow McCrary (2008) to compute standard errors and the optimal bandwidth.

estimate models, and we report the marginal effects at mean values.[22] For the continuous dependent variables, we estimate OLS models. Second, our dataset includes a detailed set of student-level covariates in addition to the test scores that we use to increase the precision of the estimated program impacts, to increase the power of significance tests, and to eliminate small sample biases (Imbens & Lemieux, 2008). Third, all standard errors are clustered by test score to account for this uncertainty in the unknown parametric part of the model, as Lee and Card (2008) suggest is appropriate in RD settings in which the assignment variable is discrete.

A fourth important practical issue involved in the estimation is using a proper specification of the function form of Y(Z) at both sides of the discontinuity. The difficulty, however, is that the true functional form is often unknown. Following recommendations made by Reichardt, Trochim, and Cappelleri (1995) and Shadish, Cook, and Campbell (2002, p. 233), we specify *f* in equations (2) through (4) as a low-order polynomial in the test score after a close graphical inspection of the empirical functional form and model fit analysis. As will be discussed in detail later, a linear or quadratic specification generally provides a good fit of the data. The introduction of higher-order polynomials does not change the conclusions presented here and in most cases cubic terms on the test score were not statistically significant or showed no improvement in terms of model fit.

It should be noted that Imbens and Lemieux (2008) suggest using a non-parametric local linear regression (LLR) approach using only the observations close to the discontinuity point to estimate RD impacts. We therefore re-estimated all our ITT remediation models using LLR, a rectangular kernel, and an estimated optimal bandwidth that varies by outcome from 15 to 20 points around the cutoff. When doing so, the estimates barely change in terms of size and statistical significance. As a result, we decided to present the impacts estimated using low-order polynomial regressions instead of LLR. For a discussion of strengths and weaknesses of each method see McCrary and Royer (2006, footnote 23).

Finally, we test the sensitivity of our estimates to different sub-samples as they appear within different bandwidths of the cutoff. We estimate our models using data from all students in the sample and also for a restricted sample of students with test scores within a 20 points band around the cutoff. We choose to report impacts only for a 20 points band around the cutoff because this was the most likely optimal bandwidth obtained from the LLR analysis as suggested by Imbens and Lemieux (2008). The results are robust to using bands of 10 or 6 points (Calcagno 2007).

---

[22] See Angrist (2001) for a discussion of models with binary outcomes and dummy endogenous regressors.

4. Results: Remedial Courses and Educational Outcomes

This section discusses estimates the impact of remediation on seven outcomes. One measure of success for remedial students is whether they can enroll and pass the first college-level course in math and English composition. It would be expected that after successfully learning the skills needed for college-level work, a remedial student would be more likely than an academically-equivalent non-remedial student to complete these courses. These courses, College Algebra (MAC 1105) and Freshman Composition Skills I (ENC 1101), are required for all standard associate degree programs, and so there should be no selection problems in terms of which students elect to take the courses.[23] Therefore, in terms of short-term educational outcomes, we first examine differences in the likelihood of passing the initial college-level course in a subject after completing remediation. A second outcome of interest is fall-to-fall (one year) persistence. A common argument against remedial classes in the literature is that placement in remediation is a barrier that discourages students from persisting in college by increasing the number of requirements needed before taking college classes (Deil-Amen & Rosenbaum, 2002; Rosenbaum, 2001). We test this discouragement hypothesis.

For longer-term outcomes, we investigate the likelihood that students on the margin of remedial placement complete a certificate, an associate degree, and/or transfer to a Florida public four-year university (i.e., within the Florida State University System). Research has shown that completion of a certificate, an associate degree, or transfer to a higher-level college has positive effects on earnings (Jaeger & Page, 1996; Kane & Rouse, 1999), and so it is important to understand how remediation may help in achieving this final goal. As noted above, remediation could lower completion and transfer by increasing the requirements students must meet. However, if remedial classes successfully teach or refresh the skills needed for college-level work, remedial students should be more likely than academically equivalent non-remedial students to complete a certificate or degree or to transfer to a four-year university. Because the sample is limited to those seeking an associate degree, two-year degree completion may be most relevant outcome as all students may not have had the intent to transfer to a four-year college. Still transfer is an important policy outcome for the state. It is also important to acknowledge that our data do not allow us to witness the transfer of students to four-year universities that are

[23]Both should be taken during the freshman year for a standard associate degree program, though the exact course might differ slightly depending on the major. Unfortunately, we know nothing about student majors or specific requirements. These courses seem to be taken by virtually all students persisting through the freshman year.

private or outside the state.[24] Remediation may also divert some students to certificate programs, and so we explore the completion of certificates. The final two outcomes under investigation are total credits earned and total college-level credits earned over six years.

Graphical Analysis of the Impact of Remediation

Figures 3 and 4 provide a visual identification of the ITT effect of math and reading remediation on six of the educational outcomes. The discontinuous relation between CPT scores and the probability of enrollment in remedial education permits one to visually identify this effect. If remedial courses had a substantial net impact on educational outcomes, one would expect to see a jump in the conditional mean of the outcome around the cutoff (Imbens & Lemieux, 2008). For example, the first row, first column graph in Figure 3 shows the relationship between completion of the first college-level math course and math CPT score. The circles are the average outcomes for students with a given CPT score. The fitted lines are predicted probabilities from a linear probability model for each educational outcome on the assignment to treatment variable and quadratic polynomial terms in the CPT score. The evidence for passing the first college-level math course as well as for associate degree completion (first row, second column) and transfer to a Florida four-year university (second row, second column) suggests a small negative ITT effect (the estimated discontinuities are listed at the top of each panel). Conversely, results for fall-to-fall persistence (second row, first column) show a positive gap between students scoring just below and above the cutoff. The last column shows two complementary graphs. In the first row the outcome is total credits earned over six years including "college-level credits" (those that count toward degree completion) and "institutional credits" (credits that count toward financial aid and full-time student status but *not* toward degree completion, i.e., remedial credits). In the second row the outcome includes only college-level credits. The comparison between these two graphs suggests that although students with scores below the cutoff earn more total credits, the statistical difference vanishes for earning credits that count toward a college degree.

Moving to remedial reading, shown in Figure 4, the pattern of estimated discontinuities are similar to the impacts found for math remediation. The ITT impact is small and negative for passing the first college-level course, for associate degree completion, and for transfer to a

---

[24]We do not observe transfers to Florida private institutions or schools outside of Florida. A study by the Florida Department of Education, Division of Community Colleges, found that among a cohort of first-time college students in 1999 who completed at least 12 hours during a six year tracking period, 8.5 percent transferred to an institution outside the Florida State University System (SUS) without earning a credential beforehand. Our estimates may be biased if remedial and non-remedial students transfer outside the SUS at different rates.(personal communication with Dr. Patricia Windham, Associate Vice-Chancellor for Evaluation, Division of Community Colleges, Florida Dept. of Education, April 2007).

Florida public four-year college. The comparison between the two graphs for credits earned over six years suggests similar conclusions: students with scores below the cutoff earn more total credits, but the statistical difference vanishes or become negative for earning credits that count toward a degree.

These graphical analyses provide important feedback regarding two empirical issues. First, there is no evidence of any other jump in the conditional expectation of the outcomes given test scores other than at the expected discontinuity at the threshold. Second, the regression model using low-order polynomials for test scores (linear or quadratic) generally provides a good track of the empirical local averages. The next two subsections explore in detail these discontinuities using the regression framework described in section 3.

The Impacts of Math and Reading Remedial Placement: Regression Analysis

Tables 3 (math remediation) and 4 (reading remediation) follow the same format. Each row focuses on a different outcome, with each cell corresponding to a different method that is detailed by the column heading. For the binary outcomes, we use the maximum likelihood probit method to estimate models, and we report the marginal effects at mean values. For the continuous dependent variables, we estimate OLS models. ITT is the intention-to-treat estimate from equation (2). RD-IV is the instrumental variable estimate from equation (4).

Columns (1) and (2) show the baseline ITT and RD-IV impacts for the complete sample of students, and columns (3) and (4) add controls for age, gender, race/ethnicity, citizenship, English limited proficiency, the test score in the opposite subject, and cohort fixed effects (all other specifications also include controls). The rest of the columns provide robustness checks. In columns (5) and (6) we test the sensitivity of our estimates by estimating our models on a restricted sample of students with test scores within a 20 points band around the cutoff. The results in columns (7) to (10) are discussed in the next subsection.

The impacts of math remediation on completion of the first college-level math class are shown in Table 3 (first row). Point estimates for students on the margin of the cutoff are negative, ranging from 1.4 to 3 percent, but they are not statistically significant (with or without controls). Note that impacts for the narrow band sample change sign, but the size is still very small and not statistically different from zero (columns [5] and [6]). Shifting to the effect on fall-to-fall persistence, the results in Table 3 do not support the discouragement hypothesis. ITT and RD-IV effects for math remediation for students on the margin of the cutoff are not statistically different from zero at conventional levels but are positive. Bettinger and Long (forthcoming) also do not find evidence of a discouragement effect in their study of Ohio students, though their results suggest a positive effect on persistence.

Results on the impact of math remediation on certificate or associate degree completion and transfer to a public four-year university in Florida are shown in rows 3 through 5. All impacts across the different samples are very small, negative, and not statistically different from zero. These results do not support the critics' hypothesis that remediation is harmful, but they are not as optimistic as previous findings by Bettinger and Long (forthcoming). The results are much more similar to Martorell and McFarlin (2007) in their RD study of Texas students.

The last two outcomes are total credits earned ("college-level" and "institutional") and total non-remedial college-level credits earned. The estimated impacts suggest that the average math remedial student earned between 3 and 7 more credits (depending on the specification) than his or her academically-equivalent, non-remedial peers. Although earning more credits is a desirable outcome, the next row of estimates shows the limitation of this measure. When only credits that count toward a college degree are included as the dependent variable, the impact of remediation is not statistically different from zero. All these impacts are robust to the different samples and controls.

The impacts for reading remediation are shown in Table 4. All of the ITT and RD-IV estimates for passing English composition, earning an associate degree, and transfer to the SUS are statistically significant, negative, and robust to student-level controls and the choice of bandwidth. For example, students on the margin of the cutoff induced to take remedial reading courses due to the cutoff policy were 9 percentage points less likely to pass English composition during the 6 years data window (column [6]). Similarly, they are 4 and 2.5 percentage point less likely to complete an associate degree or transfer to a public four-year college, respectively. We also found no statistical impact for fall-to-fall persistence or earning a certificate. The effect of reading remediation on credits earned is smaller than the impact for math remediation and sensitive to the choice of bandwidth around the passing cutoff. The average remedial reading student earned between 1 and 3 more total credits that his or her non-remedial peers, although the impacts are negative for earning non-remedial credits.

Accounting for Endogenous Sorting: The Retesting Problem

Results for the McCrary manipulation test per institution are presented in Table 5. Each row in the table represents each community college in Florida. For each subject, column (1) is the estimated bandwidth *h*, the window width defining which observations are included in the regression. Column (2) is the estimated bin size used for each institution. Columns (3) and (4) are the estimated discontinuity theta and its standard error, respectively, and these last two parameters are combined to compute the t-test in column (5) using traditional formulas. As is conventional, t-test values lower than 1.96 (bolded) are associated with a 5 percent level of significance and suggest that there is no statistical evidence of a discontinuity in the CPT distribution at the cutoff. The t-test of the null hypothesis of continuity fails to reject for nineteen community colleges for math and for seven for reading. These community colleges are considered *non-retesting* institutions in the analysis that follows, and we expect the RD identification assumption to hold for this subsample of colleges free of endogenous sorting around the cutoff. Note also that the fact that only seven institutions pass the manipulation test for reading remediation versus nineteen colleges for math suggests that test re-taking is more likely for reading; thus, we anticipate more bias in remedial reading impacts due to re-testing.

As shown in the last two columns of Table 1, the institutions that do not allow retesting serve students with characteristics that are only slightly different than the entire research sample. The average age of students in the subset of non-retesting schools is slightly higher, and students are slightly more likely to be recommended for remedial placement (as expected). In terms of the schools that do not allow retesting in math, a slightly larger proportion of the students were African American or Hispanic, but the average CPT scores are very similar to the overall research sample. Schools that did not allow retesting in reading had more African-American students but fewer Hispanic students. As is evident from the number of observations, many more students attended schools that allowed retesting in reading than in math.

The regression results for this robustness test are presented in the last columns of Tables 3 and 4. Columns (7) and (8) show estimates for no-retesting colleges only, and columns (9) and (10) combine no-retesting colleges and the narrow band sample. Impacts for math remediation are remarkably similar in terms of size, sign, and statistical significance. This is supportive of our argument that test re-taking is less likely for math based on the evidence that the density for the math exam scores is strongly skewed to the left and fairly continuous at the cutoff (Figures 2 and 3). Additionally, 19 out of 28 institutions passed the McCrary manipulation test. The one exception is in terms of fall-to-fall persistence. Columns (7) and (8) suggest that once focusing on the no-retesting institutions, remedial students are more likely to persist.

To sum up the overall results for math remediation, we did not find that it has a statistically significant impact on the likelihood of passing the first college-level algebra course, earning a certificate or associate degree, or transfering to four-year public university in Florida when comparing outcomes for academically-equivalent students with scores on the margin of the cutoff. However, we find some evidence of a positive impact in terms of fall-to-fall persistence and in the overall credits earned over six years, but this statistical gap does not hold for credits that count toward a college degree.

When limiting the results to the no-retesting sample, estimates for remedial reading in Table 4 show different results. The negative estimates previously found (columns [1] to [6]) now move toward zero. For example, the estimated impacts for passing English composition are still negative and statistically significant but now range between a 3 and 5 percentage point difference, much smaller than before. More importantly, the negative statistically significant impact for associate degree completion and transfer to a four-year college vanishes for this subsample where the RD identifying assumption holds, especially when limiting the sample to the narrow band of 20 points around the cutoff. As was mentioned above, the density of the reading exam is centered near the cutoff point, and students seem to be more likely to sort themselves just above the cutoff as judged by the larger discontinuity found in Figure 2 and the results of the McCrary test in Table 5. The direction of this change suggests that previous estimates of the impact of reading remediation were biased downward, a result consistent with a positive correlation between re-taking the exam and the outcomes of interest. The conclusions for other outcomes hold: we find no impact for fall-to-fall persistence and earning a certificate, and remedial students earn more credits overall but not credits that count toward a degree.

5. Conclusions and Implications

This study provides a comprehensive evaluation of postsecondary remediation in a large, important state system that reflects broader national trends in remediation policy and student diversity. The study addresses limitations in the previous literature by first using a quasi-experimental regression discontinuity (RD) research design on a sample of nearly 100,000 students at the 28 community colleges in Florida. We discuss the application of the RD approach, address the issue of noncompliance, and implement a solution to deal with concerns about endogenous sorting around the policy cutoff. The application of the RD design is particularly beneficial to the study of college remediation. Moreover, our application of techniques to deal with threats to the assumptions of a sharp RD design could also help inform other research using the approach. This study also contributes additional evidence on the effectiveness of postsecondary remediation. While remedial education is a major investment at many colleges and universities, the literature provides very little information about the causal impact of remedial courses, and much of the recent evidence has been conflicting.

The results of this study suggest that remediation has both benefits and drawbacks as a strategy to address the needs of underprepared students. After controlling for noncompliance and endogenous sorting around the placement test cutoff score, students on the margin of requiring math remediation were slightly more likely to persist to their second year with estimates suggesting a 2.0 to a 3.8 percentage point difference. Similarly, the impacts of both math and reading remediation were positive in terms of the total (remedial and college-level) credits earned over six years. After dealing with endogenous sorting, our best estimates (Table 3 & 4, column 10) suggest that students in math and reading remediation earned 7.2 and 2.8 more credits than non-remedial students, respectively. However, no effect was found on total college-level (non-remedial) credits completed. Meanwhile, the likelihood of passing subsequent college-level English composition was slightly lower for reading remedial students while no difference was found in future math course performance for math remedial students. No discernable impact was found in terms of certificate or associate degree completion or transfer to a public four-year college. Overall, the results suggest that remediation might promote early persistence in college, but it does not necessarily help students on the margin of passing the cutoff to make progress toward a degree.

By studying a large, diverse student group and providing information on several outcomes not previously examined, this paper gives a larger perspective on the impacts of remediation than previous work and reconciles some of the mixed results found in other causal studies. Although much more positive effects were found in Ohio (Bettinger & Long, forthcoming), we also find that remediation appears to increase student persistence, but similar to the study on students in Texas (Martorell & McFarlin, 2007), we find that this increased persistence has only a minimal impact on degree completion. The differences that do exist in the effects across these studies may be partly due to the different student populations under analysis. For example, this study includes nearly the entire universe of first-time degree-seeking students in Florida, while Bettinger and Long (forthcoming) focused on traditional-age college students who were allowed to complete their remediation at either two- or four-year public institutions. Additionally, states differ in where they locate the cutoff for placement into remediation, and so this is likely to generate slightly different populations of "students on the margin of passing the cutoff." As all three studies (Florida, Ohio, and Texas) all focus on this marginal student, differences in the cutoff could potentially explain differences found in the results.

The results suggest that the costs of remediation should be given careful consideration in light of the limited benefits. While there may be an initial return in terms of the increased likelihood of persistence, under the current design and implementation of remedial programs, it is questionable whether the additional costs to students, institutions, and the state are justified given that little to no effect has been found in terms of degree completionfor students near the cutoff placement. As noted above, students who require remediation incur additional monetary and opportunity costs, and in Florida community college students who required remediation paid an additional $504 for college prep coursework during their first year (OPPAGA, 2006, p. 4). However, because even a year of college without completing a degree has a return, the investment in remediation may not be wasted. Additional research is needed to carefully examine the full scope of costs and benefits. Moreover, by increasing early persistence, remediation may give colleges an opportunity to reach students with other types of programming and skill development that might keep them progressing toward a degree and other long-term benefits.

It is worth emphasizing that the research design we used only allows the identification of the effect of remediation on a subset of students who scored just above and just below the cutoff score. Estimates should not be extrapolated to students with academic skills so weak that they scored significantly below the cutoff point. Moreover, our analysis is a "black box" evaluation of the effectiveness of remediation in Florida. Successful interventions for specific remediation programs might already be in place at certain institutions, but unfortunately our data do not contain the necessary information to link remedial students to specific interventions.

The results also have important policy implications about the institutional implementation of remedial placement procedures. The analysis provides evidence that, although a state may have a common placement exam and statewide cutoff scores, the *actual* implementation of these policies could differ at the institutional level. In the case of Florida, mandated assignment to remedial courses and actual remedial enrollment rates differed at most institutions, especially below the cutoff. A surprising number of students with assessments below those necessary to be exempt from remediation did not in fact enroll in the courses and instead directly entered college-level courses in the relevant fields.

State Departments of Education should explore this issue of noncompliance and consider the potential consequences of this practice. States could focus on creating better mechanisms to enforce statewide placement rules at each institution. Alternatively, policymakers could reconsider whether the current set of placement cutoffs accurately reflects the preparedness levels institutions deem suitable for deciding which students are ready for college-level material. Moreover, given the evidence presented in this paper, it may be the case that students who do not comply with the placement policy are actually saving themselves the costs of remediation while losing little in long-term benefits. By examining institutional practices more closely, states could gain much more understanding about whether it would be better to focus time and resources on enforcing compliance or to reconsider the remediation courses or affiliated programs offered to remedial students.

This study also documents the fact that retesting practices are not standard across the state nor even across remedial subject areas (retesting is more common for reading). The likelihood of allowing a student to retake the placement exam differs substantially by institution. As a result, the ability to routinely retest students at some institutions may threaten the validity of the test as a tool for accurate placement. Moreover, a policy that allows retests effectively lowers the relevant cutoff score and thereby weakens the original policy intent. It is also worth noting that the likelihood of retaking the remedial assessment appears to differ by student background (as shown in Figure 2). This suggests that the enforcement of placement policy differs by student group, thereby stoking concerns about equity across groups. To deal with this concern, states should consider explicit rules concerning retesting policies. Additionally, they should collect information on retesting by including in their databases all placement test scores, the number of attempts, and the time elapsed between each attempt. This would allow one to assess the implications of different retesting policies.

Besides providing a statewide evaluation of remedial programs in higher education, this study reveals several methodological issues that should be considered for further research. Researchers using quasi-experimental methods such as an RD design should be aware of multiple potential sources of bias that might invalidate the underlying assumptions of the statistical model (McCrary, 2008; Lee, 2008). As noted above, noncompliance and retesting (i.e., endogenous sorting around the policy cutoff) are serious concerns likely to appear in the postsecondary remediation context as well as other research settings. Non-experimental techniques such as instrumental variables can be used to deal with noncompliance. We suggest that endogenous sorting be analyzed case-by-case, although a non-parametric estimation of density functions for the assignment variable can help to identify potential manipulation in any evaluation setting. Researchers should also conduct robustness checks by using available covariates as well as by focusing narrowly around the cutoffs.

While we have extended the research on postsecondary remediation through this study, additional effort is needed to estimate the impact of remedial courses on weaker students who are not necessarily close to the placement cutoff. Additionally, more work is needed on the effects of remediation relative to its costs. Future research should also focus on institutional policies, practices, additional services, and classroom strategies in order to explore differences in the effects of remediation by college and by particular ways of conducting remediation programs. It would be extremely useful to identify institutional characteristics and innovative approaches that appear to improve the success of remedial students and to evaluate them using rigorous research designs.

References

Aiken, L., West, S., Schwalm, D., Carroll, J., and Hsiung, S. (1998). Comparison of a randomized and two quasi-experiments in a single outcome evaluation: Efficacy of a university-Level remedial writing program. *Evaluation Review, 22*(2), 207-244.

Angrist, J. (2001). Estimation of limited-dependent variable models with binary endogenous regressors: Simple strategies for empirical practice. *Journal of Business and Economic Statistics*, *19*(1), 2-28.

Angrist, J., & Imbens, G. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association, 90*(430), 431-442.

Angrist, J., Imbens, G., & Rubin, D. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444-472.

Angrist, J., & Lavy, V. (1999). Using Maimondies' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics, 114*(2), 533-575.

Attewell, P., Lavin, D., Domina, T., & Levey, T. (2006). New evidence on college remediation. *Journal of Higher Education, 77*(5), 886-924.

Battistin, E., & Rettore, E. (2002). Testing for programme effects in a regression discontinuity design with imperfect compliance. *Journal of the Royal Statistical Society, Series A*, *165*(1), 39-57.

Bettinger, E., & Long, B. (2004). *Shape up or ship out: The effects of remediation on students at four-year colleges* (NBER Working Paper No. 10369). Cambridge, MA: National Bureau of Economic Research.

Bettinger, E., & Long, B. (2005). Remediation at the community college: Student participation and outcomes. *New Directions for Community Colleges*, *129*(1), 17-26.

Bettinger, E., & Long, B. (2007). Institutional responses to reduce inequalities in college outcomes: Remedial and developmental courses in higher education. In S. Dickert-Conlin & R. Rubenstein, (Eds.) *Economic Inequality and Higher Education: Access, Persistence and Success* (pp. 69-100). New York: Russell Sage Foundation.

Bettinger, E., & Long, B. (forthcoming). Addressing the needs of under-prepared college students: Does college remediation work? *Journal of Human Resources*.

Bloom, H. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review, 8*(2), 225-46.

Boylan, H., & Saxon, D. (1999). What works in remediation: Lessons from 30 years of research. [Prepared for The League for Innovation in the Community College.] Retrieved April 7, 2005, from http://www.ncde.appstate.edu/reserve_reading/what_works.htm

Breneman, D., & Haarlow, W. (1998). *Remedial education: Costs and consequences.* Paper presented at Remediation in Higher Education: A Symposium, Washington, DC.

Calcagno, J. C. (2007). *Evaluating the impact of developmental education in community colleges: A quasi-experimental regression-discontinuity design*. Ph.D. dissertation, Columbia University, Graduate School of Arts and Sciences.

Campbell, D. (1969). Reforms as experiments. *American Psychologist*, *24*(4), 409-429.

College Board. (2003). *ACCUPLACER online: Technical manual.* New York: College Board.

Deil-Amen, R., & Rosenbaum, J. (2002). The unintended consequences of stigma-free remediation. *Sociology of Education*, *75*(3), 249-268.

Dougherty, K. (1994). *The contradictory college: The conflicting origins, impacts, and futures of the community college.* Albany: State University of New York Press.

Gennetian, L., Morris, P., Bos, J., & Bloom, H. (2005). Constructing instrumental variables from experimental data to explore how treatments produce effects. In H. Bloom (Ed.), *Learning more from social experiments: Evolving analytic approaches* (pp. 75-114). New York: Russell Sage Foundation.

Greene, J., & Foster, G. (2003, September). *Public high school graduation and college readiness rates in the United States* (Manhattan Institute, Center for Civic Information, Education Working Paper, No. 3). New York: Manhattan Institute.

Grubb, N. (2001). *From black box to pandora's box: Evaluating remedial/developmental education.* New York: Columbia University, Teachers College, Community College Research Center.

Hahn, J., Todd, P., & van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica, 69*(1), 201-209.

Heckman, J., LaLonde, R., & Smith, J. (1999). The economics and econometrics of active labor market programs. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics, Volume 3A* (pp. 1865-2073). Amsterdam: Elsevier Science.

Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945-970.

Imbens, G., & Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica, 62*(2), 467-476.

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

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

Jaeger, D., & Page, M. (1996). Degrees matter: New evidence on sheepskin effects in the returns to education. *Review of Economics and Statistics*, *78*(4), 733-740.

Jenkins, D., & Boswell, K. (2002). *State policies on community college remedial education: Findings from a national survey*. Denver: Education Commission of the States.

Kane, T., & Rouse, C. (1999). The community college: Educating students at the margin between college and work. *Journal of Economic Perspectives, 13*(1), 63-84.

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

Lee, D., & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, *142*(2), 655-674.

Lesik, S. (2006). Applying the regression-discontinuity design to infer causality with non-random assignment. *Review of Higher Education*, *30*(1), 1-19.

Lesik, S. (2007). Do developmental mathematics programs have a causal impact on student retention? An application of discrete-time survival and regression-discontinuity analysis. *Research in Higher Education*, *48*(5), 566-591.

Levin, H., & Calcagno, J. C. (2008). Remediation in the community college: An evaluator's perspective. *Community College Review*, *35*(3), 181-207.

Martorell, P., & McFarlin, I. (2007). *Help or hindrance? The effects of college remediation on academic and labor market outcomes.* Unpublished manuscript.

McCabe, R. (2000). *No one to waste: A report to public decision-makers and community college leaders.* Washington, DC: Community College Press.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), 698-714.

McCrary, J., & Royer, H. (2006). *The effect of female education on fertility and infant health: Evidence from school entry policies using exact date of birth* (NBER Working Paper No. 12329). Cambridge, MA: National Bureau of Economic Research.

Merisotis, J., & Phipps, R. (2000). Remedial education in colleges and universities: What's really going on? *Review of Higher Education*, *24*(1), 67-85.

Moss, B. G., & W. Yeaton. 2006. Shaping policies related to developmental education: An evaluation using the regression-discontinuity design. *Educational Evaluation and Policy Analysis*, *28*(3), 215-229.

National Center for Education Statistics [NCES]. (2003). *Remedial education at degree-granting postsecondary institutions in fall 2000.* Washington, DC: Department of Education.

National Center for Education Statistics [NCES]. (2004). *Digest of education statistics, 2003*. Washington, DC: U.S. Department of Education.

Office of Program Policy and Government Accountability [OPPAGA]. (2006). *Steps can be taken to reduce remediation rates* (Report 06-40). Tallahassee: Florida Legislature.

O'Hear, M., & MacDonald, R. (1995). A critical review of research in developmental education, Part I. *Journal of Developmental Education, 19*(2), 2-6.

Perin, D. (2006). Can community colleges protect both access and standards? The problem of remediation. *Teachers College Record, 108*(3), 339-373.

Reichardt, C., Trochim, W., & Cappelleri, J. (1995). Reports of the death of the regression-discontinuity design are greatly exaggerated. *Evaluation Review, 19*(1), 39-63.

Rosenbaum, J. (2001). *Beyond college for all*. New York: Russell Sage Foundation.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology, 66,* 688-701.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton-Mifflin.

Trochim, W. (1984). *Research design for program evaluation: The regression-discontinuity approach*. Newbury Park, CA: Sage.

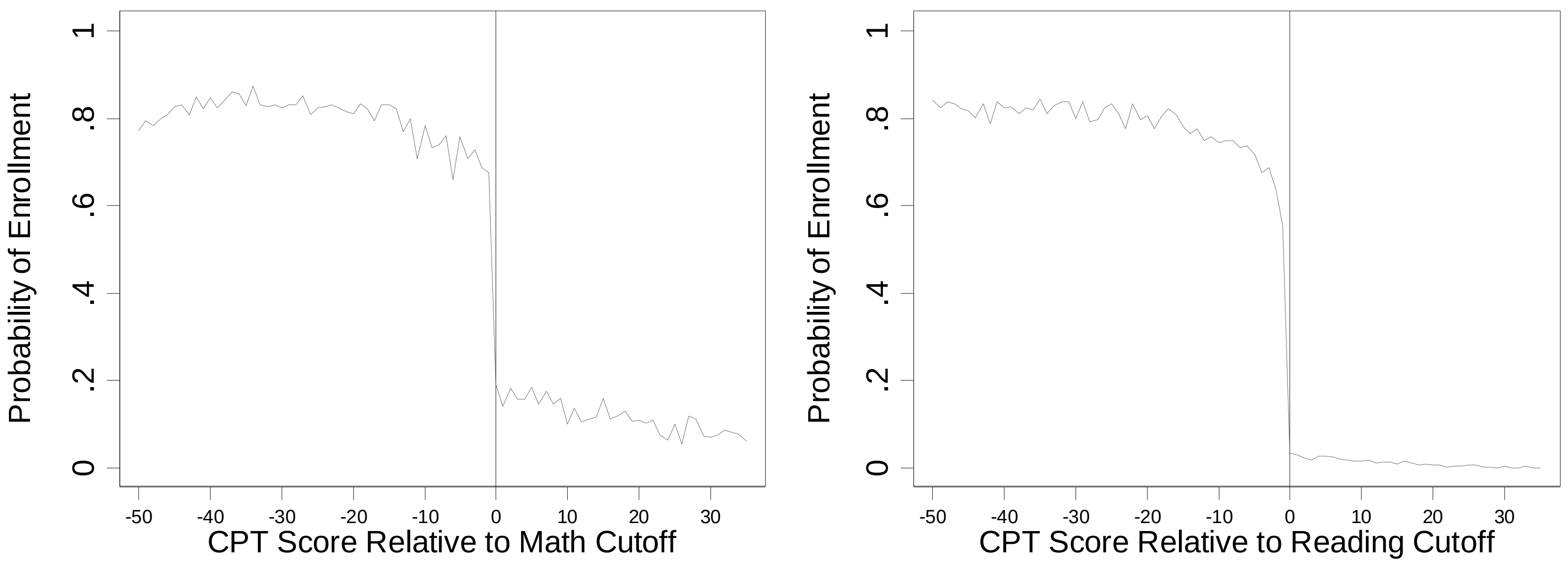
Urquiola, M., & Verhoogen, E. (2007). *Class size and sorting in market equilibrium: Theory and evidence* (NBER Working Paper No. 13303). Cambridge, MA: National Bureau of Economic Research.

Van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment decisions: A regression-discontinuity approach. *International Economic Review, 43*(4), 1249-1287.

Windham, P. (2005). Developmental education in Florida community colleges. *Program Review 2005-05.* Tallahassee: Florida Department of Education.

Appendix: Figures and Tables

Figure 1: Probability of Enrollment in Remedial Math and Reading by College Placement Test (CPT) Score



Notes: Each graph corresponds to a different remedial subject. The lines join together the mean probability of enrollment in remediation for students with a given CPT score.

Figure 2: College Placement Test (CPT) Distributions by Subject and Race/Ethnicity

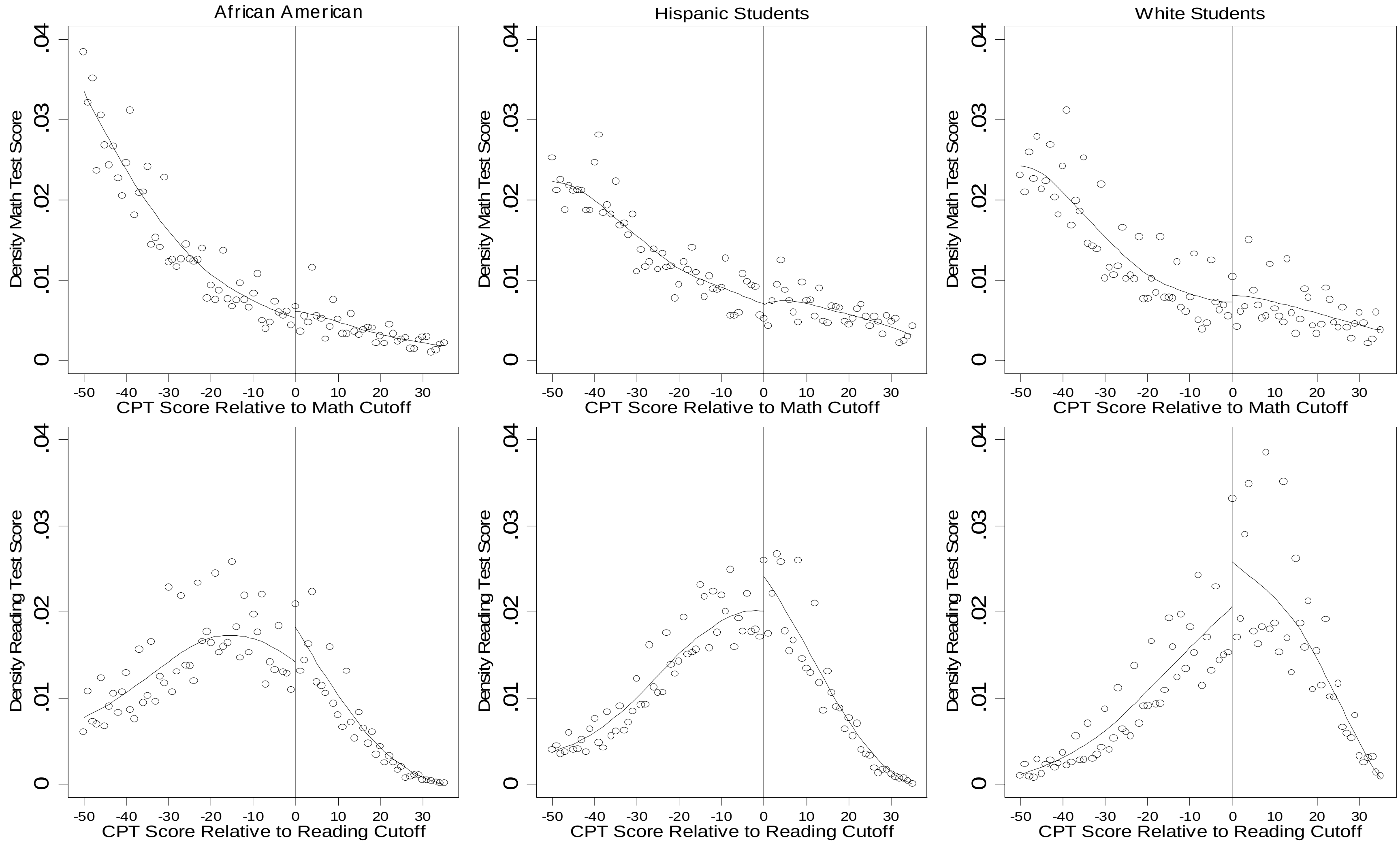
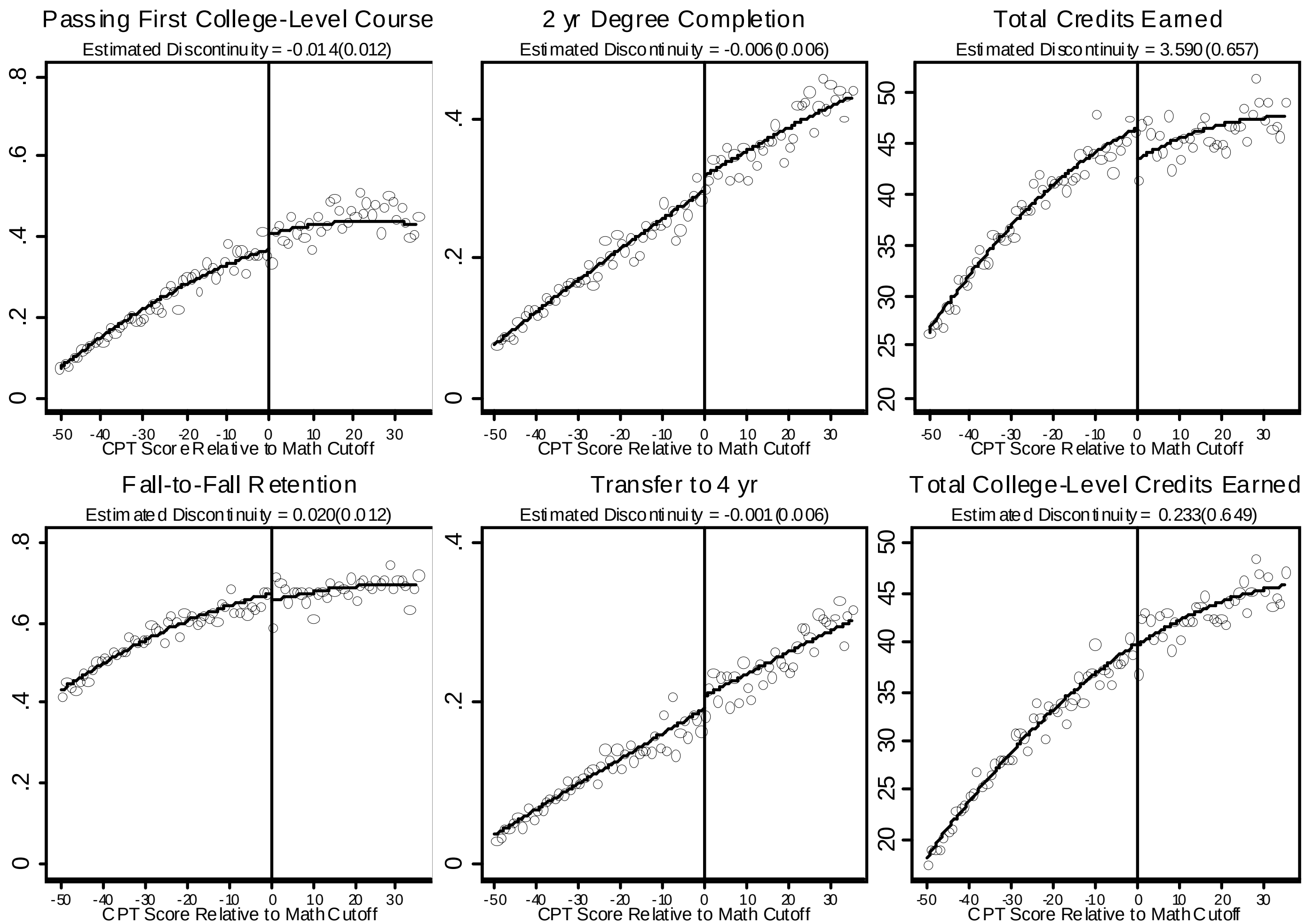
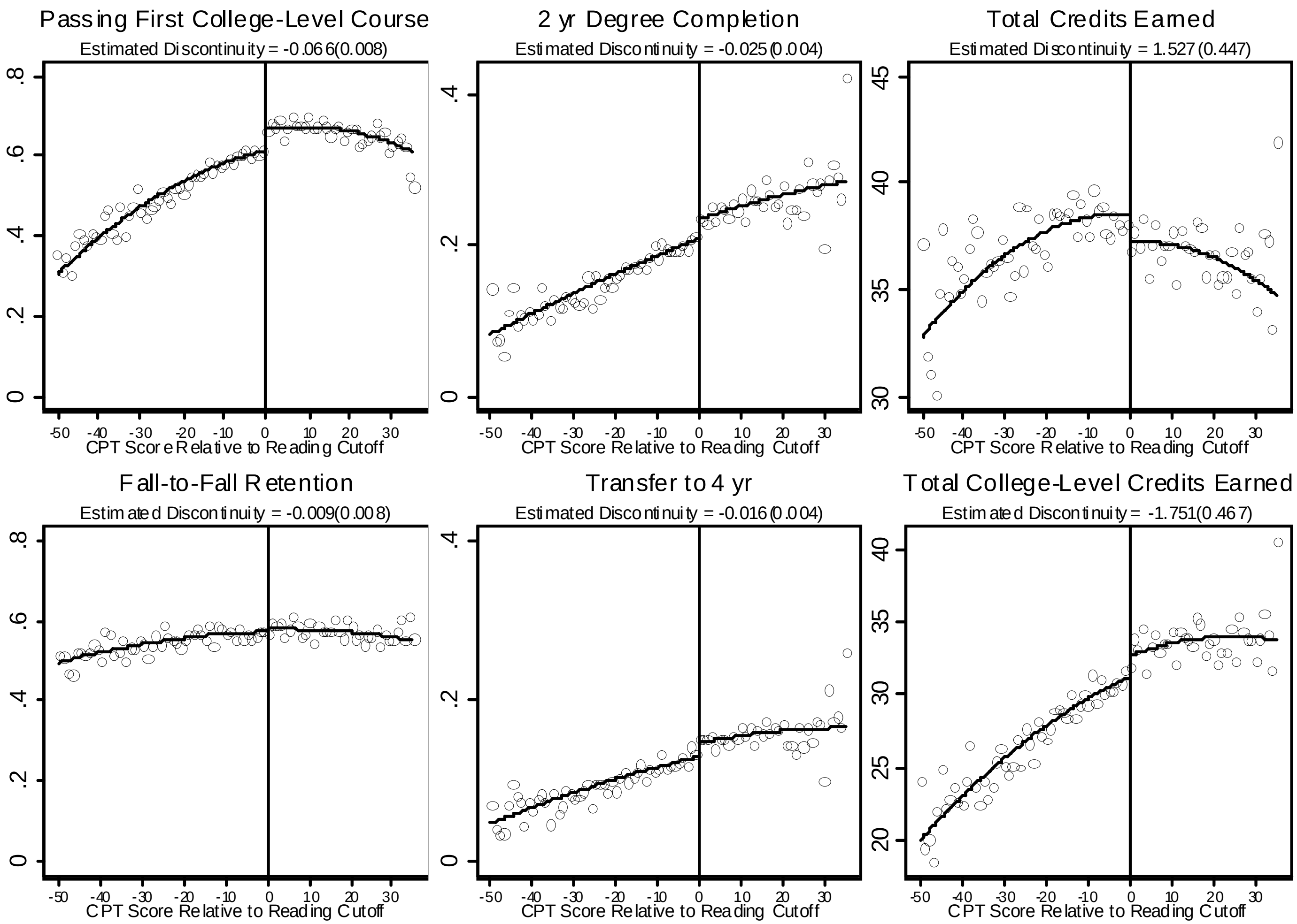


Figure 3: Educational Outcome by Math CPT Score and Estimated Discontinuity



Notes: Each graph corresponds to a different educational outcome. The circles are the mean of the binary dependent variable for students with a given CPT score. The fitted lines are predicted probabilities from a linear probability model for the educational outcome on the assignment to treatment variable and quadratic polynomial terms in the CPT score. Estimated effects around the discontinuities are shown as the baseline intention-to-treat (ITT) estimates in Table 3.

Figure 4: Educational Outcome by Reading CPT Score and Estimated Discontinuity



Notes: Each graph corresponds to a different educational outcome. The circles are the mean of the binary dependent variable for students with a given CPT score. The fitted lines are predicted probabilities from a linear probability model for the educational outcome on the assignment to treatment variable and quadratic polynomial terms in the CPT score. Estimated effects around the discontinuities are shown as the baseline intention-to-treat (ITT) estimates in Table 4.

Table 1: Descriptive Statistics – Entering Community College Students (fall 1997 to 2000)

Variable	Full Sample	Research Sample	Restesting *Not* Allowed (Math)	Restesting *Not* Allowed (Reading)
Demographics				
Age	20.10	20.89	20.95	21.13
Female	0.54	0.54	0.55	0.54
African-American	0.16	0.19	0.21	0.22
Hispanic	0.18	0.19	0.22	0.13
Asian	0.03	0.03	0.03	0.03
U.S. Citizen	0.89	0.87	0.85	0.86
Limited English Proficiency	0.05	0.06	0.05	0.08
Began Fall 1997	0.23	0.23	0.25	0.25
Began Fall 1998	0.25	0.25	0.26	0.24
Began Fall 1999	0.26	0.25	0.24	0.25
Began Fall 2000	0.27	0.26	0.25	0.26
Test Scores and Remedial Placement				
Math CPT Score (range 20-120) [98,370 observations]	46.14 (27.98)	46.14 (27.98)	46.34 (27.71)	43.22 (27.40)
Reading CPT Score (range 20-120) [98,370 observations]	77.16 (19.55)	77.16 (19.55)	76.47 (19.74)	76.42 (20.18)
SAT Math Score (range 200-800) [15, 745 observations]	489.56 (75.92)	---	---	---
SAT Verbal Score (range 200-800) [15, 745 observations]	489.63 (75.05)	---	---	--
ACT Math Score (range 1-36) [16,747 observations]	18.95 (3.42)	---	---	---
ACT Reading Score (range 1-36) [16,747 observations]	20.73 (4.62)	---	---	--
Recmd. for Math Remediation	0.61	0.79	0.80	0.83
Recmd. for Reading Remediation	0.43	0.55	0.57	0.57
College Outcomes				
Passed 1st College Course (Math)	0.30	0.24	0.23	0.22
Passed 1st College Course (Reading)	0.64	0.59	0.59	0.57
Fall-to-Fall (one year) Persistence	0.61	0.56	0.56	0.55
Two-Year Degree Completion	0.27	0.20	0.20	0.18
Transfer to a Four-Year University	0.18	0.13	0.12	0.12
Total Credits Completed	40.40 (32.16)	37.11 (32.62)	36.51 (31.93)	34.18 (31.14)
Total Non-Remedial Credits Earned	34.73 (30.20)	30.18 (29.63)	30.01 (29.61)	27.49 (29.06)
Number of Observations	130,862	98,370	68,337	24,151

Notes: Standard deviations are shown in parentheses. The Research Sample contains all degree-seeking students who took the CPT taker and enrolled in a Florida community college between fall 1997 and fall 2000.

Table 2: Descriptive Statistics by Remedial Subject: Group Means and Group Differences

Variable	Band around cutoff (all range)			Band around cutoff (+/-10)			Band around cutoff (+/-5)		
	All below	All above	Difference	(-10 to -1)	(0 to 9)	Difference	(-5 to -1)	(0 to 4)	Difference
MATHEMATICS									
Age	21.28	19.36	1.924*	19.33	19.25	0.086	19.38	19.19	0.189
Female	0.563	0.496	0.067*	0.54	0.53	0.011	0.536	0.539	-0.003
African-American	0.209	0.120	0.089*	0.16	0.14	0.017	0.145	0.151	-0.006
Hispanic	0.194	0.199	-0.006	0.22	0.19	0.028	0.229	0.190	0.039*
Asian	0.021	0.053	-0.031*	0.03	0.04	-0.005	0.035	0.038	-0.004
American Indian	0.005	0.004	0.001*	0.00	0.00	0.001	0.004	0.004	0.001
U.S. Citizen	0.891	0.821	0.07*	0.85	0.87	-0.015*	0.853	0.867	-0.014
Limited English Proficiency	0.056	0.083	-0.027*	0.06	0.06	0.001	0.065	0.063	0.001
Number of Observations	74,295	22,863		7,177	7,390		3,700	3,876	
READING									
Age	20.27	21.65	-1.379*	20.326	20.689	-0.363*	20.48	20.62	-0.146
Female	0.561	0.515	0.046*	0.547	0.545	0.002	0.544	0.555	-0.011
African-American	0.263	0.101	0.162*	0.172	0.129	0.043	0.162	0.138	0.024*
Hispanic	0.222	0.157	0.065*	0.221	0.187	0.034	0.222	0.189	0.033*
Asian	0.036	0.021	0.015*	0.028	0.024	0.004	0.026	0.025	0.001
American Indian	0.004	0.005	-0.001*	0.005	0.005	0.000	0.005	0.004	0.000
U.S. Citizen	0.837	0.918	-0.08*	0.876	0.906	-0.029*	0.879	0.902	-0.023*
Limited English Proficiency	0.074	0.046	0.028*	0.054	0.047	0.007	0.054	0.048	0.006
Number of Observations	54,085	44,283		16,736	21,171		7,900	11,839	

* Denotes significant difference at 1 percent level, two-tailed test, unequal variances.
Notes: The sample contains all degree-seeking students who took the CPT taker and enrolled in a Florida community college between fall 1997 and fall 2000.

Table 3: Impact of Math Remediation on Educational Outcomes

	All Students				Narrow Band Sample		No-Retesting Sample		No-Retesting & Narrow Band Sample	
	Without Controls		With Controls							
	ITT	RD-IV	ITT	RD-IV	ITT	RD-IV	ITT	RD-IV	ITT	RD-IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Completion of First College-Level Course	-0.014	-0.022	-0.018	-0.030	0.006	0.012	-0.011	-0.018	-0.001	-0.002
	(0.012)	(0.020)	(0.011)	(0.019)	(0.029)	(0.057)	(0.011)	(0.018)	(0.031)	(0.064)
Fall-to-Fall Persistence	0.020	0.035	0.014	0.026	0.004	0.008	0.020*	0.038*	0.007	0.015
	(0.012)	(0.021)	(0.011)	(0.019)	(0.032)	(0.062)	(0.009)	(0.018)	(0.025)	(0.051)
Earning a Certificate	-0.004	-0.006	-0.003	-0.006	-0.004	-0.008	-0.004	-0.007	-0.007	-0.014
	(0.002)	(0.004)	(0.002)	(0.004)	(0.005)	(0.009)	(0.003)	(0.004)	(0.006)	(0.012)
Associate Degree Completion	-0.006	-0.010	-0.006	-0.011	-0.016	-0.032	0.003	0.005	-0.014	-0.027
	(0.006)	(0.011)	(0.006)	(0.011)	(0.016)	(0.031)	(0.007)	(0.012)	(0.012)	(0.025)
Transfer to 4-year University (SUS)	-0.001	-0.002	-0.003	-0.005	-0.022	-0.043	-0.003	-0.006	-0.033	-0.067
	(0.006)	(0.010)	(0.006)	(0.010)	(0.017)	(0.033)	(0.006)	(0.010)	(0.018)	(0.037)
Total Credits Earned	3.590**	6.169**	3.290**	5.690**	3.797*	7.453*	3.741**	5.930**	3.515*	7.282*
	(0.657)	(1.099)	(0.613)	(1.023)	(1.698)	(3.425)	(0.650)	(1.025)	(1.621)	(3.252)
Total Non-Remedial Credits Earned	0.233	0.400	0.011	0.019	1.398	2.744	0.884	1.204	-0.118	-0.244
	(0.649)	(1.113)	(0.596)	(1.031)	(1.836)	(3.622)	(0.578)	(0.954)	(1.759)	(3.641)
Institutions	28	28	28	28	28	28	19	19	19	19
Observations (students)	96,724	96,724	96,724	96,724	14,493	14,493	68,337	68,337	9,593	9,593

* Significant at 5%. ** Significant at 1%.

Notes: Each row focuses on a different outcome, with each cell corresponding to a different method that is designated by the column heading. For the binary outcomes, we use the maximum likelihood probit method to estimate models, and we report the marginal effects at mean values. For the continuous dependent variables, we estimate OLS models. ITT is the intention-to-treat estimate from equation (2). RD-IV is the instrumental variable estimate from equation (4). Columns (1) and (2) show the baseline ITT and RD-IV impacts, and columns (3) and (4) add controls for age, gender, race/ethnicity, citizenship, English limited proficiency, test score in the opposite subject, and cohort fixed effects (all other specifications also include controls). In columns (5) and (6) we estimate our models on students with test scores within a 20 points band around the cutoff. Columns (7) and (8) include estimates that are robust to the retesting problem. Columns (9) and (10) combine no-retesting colleges and the narrow band sample.

Table 4: Impact of Reading Remediation on Educational Outcomes

	All students				Narrow Band Sample		No-Retesting Sample		No-Retesting & Narrow Band Sample	
	Without Controls		With Controls							
	ITT	RD-IV	ITT	RD-IV	ITT	RD-IV	ITT	RD-IV	ITT	RD-IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Completion of First College-Level Course	-0.066**	-0.095**	-0.060**	-0.086**	-0.053**	-0.090**	-0.039**	-0.049**	-0.028*	-0.036*
	(0.008)	(0.012)	(0.008)	(0.012)	(0.009)	(0.017)	(0.012)	(0.016)	(0.013)	(0.017)
Fall-to-Fall Persistence	-0.009	-0.012	-0.003	-0.003	-0.017	-0.029	-0.005	-0.006	-0.009	-0.013
	(0.008)	(0.011)	(0.008)	(0.011)	(0.010)	(0.017)	(0.014)	(0.019)	(0.018)	(0.028)
Earning a Certificate	-0.002	-0.003	-0.002	-0.002	-0.004	-0.007	0.002	0.002	-0.003	-0.005
	(0.002)	(0.003)	(0.002)	(0.003)	(0.004)	(0.007)	(0.002)	(0.003)	(0.005)	(0.008)
Associate Degree Completion	-0.025**	-0.037**	-0.020**	-0.029**	-0.024**	-0.040**	-0.022**	-0.031**	-0.020	-0.031
	(0.004)	(0.006)	(0.004)	(0.006)	(0.009)	(0.014)	(0.008)	(0.010)	(0.017)	(0.026)
Transfer to 4-year University (SUS)	-0.016**	-0.024**	-0.009*	-0.013*	-0.015*	-0.025*	-0.005	-0.008	-0.004	-0.005
	(0.004)	(0.005)	(0.004)	(0.006)	(0.007)	(0.011)	(0.008)	(0.011)	(0.016)	(0.022)
Total Credits Earned	1.527**	2.266**	2.048**	3.025**	0.854	1.437	2.370**	3.178**	1.858	2.889
	(0.447)	(0.647)	(0.461)	(0.653)	(0.496)	(0.818)	(0.682)	(0.912)	(1.158)	(1.740)
Total Non-Remedial Credits Earned	-1.751**	-2.599**	-1.190**	-1.758**	-1.182	-2.159	-1.662**	-2.225**	-0.935	-1.590
	(0.467)	(0.685)	(0.431)	(0.636)	(0.684)	(1.271)	(0.563)	(0.749)	(1.252)	(2.124)
Institutions	28	28	28	28	28	28	7	7	7	7
Observations (students)	97,938	97,938	97,938	97,938	37,747	37,747	24,151	24,151	8,775	8,775

* Significant at 5%. ** Significant at 1%.

Notes: Each row focuses on a different outcome, with each cell corresponding to a different method that is designated by the column heading. For the binary outcomes, we use the maximum likelihood probit method to estimate models, and we report the marginal effects at mean values. For the continuous dependent variables, we estimate OLS models. ITT is the intention-to-treat estimate from equation (2). RD-IV is the instrumental variable estimate from equation (4). Columns (1) and (2) show the baseline ITT and RD-IV impacts, and columns (3) and (4) add controls for age, gender, race/ethnicity, citizenship, English limited proficiency, test score in the opposite subject, and cohort fixed effects (all other specifications also include controls). In columns (5) and (6) we estimate our models on students with test scores within a 20 points band around the cutoff. Columns (7) and (8) include estimates that are robust to the retesting problem. Columns (9) and (10) combine no-retesting colleges and the narrow band sample.

Table 5: McCrary Manipulation Test per Institution Log Discontinuity Estimates

Instit.	Math					Reading				
	Band-width	Bin size	Theta	Std. Error	T-Test	Band-width	Bin size	Theta	Std. Error	T-Test
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
A	26	1	0.264	0.135	**1.959**	27	1	0.283	0.076	3.706
B	25	1	0.051	0.082	**0.628**	8	1	0.159	0.087	**1.815**
C	33	1.338	0.082	0.142	**0.581**	38	1	0.567	0.084	6.741
D	37	2.365	1.267	0.311	4.072	23	1.750	0.470	0.235	2.001
E	31	1	0.255	0.115	2.212	64	1	0.775	0.046	16.855
F	44	1	0.123	0.145	**0.848**	36	1	0.151	0.074	2.033
G	30	1	0.134	0.085	**1.587**	23	1	0.232	0.054	4.326
H	30	3.733	0.404	0.541	**0.746**	20	3.382	0.806	0.414	**1.947**
I	31	1	0.285	0.153	**1.864**	35	1	0.772	0.073	10.550
J	61	1	0.879	0.067	13.183	31	1	0.747	0.059	12.669
K	31	1	0.942	0.172	5.488	27	1	0.525	0.101	5.173
L	40	1.963	0.313	0.201	**1.558**	24	1.335	0.765	0.165	4.631
M	67	1.602	-0.155	0.127	**-1.215**	30	1.074	0.732	0.120	6.091
N	49	1	0.170	0.121	**1.405**	37	1	0.152	0.083	**1.837**
O	28	1	-0.142	0.055	**-2.584**	15	1	0.217	0.045	4.864
P	24	2.515	-0.583	0.455	**-1.280**	25	2.270	-0.100	0.293	**-0.340**
Q	38	1	1.039	0.156	6.658	31	1	1.074	0.106	10.112
R	22	1	0.133	0.119	**1.115**	17	1	0.385	0.078	4.947
S	46	1	0.587	0.147	3.996	28	1	0.674	0.096	6.997
T	27	1	0.481	0.158	3.047	23	1	0.330	0.084	3.954
U	27	1	0.127	0.117	**1.082**	26	1	0.169	0.067	2.539
V	41	1.101	0.296	0.165	**1.797**	29	1	0.029	0.099	**0.295**
W	35	1	0.373	0.085	4.410	22	1	0.416	0.057	7.269
X	26	1	0.000	0.125	**0.002**	28	1	0.052	0.074	**0.694**
Y	39	1	0.079	0.120	**0.659**	36	1	0.172	0.072	2.395
Z	38	2.300	0.541	0.299	**1.807**	29	1.316	1.939	0.202	9.590
AA	40	1	0.220	0.116	**1.894**	60	1	0.105	0.055	**1.919**
BB	31	1	0.290	0.063	4.625	30	1	0.305	0.040	7.564

Notes: Each row represents a community college in Florida. For each subject (math and reading), column (1) is the estimated bandwidth *h*, column (2) is the estimated bin size that is used for each institution, columns (3) and (4) are the estimated discontinuity theta and its standard error, respectively, and these last two parameters are combined to compute the t-test in Column (5). T-test values lower than 1.96 (bolded) are associated with a 5 percent level of significance, indicating that there is no statistical evidence of a discontinuity in the CPT distribution at the cutoff.