Help or Hindrance? The Effects of College Remediation on Academic and Labor Market Outcomes

– Working Paper Version with Appendices –

Paco Martorell*
RAND
paco@rand.org

Isaac McFarlin Jr.
University of Michigan
imcfar@umich.edu

April 2010

Abstract

Providing remedial (also known as developmental) education is the primary way colleges cope with students who do not have the academic preparation needed to succeed in college-level courses. Remediation is widespread, with nearly one-third of entering freshman taking remedial courses at a cost of at least $1 billion per year. Despite its prevalence, there is uncertainty surrounding its short- and longer-run effects. This paper presents new evidence on this question using longitudinal administrative data from Texas and a regression discontinuity research design. We find little indication that remediation improves academic or labor market outcomes.

* The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B07581 to the University of Texas at Dallas. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. The authors also gratefully acknowledge support from the Smith Richardson Foundation and W.E. Upjohn Institute. We would like to thank David Bjerk, Juan Carlos Calcagno, David Card, Kerwin Charles, Sheldon Danziger, John DiNardo, Janet Hansen, Peter Hinrichs, Paul Jargowsky, Chris Jepsen, Steve Rivkin, Karl Scholz, and numerous seminar participants for helpful comments. Greg Branch and Lee Holcombe were helpful in answering questions about the data. Tammy Leonard, Yu Xue, and Irene Ngugi provided valuable research assistance. All errors are our own.

1. Introduction

Many new college students arrive on campus lacking the preparation to successfully pursue postsecondary education. A recent study shows that only one-third of high school graduates possess the minimum qualifications for a four-year college (Greene and Foster, 2003). This widespread phenomenon is seen as a key reason why large numbers of college students, particularly those from disadvantaged backgrounds, drop out (Venezia et al., 2003). Given the sizable earnings premium associated with college attainment (Kane and Rouse, 1995), understanding how to help under-prepared students complete college is an important question.[1]

Remediation is the most common approach used by colleges to assist students who possess weak academic skills. Remedial – also known as developmental – education consists of courses and other services (such as tutoring) that are designed to foster skills generally acquired in high school. It is a central feature of American higher education, especially in two-year (or community) colleges. Among entering freshman in fall 2000, 30 percent overall and 42 percent of two-year college students participated in remediation (NCES, 2003).

Despite the prevalence of remediation, substantial controversy surrounds its use. Supporters contend that it helps poorly-prepared students succeed in college by allowing them a chance to catch up to their peers. According to this view, under-prepared students are better served in remedial courses than they would be floundering in college-level courses (Lazarick, 1997). In contrast, opponents argue that any benefits of remediation are outweighed by its high cost. Estimates from a decade ago suggest public colleges spent $1 billion per year on remediation ($172 million per biennium in Texas alone; Breneman and Haarlow, 1997), and some even argue that the costs are higher (Steinberg, 1998). Worries about cost partly explain why some states have cut funding for remediation programs (Bettinger and Long, 2009).

The disagreements over college remediation also fit into the debate in labor economics about policies to improve the human capital of low-skilled individuals. One side argues that little can be done to increase human capital of low-ability individuals once they reach a certain age (Carneiro and Heckman, 2003). According to this view, "skills beget skills" and effective interventions must therefore occur very early in life. Others maintain that some "second chance" programs (especially those aimed at individuals from disadvantaged backgrounds) have been successful even though the interventions did not take place at a young age (Krueger, 2003). Since college remediation is an important example of a later-life intervention, understanding whether

[1] It is important to recognize that dropping out of college can be optimal for some students (Manski, 1989). Nonetheless, as discussed in Angrist et al. (2009), poor performance in college might be a cause for concern if students misjudge the costs and benefits of leaving school (Dominitz and Manski, 2000) or if they have time-inconsistent preferences (Oreopoulos, 2007).

remediation actually helps students develop economically valuable skills is informative about which view of human capital formation is more accurate.

Currently, there is considerable uncertainty surrounding the effectiveness of remediation. Assessing the impact of remediation is difficult because remediated students would likely have worse outcomes than non-remediated students in the absence of the program. Three recent papers in the economics literature seriously address this challenge, but the evidence remains mixed. Bettinger and Long (2009) and Jepsen (2006) generally find positive effects of remediation on college persistence and attainment. However, using a sample of Florida community college students and a research design similar to that used in this paper, Calcagno and Long (2008) find little consistent evidence that remediation has positive effects. Moreover, as we explain below, all of these studies potentially face methodological limitations that make it difficult to interpret their findings.

This paper presents new evidence on the effect of remediation using a large longitudinal dataset of Texas students. Our regression discontinuity (RD) research strategy exploits the fact that during the time period of the study, Texas used placement test score results to assign students to remediation.[2] Consistent with this policy, we find clear evidence that students who barely failed this exam were more likely to be in remediation than were students who passed. This discontinuity generates exogenous variation in the probability of remediation that is used to generate instrumental variables (IV) estimates of the effect of remediation.

The administrative records that comprise our data contain rich information on several measures of student success in college such as academic credit hours, years of college completed, and degree attainment. In addition, we examine the impact of remediation on labor market earnings using Unemployment Insurance (UI) earnings records. Studying the labor market effects of remediation is important because it allows us to directly test whether remediation helps students become economically successful. To the best of our knowledge, no other study has examined the labor market effects of college remediation.

Our results provide little indication that students benefit from remediation. For a wide range of academic outcomes, and across a variety of subgroups, the estimated effects of remediation are small in magnitude and statistically insignificant. If anything, we find some evidence that remediation might worsen the outcomes of some students. We also find no

[2] A similar empirical strategy is used by Jacob and Lefgren (2004) to study the effect of remedial education in elementary school and by Matsudaira (2009) to study the effect of English as a Second Language programs.

evidence that remediation confers longer run economic benefits in the form of higher earnings, although these estimates are somewhat imprecise.

Although our empirical strategy delivers credible estimates of the impact of remediation, it is important to recognize that our estimates are most pertinent for students scoring close to the remediation-placement cutoff and whose participation in remediation is affected by falling above or below the cutoff. As with all discontinuity-based designs, our estimates might not be applicable to students scoring far away from the passing cutoff, and in particular the effect of remediation for the very lowest ability students could be different from what we report in this paper. However, the "marginal" group is interesting because a large fraction of remediated students score relatively close to the passing score. Furthermore, the passing cutoff changed during our study period, allowing us to examine effects at different points in the ability distribution. Since we find no evidence of positive effects of remediation under either cutoff, our findings may have greater external validity than is typical with a regression discontinuity design.

We begin with background information on college remediation in the following section. Section 3 describes the data we use. Section 4 presents our econometric strategy. Section 5 contains the results. Section 6 discusses our findings in terms of existing research. Section 7 concludes.

2. Background

2.1 What is College Remediation?

"College remediation" is an umbrella term that describes services provided by postsecondary institutions to help academically underprepared students succeed in college. Remediation in its most common form consists of coursework that mirrors a college-preparatory curriculum in reading, writing, and mathematics. In addition to covering less-advanced material, remedial courses generally do not confer degree-counting credits and are often taught by different instructors than the ones who teach college-level – also known as "academic" – courses.[3] Less formal types of instruction are also used. Information provided by several developmental education practitioners at two- and four-year institutions suggest that the types of non-course services offered across schools are remarkably similar. For instance, most schools offer free tutoring, learning assistance centers, and supplemental instruction. School officials indicated that these services are also available to students who are not in mandatory remediation. However, these services are typically supplemental and offered in conjunction with formal

[3] A survey of Texas public institutions found that over 70 percent college-level mathematics courses were taught by full-time faculty compared to 48 percent of remedial math courses (THECB, 2005)

coursework. For example, in our sample 90 percent of students in remediation for math during their first year were in course-based remediation, although informal remediation is somewhat more common in four-year schools and for reading and writing.

Assignment to remediation is mainly determined by performance on a placement test, although more individualized means such as referral from an advisor are also used. States vary widely in the rules governing assignment to remediation. Some states (such as Ohio, and as of 2003, Texas) grant institutions considerable latitude in the choice of passing standards and examinations used to make remedial placements (Abraham, 1986). Other states (such as Florida and prior to 2003, Texas) use a statewide test and passing standard.

Although our data do not contain information on particular organizational approaches or instructional techniques, we briefly describe the predominant methods to provide a sense of what constitutes remedial education in Texas. One approach, known as "mainstreaming", is to offer remedial services in conventional academic departments. Other schools have "centralized" administrative units that provide remedial education. A recent survey indicates that 74 percent of schools in Texas use mainstreaming as opposed to centralized remedial services (THECB, 2005). Although little evidence exists on the efficacy of either approach (Perin, 2002), mainstreaming is thought to improve quality of instruction by helping align remedial and academic coursework. On the other hand, a centralized approach helps enable colleges to embed ancillary support services such as learning assistance centers into their remedial programs.

Finally, practitioners of college remediation may use a variety of instructional techniques. The most common method is the so-called "skills and drills" approach (Grubb and Kalman, 1994), which emphasizes basic skills (such as grammar and vocabulary) then building up to more complex skills (such as reading for meaning). Alternatively, in "student-centered" approaches, instructors mold the curriculum to match what the students in a particular class hope to get out of attending college (Grubb, 2001). Remedial strategies may also differ in breadth. Some have a narrow focus on academic skills, while others take a more holistic approach and also foster non-cognitive competencies such as social skills (Boylan, 2001).

2.2 How Might Remediation Affect Student Outcomes?

The most obvious effects of remediation result from what actually transpires in remedial education. Ideally, the material covered in remediation helps students develop strong academic skills thereby improving college performance. In addition to these instructional effects, remediation also affects student peer groups since remedial courses are populated by low-ability students. If having low-ability peers has a negative effect on own-outcomes, then peer effects could offset positive instructional effects. A related possibility is that assignment to remediation could worsen academic outcomes by stigmatizing students, possibly by triggering the phenomenon psychologists refer to as "stereotype threat" (Steele and Aaronson, 1995).

Remediation can also have other unintended consequences. Since remediated students spend at least some time taking courses that do not count towards a degree, remediation might increase the time necessary to graduate. The cost of such a delay is foregone labor market earnings. In the analysis below we test for such a delaying effect by examining whether the effect of remediation on graduating within 4 years differs from the effect of graduating within 5 or 6 years. Another possibility is that remediation could crowd out attainment of academic credits. For instance, consider a remediated and non-remediated student who each spends 3 semesters in college and do not complete a degree. The remediated student would have completed fewer academic credits since they spent part of their time in remedial courses that do not earn academic credits. We test for these effects by examining the impact of remediation on total academic credits as well as on credits in the first year, when crowding out is most likely to occur.

Finally, the policy of offering remediation could have system-wide impacts that extend beyond students who actually are in remediation. Remediation could allow instructors of academic classes to maintain the rigor of their courses, improving outcomes for non-remediated and possibly even remediated students. On the other hand, offering remedial courses might lower college quality by diverting resources away from standard academic courses. Despite the potential importance of these effects, we cannot examine them empirically since all colleges offered remediation throughout our study period. Therefore, like virtually all studies on college remediation, we focus on the effects of assignment to remediation.

2.3 What has Prior Research on College Remediation Shown?

Most early studies of the effect of remediation suffer from serious methodological and data limitations. Chief among these is an inability to account for the differences between remediated and non-remediated students (O'Hear and MacDonald, 1995). However, several recent studies use innovative empirical strategies that address selection bias.[4] Bettinger and Long (2009) use an IV strategy that exploits the differing remediation placement standards across colleges in Ohio. Because students are more likely to enroll in a college that is near their home, living nearby a college that has either stringent or lax remedial placement standards generates

[4] A related study is Angrist et al. (2009) who conduct a randomized evaluation of a program that offers monetary rewards for good grades as well as educational services that resemble remediation (peer advising and organized study sessions). They find no effect of offering educational services alone, but for women they do find that incentives combined with services had positive effects. Lavy and Schlosser (2005) find positive effects on high school completion of a remediation program in Israel.

variation in the probability of remediation. They find that remedial education has positive effects on transferring to a more selective college and earning a college degree. One potential limitation of their approach is that colleges may differ along dimensions other than their remediation policies. In particular, the estimates of the impact of remediation might be overstated if schools with stricter remediation placement policies have other (non-remediation) features that improve the outcomes of remediated students. Jepsen (2006) compares California community college students who were in remediation to those who were referred by staff to be in remediation but who chose not to enroll in remediation. Like Bettinger and Long (2009), he finds some evidence that remediation has positive effects on college persistence and degree completion. However, these estimates could be confounded if, for instance, referred students who enter remediation have relatively high levels of unmeasured academic motivation.

Calcagno and Long (2008) use data on Florida community college students and a regression discontinuity design to isolate the causal impact of remediation.[5] In contrast to Bettinger and Long (2009) and Jepsen (2006), this study's results are somewhat inconclusive. It finds that remediation improves fall-to-fall retention. However, it also finds little evidence that remediation increases the likelihood of completing college-level courses, transferring to a four-year college or completing a degree. We expand on this study in two important ways.[6] First, we examine the effect of remediation at four-year schools as well as at community colleges (Texas has remediation in two- and four-year schools, while remediation in Florida is almost exclusively offered at two-year schools). Second, our data allow us to examine the effect of remediation on labor market outcomes.

2.4 College Remediation Placement Policies in Texas

Our study examines the impact of remediation on Texas college students who entered college in the 1990's. Texas is a good state to study the effect of remediation because it is a populous state (only California has more students enrolled in public colleges, NCES, 2007) that

[5] Three earlier studies also used a regression discontinuity design (Aiken et al. 1998; Lesik, 2006; Moss and Yeaton, 2006) but they all had small sample sizes, making it difficult to carry out narrow comparison at the treatment assignment threshold.

[6] One important limitation of the data they use is that it only has information on the student's most recent placement exam score. Because some students who initially fail can retake and pass the test, the assumption that placement exam barely-failers are comparable to barely-passers might be violated. Results based on a limited sample of students attending schools that appear to not allow retesting are generally, but not uniformly, suggestive that remediation improves fall-to-fall retention, but this result is sensitive to the choice of bandwidth around the passing cutoff and the years included in the sample. In contrast, we have data on all of a student's test attempts. By focusing on the initial test score, we avoid the "retesting bias" problem. However, as explained below, this approach weakens the "first-stage" relationship between placement score performance and remediation, as students who initially fail the exam but retest and pass it will not be subject to the remediation requirements

is diverse in terms of economic status, ethnicity and geography. Moreover, Texas has a large number of public two- and four-year institutions that are distributed throughout the entire state. During this period, Texas law required all students pursuing academic degrees to enter remediation if they could not demonstrate college readiness. This policy, known as the Texas Academic Skills Program (TASP), stipulated that college readiness could be shown by passing the statewide TASP test or one of the state-approved alternative tests.[7] Students who have sufficiently high scores on the state's high school exit exam, the SAT or the ACT were exempt from the TASP testing requirement, although such exemptions were relatively rare (about 12 percent in our sample).

The TASP test consists of three sections: math, reading and writing. Raw scores on each section were translated into a scale score that ranges from 100 to 300. The passing standard prior to September 1995 for all sections was a scale score of 220. Afterwards it was increased to 230 for math and reading.[8] This change meant that the "marginal" student changed over our study period. Below, we report results by whether or not students faced the higher or lower passing standard, which provides information about whether the impacts of remediation differed for higher- or lower-ability marginal students. The main portion of the writing exam is an essay which is scored by two graders on a 1-4 scale.[9] This poses two challenges for the regression discontinuity design. First, it introduces "lumpiness" in the score. Second, since graders know the scores needed to pass, students receiving the minimum passing score might be very different from students with the next lowest score. We discuss evidence below consistent with this possibility, and we therefore focus our analysis on the math and reading exams.

Failing the TASP test is fairly common. For instance in our sample, 21 percent of students initially enrolling at a four-year college and 40 percent of two-year college students failed at least one section of the TASP. Students who failed any section of the exam were required to participate in remediation in each semester in which they were enrolled until all sections of the test were passed, although students who failed multiple sections were required to enter remediation in only one of the failed subject areas (THECB, 1995). In practice, however,

[7] We are able to locate TASP test score records for about two-thirds of students entering during the study period. Although some students who did not take the TASP test took an alternative to the TASP, others may have dropped out before taking any test. On the other hand, some students who took the TASP test may have initially taken and failed one of the alternative tests.

[8] Discussions with officials at the THECB indicate that they originally planned to increase the passing standard to 250, but the cut score never increased beyond 230. Note that the contractor responsible for test development recommends a score of 270, more than one standard deviation above the actual cutoff, be used as the standard required to demonstrate readiness for college-level algebra (McDonough, 2006).

[9] If the sum of the two scores equals 5 then they must score sufficiently well on a multiple choice section. If the combines score is less than 5, then they fail automatically.

not all students who failed entered remediation. For example, consider a student who took the TASP test during her first semester in college, but after enrolling in college-level courses. If she failed the exam on the first try but retook it and passed before the start of the following semester, she would not have to enter remediation.[10] Alternatively, she would also appear in our data as not entering remediation if she dropped out of college following her first semester. By the same token, students who pass all sections of the TASP may nevertheless enroll in developmental education. This could occur because their advisors encourage them to do so, because they failed a local placement exam at their school, or if they initially took and failed an alternative to the TASP test.[11] As shown in Section 5, however, failing the TASP strongly affects the likelihood of entering remediation.

3. Data and Descriptive Statistics

This paper uses data from the Texas Schools Microdata Panel (TSMP), which is a collection of administrative records from the state agencies that oversee K-12 public schools, public postsecondary institutions, and the state's Unemployment Insurance system. We examine entering freshman who enrolled in a public two- or four-year college in Texas between the.1991-92 and 1999-00 school years. Throughout this paper we refer to "four-year" and "two-year" students based on the type of school they initially attended even though transferring from one type of institution to the other is fairly common (and in fact is an outcome that we will examine). For these students, longitudinal data files can be made by linking records across files on the basis of an individual's encrypted social security number. Additional details about the data can be found in Appendix A.

Most of the data we use comes from student-level reports colleges submit to the Texas Higher Education Coordinating Board (THECB). These data contain background demographic and other baseline characteristics, and are used to generate academic outcome measures. We supplement data from the THECB files with information provided by the THECB's testing contractor on all TASP test scores taken since the inception of the testing program. It includes the scale score received on each section of the test as well as the administration date, which

[10] The timing of the TASP test changed over our study period. Starting in the 1993 fall semester students had to take the TASP test before completing the 9 academic credit hours while before that they could wait until completing 15 credit hours (approximately one academic term). In the fall of 1998, students were required to take the TASP test before enrolling in college for the first time. Thus, the type of scenario whereby a student could avoid remediation by retesting in the first semester became less common over time.

[11] Conversations with college remediation program officers suggest that retesting is encouraged primarily for students who barely failed the placement exam. Some schools offer short review classes for these students.

enable us to determine the student's performance relative to the passing standard and identify the scores received on the *initial* attempt.

Our primary independent variable of interest is whether a student was in remediation. In Texas, students can be in remediation for different (and possibly multiple) subjects. In this paper we focus on whether a student is in remediation for any subject. We also present results where the "treatment" is defined as remediation for math or remediation for reading (which does not depend on remediation status in other subjects). Because some students take the TASP exam after beginning their first semester, our definition of treatment status includes remediation in the student's first or second semester. Finally, we discuss in Section 6 results where the measure of exposure to remediation is the number of remedial courses a student takes.

The THECB data goes through the 2004-05 school year, allowing us to follow all students in our sample up to six years after first enrolling. These data only cover Texas public colleges, but mobility tends to be rare among lower-ability students (Kain and O'Brien, 2000). We use these data to generate four types of academic outcomes. The first is the number of academic credits a student attempts. We consider total credits attempted in six years as well as the number of academic credits attempted during the first year. Note that we only observe attempted rather than completed credits. Attempted credits include courses the student failed, but excludes courses "dropped" before a grade was assigned. Therefore, we expect the discrepancy between attempted and completed credits to be small. Second, we examine whether a student initially enrolling in a two-year college "transfers up" to a four-year college and conversely, whether a student initially attending a four-year school "transfers down" to a two-year college. Third, we examine the highest grade a student completed. Finally, we examine college graduation.

We also examine the effect of remediation on earnings. These analyses use Unemployment Insurance (UI) earnings records from the Texas Workforce Commission (TWC). The TSP has UI data through the 3rd quarter of 2004. Although not all employment in Texas is covered in the TWC data, estimates suggest the vast majority of workers are covered by the state's Unemployment Insurance system (Stevens, 2002; King and Schexnayder, 1999). The analyses we will conduct examine total earnings received in the 5th, 6th and 7th year after a student first enrolls in college, where annual earnings are converted to year 2000 dollars using the CPI-U.

Our final analysis dataset was generated by making a number of sample restrictions. Students were included in the sample if they: (1) were not exempt from the TASP and who took the placement exam, (2) have non-missing data for date of birth and ethnicity, (3) were pursuing academic degrees when first enrolled, and (4) took the placement exam by the end of their first semester. We exclude students who were not pursuing a degree because the TASP requirements do not apply to them, although we find similar results when they are included. The rationale for the final restriction is that it allows us to focus on remediation taken within the first year, during a student's college career, when it is most likely to have an impact on academic outcomes. For analysis of labor market earnings, we exclude students who entered college after the fall of 1998 to allow a six year follow-up, and students who entered after fall of 1997 to analyze earnings 7 years after entering college.

The research design used here also requires that the sample be limited to students with valid placement exam scores in all three subject areas. In our main analyses that examine the impact of remediation in any subject area, whether the minimum of the math, reading, and writing scores is high enough to pass determines whether a student is required to be in remediation. This minimum score can therefore be used as the assignment variable in a regression discontinuity analysis. However, as noted earlier, scores on the writing section take on only a few values, making it inappropriate for use in a regression discontinuity analysis. Therefore, we further limit the sample to students who passed the writing section (75 percent and 86 percent for two-year and four-year college students). This sample has 255,878 two-year college students and 197,502 four-year college students. A drawback of excluding students who fail the writing section is that these students might have the largest benefits of remediation. However, the analyses which specifically focus on the effect of math or reading remediation include all students regardless of their score on the writing section. Thus, the samples used to analyze the effect of math or reading remediation are 14 to 25 percent larger (for two-year and four-year college students, respectively) than the sample used to analyze the effect of remediation in any subject.

Table 1 reports descriptive statistics for students by whether they were in remediation for at least one subject. There are three important features of the data that bear mention. First, remediation rates in our sample are slightly higher than the national statistics cited in Section 1. For instance, 23 percent of Texas students at public four-year colleges began their academic careers in remediation compared to only 20 percent nationally (NCES, 2003). However, in both Texas and throughout the nation, more students were in remediation for math than for either reading or writing. Second, there are clear differences in baseline characteristics by remediation status. For example, at two-year colleges, students in remediation are 14 percentage points less likely to be non-Hispanic white than are non-remediated students. In addition, compared to non-remediated students, individuals in remediation tend to be older when first enrolling, are more likely to be economically disadvantaged, and have much lower test scores. Finally, across all measures of academic success, remediated students have worse outcomes. For instance, only 23 percent of remedial students starting at two-year colleges earn an Associates or Bachelor's degree within six years, compared to 38 percent of non-remediated students.

4. Econometric Strategy

4.1 Research Design

Isolating the causal impact of remediation is difficult since students participate in developmental education precisely because their academic skills are weak. We therefore use a fuzzy regression discontinuity approach (Hahn, Todd and van der Klaauw, 2001) based on the TASP score passing cutoffs which generates variation in the probability of being in remediation at the passing cutoff. In particular, we argue that, conditional on the TASP score, an indicator for passing the TASP is a valid instrumental variable (IV) for remediation status. The system of simultaneous equations that we will estimate is:

$$Y = \theta R + f(S) + \varepsilon \qquad (1)$$

$$R = \pi F + g(S) + u \qquad (2)$$

where Y is a student outcome, R is an indicator for remediation status, S is the TASP test score, F is an indicator for failing the exam, and ε and u are mean zero random terms. The functions f and g capture the relationship between Y and S and R and S, respectively, away from the passing cutoff. The parameter θ represents the effect of remediation on Y (we discuss the more realistic heterogeneous effects case below). The parameter π is equal to the discontinuity in the remediation rate at the passing cutoff, and represents the magnitude of the shift in the remediation rate induced by failing the TASP.

Using F as an instrumental variable for R, two-stage least squares estimates of θ will be consistent if there is a discontinuity in the remediation rate at the passing threshold (i.e., $\pi \neq 0$) and failing the placement exam is related to Y only through its effect on the likelihood of being in remediation. This "instrument validity" condition can be expressed formally as:

$$E(\varepsilon F \mid S) = 0 \,. \qquad (3)$$

Intuitively, this condition implies that unobserved determinants of Y do not vary discontinuously between students just above and just below the passing threshold. This condition could be violated if students or schools could manipulate the test scores near the passing threshold. Students do not know the exact number of questions they must answer correctly to pass the test, so this scenario is unlikely on a given test attempt.[12] However, students who initially score just below the remediation placement cutoff can retake the test. Consequently retesting could lead to differences between students just above and below the passing cutoff if the most recent or highest score is used as the "running variable" in the regression discontinuity estimation (Calcagno and Long, 2008). Therefore, we use the student's initial test score to avoid bias induced by retesting. Below, we show evidence consistent with our identification assumption when using this approach. However, using the first test score reduces the power of the "first stage" since some students initially assigned to remediation avoid remediation through retesting.

4.2 Estimation

The estimation problem in this context involves estimating the functions f and g on either side of the passing cutoff. Following a number of recent examples in the applied RD literature, we model these functions as low-order polynomials and use data away from the cutoff to estimate the parameters of the polynomial.[13] Specifically, we use a third-order polynomial with interactions between the polynomial terms and an indicator for failing the exam. Effectively, this allows separate cubic functions to be estimated on either side of the passing cutoff. To gauge the quality of the polynomial approximation, we will plot the "local averages" as a function of S and superimpose the regression fit onto this graph; if the parametric fit performs well, it should closely track the local averages.[14] We also assess the robustness of our results by estimating the models on a "narrow band" sample around the passing cutoff (equal to 10 scale-score points above and below the cutoff).[15]

[12] The condition in Equation (3) could also be violated if failing the TASP test affects students in ways other than by affecting the likelihood of remediation. For instance, students who fail might decide to delay college for a semester while they prepare to retake the test, or if it causes them to drop out *before* entering remediation. We found little evidence of either possibility, although the estimated effect on delaying enrollment was statistically significant for four-year colleges in the narrow band sample. However, the estimate was small and statistically insignificant in the global polynomial specification, and the graphical evidence did not reveal a clear discontinuity. Moreover, such an effect would be part of the reduced-form effect of being assigned to remediation, which is clearly relevant for policy.

[13] This approach has been used by DiNardo and Lee (2004), Jacob and Lefgren (2004) and Matsudaira (2009). An alternative approach uses local polynomial regression methods (Porter, 2008). This approach has the advantage of not requiring parametric assumptions about the underlying conditional expectations, but it does require the researcher to choose the bandwidth. McCrary and Royer (2008) discuss the relative advantages of the global versus local polynomial regression in the context of implementing an IV strategy based on a RD design. See also Imbens and Lemieux (2008) for a review of applied regression discontinuity methods.

[14] All standard errors will be adjusted for "clustering" at the test score level, which can be induced by misspecification of the functional form (Lee and Card, 2008).

[15] A more parsimonious specification - linear in the test score with an interaction between the passing indicator variable and the test score – is used when constraining the sample to the narrow band around the passing cutoff. A linear function appears to fit the data reasonably well close to the passing cutoff.

Unlike the specification of Equations (1) and (2), the TASP consists of three subject-specific scores rather than a single score. As explained in Section 3, our primary analyses use the minimum of the math and reading score as the assignment variable. Among students who passed the writing section, the minimum of the math and reading scores determines whether a student will be required to be in remediation in at least one subject area. To analyze the effect of subject-specific remediation, S will simply be the scale score for that subject.

4.3 Interpreting the IV Estimates

Equation (1) implies that the effect of remediation is constant across students, which is unlikely to be true. With heterogeneous returns to remediation, IV will consistently estimate the "Local Average Treatment Effect" (LATE; Imbens and Angrist, 1994).[16] In the current context, the LATE will pertain to "marginal" students who score close to the passing cutoff and whose participation in remediation was manipulated by whether or not they passed the placement exam. If the effect of remediation varies with ability, then our results may not be applicable to students who score well below the passing cutoff. However, results for the marginal group are interesting for at least four reasons. First, many remediated students were fairly close to the passing cutoff – 24 percent of students at two-year colleges and 30 percent of students at four-year colleges who are in remediation score within about one-third of a standard deviation of the passing cutoff. As long as the treatment effects do not vary in a highly non-linear manner away from the passing cutoff, our estimates should closely approximate the effects for a large number of remediated students. Second, the change in the passing standard that took place during the study period allows us to estimate effects for two marginal groups of differing ability. If the results are similar for both cutoffs, then it would suggest that our estimates are applicable for a wider range of students than would be possible in typical RD analyses. Third, policymakers clearly intended for remediation to help students who fall just below the passing level they set. In contrast, students who score well above the cutoff do not need remediation and students far below it arguably have little chance of completing college even with remediation. Finally, the results for this group are informative about whether the passing cutoff is set at an appropriate level.

Our results might also not be informative about the effect of remediation for students who fail the placement exam but avoid remediation by passing on a retest attempt, or who pass the exam but nonetheless enter remediation. Nonetheless, the LATE we estimate will be

[16] With heterogeneous effects, the instrument must also affect the probability of remediation in one direction for the IV estimates to consistently estimate the LATE (Imbens and Angrist, 1994). The "monotonicity" assumption in this context rules out the existence of students who would be in remediation if they passed the placement exam but who would not if they failed it, which we view as reasonable.

informative about the students who are directly affected by the mandatory remediation requirements of the TASP. Our results are therefore clearly relevant to policymakers. To further understand what group the LATE we estimate pertains, we examined the prevalence of retesting during the latter years of our study period (1998 and later) when initial TASP testing was required before students entered college and when retesting to avoid remediation was likely to be most common. Among all students initially failing the TASP prior to fall semester, 29 percent of four-year college students and 16 percent of two-year college students retook the test by September, with slightly higher retesting rates in the "narrow band" subsample. Interestingly, the retesting rates were similar across racial groups and between men and women. This suggests that the LATE is not disproportionately weighted toward the effect for a particular observable student subgroup.

Finally, it is important to recognize that remediation involves several "treatments" occurring simultaneously, and our estimates cannot isolate the impacts of any particular aspects of remediation. For instance, we cannot determine the relative importance of direct instructional effects versus peer or stigma effects. However, since assignment to remediation involves all of these different effects, the total effect which we estimate is arguably most relevant for policy.[17]

5. Results

5.1 Tests of the Validity of the Research Design

The assumptions required for the validity of our research design have two testable implications: (1) the test score distribution and (2) the conditional expectations of pre-determined covariates should be continuous at the passing cutoff (Lee, 2008; McCrary, 2008). To test the first of these, Figure 1 plots the number of observations in a test score cell (re-centered to be zero at the passing cutoff) as a function of the test score, along with the superimposed fitted values obtained by running a cell-level regression of the cell size on a cubic polynomial in the test score (weighted by the cell size).[18] The estimated discontinuities along with standard

[17] Moreover, there are reasons to suspect that peer effects might be relatively unimportant. First, while residential college peers do seem to matter quite a bit (Sacerdote, 2001; Carrell et al., 2007; and Zimmerman, 2003), only weak evidence of peer effects in college classrooms has been found (Hoel et al., 2006) . In addition, we show evidence that the effects of remediation are less positive (or more negative) in colleges where remediation is more common. The peer effects story would suggest the opposite because the "shock" to peer quality of being placed in a remedial course is smaller in high remediation schools.

[18] The analyses examining the smoothness of the test score distribution exclude scale scores more than 80 points below the passing threshold while the rest of the results in this paper only exclude observations more than 100 scale score points below passing. The additional restriction is made here because the parametric fit of the average cell size tracks the data poorly for extremely low scores. The estimated discontinuities in the mean cell size yield qualitatively similar conclusions when the full range of test score values is used.

errors are reported in the upper-right of the figure, and are consistent with a continuous distribution of test scores for both two- and four-year colleges.[19,20]

To test the second implication, we examined whether baseline characteristics exhibited discontinuities at the passing cutoff. A convenient way of summarizing the baseline covariates into an informative "single index" of academic preparation is to compute the conditional expectation of a student outcome as a function of observables. We did this for college graduation, completion of at least one year in college, and Year 7 earnings. Figure 2 shows that the conditional probability of college graduation is strongly related to the TASP score, but that it is smooth through the passing cutoff. Table 2 shows that for this and all other baseline covariates, the estimated discontinuities are all small and statistically insignificant.

5.2 Effect of Performance on the TASP Test on the Probability of Remediation

Figure 3 plots the likelihood of being in remediation in at least one subject as a function of the minimum of the math and reading TASP test score. The open symbols represent the remediation rate for students in a particular test score cell. In both two- and four-year colleges, there is a sharp fall in the remediation rate at the passing cutoff. These results provide strong evidence that failing at least one section of the TASP has a strong causal effect on the likelihood of being in remediation. The point estimate for the four-year colleges is somewhat larger than it is for two-year colleges (0.421 compared to 0.360), but both are precisely estimated. Thus, the IV estimates discussed below are not subject to the statistical problems associated with "weak instruments" (Bound et al., 1995). An important consideration when evaluating the point estimates is the degree to which the polynomial in the test score (the function f in Equation 3) accurately approximates the underlying conditional expectation, $E(R|S)$. Misspecification of f is unlikely to be driving these results since the visual evidence in Figure 3 indicates that the polynomial fit closely "tracks" the cell means and the estimates in a "narrow band" sample (top row of Table 3) are similar to those using the full test score range. In addition, adjusting for baseline covariates has a minimal effect on the point estimates.[21]

---

[19] The way the NES assigned scale scores resulted in very few students receiving a scale score one point below the passing cutoff (S=-1) and a relatively large number of students receiving a scale score equal to 2 points below the cutoff (S=-2). Therefore, in Figure 1 we collapsed the data in the S=-1 and S=-2 cells and replaced it with the average of the two cell sizes. The first row of Table 2 shows the estimated discontinuity when the "unadjusted" cell size is used. The statistically significant negative discontinuity is entirely driven by the large number of observations in the S=-2 cell. When examining the unadjusted cell size and weighting all observations equally, the estimated discontinuity is small and statistically insignificant.

[20] This test is similar in spirit to the test for a discontinuous density proposed by McCrary (2008). Appendix A, discusses why McCrary's specific procedure does not appear to yield valid statistical inference in this context.

[21] Additional covariates include the maximum of the math and reading scores and dummy variables for white, Hispanic, starting in fall semester, being 21 or older when enrolling in college, academic year of enrollment,

Although these results show clear evidence that performance on the TASP affects the probability of entering remediation, it is important to note that there are many students who fail a section of the TASP but do not go into remediation ("never-takers" in the parlance of Imbens and Angrist, 1994) and others who enter remediation even though they passed all sections ("always-takers"). There may be always-takers because some students who pass the TASP might fail a local placement exam (which we do not observe in our data), or be advised by their college counselor to enter remediation. There are also reasons why one would expect sizable numbers of never-takers to exist as well. Recall that Figure 3 uses the student's initial exam score. As noted in Section 2, students who initially fail the test can retake and pass it, thereby avoiding the mandatory remediation requirement. Another explanation lies in the fact that some students who fail the test might drop out of school before entering remediation (this type of student would be treated as non-remediated). In our data, 86 percent of two-year college students and 93 percent of four-year college students enroll in a Texas university in the semester immediately following the semester during which they initially enrolled, although some might drop out before the completion of this semester and thus not enter remediation. The decision to drop out is potentially affected by failing the TASP test. If it is, then failing the TASP might affect student outcomes directly (i.e., not through its effect on the likelihood of being in remediation), thereby threatening the validity of our IV strategy. However, in results not reported, we find no evidence of an effect on the likelihood that a student leaves school at the end of their first semester.

5.3 Estimates of the Impact of Remediation on Academic Outcomes

We now turn to the results for academic outcomes, starting with academic credits reported in the lower panels of Table 3.[22] One of the criticisms leveled against remediation is that it reduces the time spent on earning credits that count towards a degree. To test this claim, we examined whether remediation is associated with fewer academic credits attempted during the student's first year, the time when "crowding out" would most likely occur. Figure 4 shows that attempted credits during the first year are slightly higher for students who passed the

academic year student initially took the TASP, economically disadvantaged, missing data on economically disadvantaged, receiving in-district tuition, missing data for in-district tuition status, distance from HS<25 miles, distance from HS>50 miles, distance from HS missing, and whether the TASP was taken at least 2 semesters before entering college.

[22] An earlier version of this paper reported results on performance in a students' first college-level math course. Among students who took such a course for a grade, our estimates suggested that remediation improved the grade a student received. However, these results were difficult to interpret because this analysis was conditional on a sample of course-takers. In the conditional sample, placement exam barely-failers would not be comparable to barely-passers if remediation affects the likelihood of taking a college-level math course. In contrast to the estimates for the other academic outcomes considered in this study, we found that the estimated effects fell sharply when adjusting for baseline covariates. This pattern is consistent with selection into course-taking which may generate biases in the estimated effects of remediation.

TASP.[23] The IV estimates reported in Table 3 show that being in remediation reduces first-year academic credits by about 2.4 for two-year college students and 1.5 for four-year college students. To see whether this effect persists, we examined the total number of credits attempted over six years. Appendix Figure 1 shows some evidence that at two-year colleges, barely-passers attempt more credits and the IV estimates in Table 3 suggest that remediation leads to a reduction of 3 to 6 academic credits (the smaller estimate in the narrow band sample is not statistically significant). The point estimate for four-year college students is about half as large and only statistically significant when controlling for baseline covariates. Most students in remediation take one or two remedial courses, with each course worth three credit hours. So viewed in relation to the number of academic credits one credit hour of remedial education crowds out, a reduction of 6 academic credit hours is sizable. However, in terms of years of college, this effect is rather small since one full year of college is 30 credits.

The next outcome we consider is whether a two-year college student transfers up to a four-year school. This outcome is important since a central mission of two-year colleges is preparing students for university-level study (Cohen and Brawer, 2003). Figure 5 shows no evidence that transferring up is more common among students who barely fail the TASP test. The IV estimates are small and statistically insignificant, and we can rule out positive effects larger than 1 percentage point (when adjusting for baseline covariates) on a base of 30 percent. Similarly, the IV estimates of the effect of remediation on the likelihood that four-year college students transfer down to a two-year college are small and statistically insignificant. In the narrow band sample, the estimates are larger in magnitude (suggesting that remediation *increases* transferring down by 4 percentage points on a base of 16 percent), but the visual evidence in Figure 5 provides little indication of a discontinuity at the passing cutoff.

The final set of academic outcomes describes a student's college attainment. Figure 6 plots the fraction of two- and four-year college students that complete at least one year in college by the TASP test score and suggests that students who barely fail the exam are slightly less likely to complete at least one year than students who barely pass. The IV estimates in Table 3 indicate that remediation for those starting at two-year colleges lowers the probability of completing at least one year in college by 6 percentage points. However, the estimates based on the narrow band sample are smaller and statistically significant only when controlling for baseline covariates. The estimated effects on the likelihood of completing two or more years are also fairly small and

[23] The sample in these analyses is limited to students who entered college in the fall because some schools only report credits attempted on an annual rather than on a semester basis. This makes it difficult to determine the number of credits attempted during a student's first year for those entering in the middle of an academic year.

statistically insignificant. Turning to the results for four-year colleges, Figure 6 and the IV estimates in Table 3 provide little evidence that passing the TASP or placement in remediation affect the number of years completed (although the estimate for completing at least one year is negative and marginally significant when adjusting for baseline covariates).

Figure 7 shows the fraction earning a college degree within six years of entering college and the TASP score. There is no evidence that the graduation rate changes sharply at the passing cutoff and the estimated discontinuities are small and statistically insignificant. These results imply that remediation has little effect on eventual degree attainment. However, it may be that remediation increases the time needed to complete a degree. We test this possibility by examining whether remediation affects the probability of graduation within 4, 5, or 6 years of entering college. Because we find no effect of remediation on the likelihood of graduating within 6 years, if remediation delayed time to graduation, then there ought to be a negative effect on graduation within 4 (or 5) years. However, we find no evidence of such a pattern. Thus, our results do not support either the hypothesis that remediation increases the time needed to finish a degree nor that it improves the chances of graduation for under-prepared students.

5.4 Estimates of the Impact of Remediation on Labor Market Outcomes

We now discuss the effect of remediation on labor market earnings. The impact on earnings is informative about whether remediation helps students develop economically valuable skills, and should therefore be considered a key measure of the program's effectiveness. Although we found little evidence that students' academic outcomes improve if placed in remediation, it is still interesting to see if there is an effect on labor market success. Remediation may improve a student's ability to learn in college courses even if it does not affect the number of years completed or the likelihood of earning a degree. Furthermore, the actual skills covered in remedial education (such as literacy) could still have value in the labor market (Johnson, 2005).

Although understanding the effect of remediation on earnings is important, this analysis is complicated by two related issues. The first is that earnings received while still attending college are likely to be a poor indication of long-run earnings potential. We address this issue by following students for up to 7 years after they first entered college, by which time most students have left college (in the 7th year, only about one-quarter of students are still enrolled). Because the earnings data end in the third quarter of 2004, to analyze earnings 7 years out, we restrict the sample to students first enrolling in fall 1997 or earlier.[24]

[24] This restriction is not made for the analyses of labor market outcomes in years 5 and 6.

A second issue is that the UI earnings records contain no information about individuals who are not employed in a job covered by the Texas UI system. Thus it is impossible to distinguish between instances where someone has zero earnings, works in an uncovered job, or has moved outside of Texas. One approach to this problem is simply to impute zero earnings when an individual has no earnings in the UI records. The disadvantage of this strategy is that the estimated effects of remediation on earnings might be confounded if remediation affects the likelihood of imputing zeros. An alternative approach is to condition the sample on having positive earnings. The drawback of this method is that it could introduce selection bias. For instance, having positive earnings might be affected by remediation if remediation makes it easier (or harder) to find a job. In that case, conditioning the sample on positive earnings could introduce differences between barely failers and barely passers, thereby violating underlying assumptions of the research design.

Since imputing zeros and conditioning the sample on having positive earnings each have drawbacks, our approach is to compare the estimates from each method. Figure 8 plots average earnings in the 7th year for two- and four-year college students, excluding those who have zero earnings. In all instances, earnings are strongly related to the TASP test score, but there is no sharp change in mean earnings apparent at the passing cutoff. The IV estimates in the first four columns of Table 4 are all statistically insignificant regardless of whether we condition on positive earnings or impute zeros. This pattern suggests that missing earnings data are probably not driving the results. Furthermore, the estimates of the discontinuity in the probability of having positive earnings in the UI data are small and statistically insignificant except for two-year colleges in the "narrow band" sample, where the IV estimates are similar in the conditional and unconditional samples. Also notable is the fact that the estimates are insignificant in the 5th, 6th and 7th years after entering college even though the fraction of students enrolled in college falls steadily over this time period. Thus, distortions arising from observing earnings while individuals still attend school do not appear to have important influence on our results.

Another way to assess whether restricting the sample to positive earners generates bias is to estimate parametric selection-correction models. Under certain conditions controlling for the inverse Mill's ratio can yield consistent estimates of regression coefficients even in the presence of non-random sample selection (Heckman, 1979; Wooldridge, 2001). A key limitation of this approach, however, is that we rely on assumptions about the functional form of the distribution of the residuals for identification since we do not have an instrument for selection into the positive earnings sample. As can be seen from the last set of estimates in Table 4, the results obtained from implementing this procedure do not change the main conclusion that remediation does not improve earnings. Because the inverse Mill's ratio is a "generated regressor," standard errors are calculated via 150 bootstrap replications, where sampling is done at the TASP score cell level and the inverse Mill's ratio is regenerated in each iteration. The resulting standard errors are therefore larger, especially in the "narrow band" sample.

Although we find no evidence that remediation improves labor market outcomes, this may be because remediation has a relatively small effect which we cannot detect. Focusing on the results from specifications that use the full range of test scores and hence have the greatest precision, the upper bound of the 95 percent confidence interval for the impacts on year 7 earnings is $1,100 for two-year colleges in the sample that includes zero earnings, or 6.1 percent of the sample mean. In the sample restricted to positive earnings, the upper limit of the confidence interval is $1,670 and $1,800 in the specifications that do and do not control for the inverse Mill' ratio, respectively, or 7.2 and 7.8 percent of the sample mean. For four-year colleges, the upper bounds are $400 in the unconditional sample and $860 in the conditional sample ($929 when controlling for the inverse Mill's ratio), or increases of 2.1 and 3.4 percent relative to the respective sample means. To place these upper bounds in context, we use the estimates of Kane and Rouse (1995) who find that an additional 30 college credits (or one year) increase earnings by 4 to 6 percent. Their estimates imply our upper-bound effects are comparable to what would be gained by completing 1 to 1.5 additional years of schooling for two year college students, and about half a year for four-year college students. Thus, we cannot statistically rule out economically significant positive effects on year 7 earnings for two-year college students, but we can rule out fairly small positive effects for four-year college students. Furthermore, for years 5 and 6 where the estimates are considerably more precise, we are able to rule out effects larger than 3.5 to 6 percent of the sample mean for two-year college students in the specification where we do not impute zeros.

5.5 Estimates for Subgroups

To examine whether the aggregate estimates mask benefits for certain groups of students, we examined the effect of remediation on four key outcomes – completing at least one year in college, the total number of credits accumulated, receipt of a degree, and labor market outcomes – separately by subgroup. First, we examined the results by whether a student initially took the TASP prior to or after September 1995, when the passing standard was increased from a scale score of 220 to 230 (about one-third of a standard deviation). The marginal student in the earlier period was of lower ability than the marginal students later in the study, so a comparison of these results is informative about whether the effect of remediation differs by ability.[25] Regardless of the passing standard level, the results in Table 5 show no significant positive effects of remediation on academic or labor market outcomes. Estimated effects in the higher-standard sample are below those for the lower-standard sample, although the hypothesis of equal effects can only be rejected for completing 1 year of college in four-year colleges. These results suggest that remediation may have "less negative" effects among lower-ability students, although even under a lower placement cutoff, remediation does not appear to improve the outcomes of marginal students.

Next, we examined results by how common remediation is at a given school. We computed the distribution of the fraction of students in remediation (separately by two- or four-year and by academic year), and split the sample into whether a student was in a school in the top quarter or bottom quarter of this distribution. For the outcomes listed in Table 5, the hypothesis that the effects are equal across the two types of institutions at the 5 percent significance level cannot be rejected. However, the pattern of coefficients for two-year schools suggests that remediation has sizable negative effects on academic outcomes in high-remediation institutions (such as a 12 percentage point reduction in the probability of completing at least one year of college) and smaller effects in low-remediation schools. We cannot definitively determine why sizable negative effects exist in the high-remediation schools, but one possibility is that the quality of remedial programs is diluted by the relatively high fraction of students in remediation. A second possibility has to do with the fact that students in high-remediation schools spend somewhat less time in college (as measured by credits attempted or years completed). For these students, being assigned to remediation could displace a greater share of academic credits since they have less time to make up the classes they did not take while in remediation. The results for year 7 earnings actually show large negative effects in low-remediation four-year schools, but these estimates are smaller in the "narrow band" sample (Appendix Table A1), and the effects on year 6 earnings are also quite small.

While by no means definitive, these results also suggest that negative peer or stigma effects of assignment to remediation may be unimportant. If peer effects had detrimental consequences for remediated students, then remediation should have smaller positive (or larger negative) effects in schools where assignment to remediation represents a bigger "shock" to

[25] About 9 percent of students who entered in the fall of 1995 or later initially took the TASP under the old passing standard. About 1 percent of students in the earlier cohorts took the TASP under the more stringent passing standard, which meant they did not take the TASP during or prior to their first semester in college. Similar estimates to those in Table 5 are obtained when stratifying the sample by whether a student started college before or after September 1995.

one's peer group. By the same reasoning, "stigma" associated with assignment to remediation would likely be stronger in schools where remediation is less common. If anything, we find the opposite pattern.

Across the other dimensions examined in Table 5, we again find no evidence of positive effects of remediation, and the estimates typically do not vary substantially along demographic lines. Notably, the results for men and women do not differ significantly, which is in contrast to what has been found in some studies of college-level interventions (Angrist et al., 2009).[26]

5.6 Estimates of the Effect of Remediation in a Particular Subject

Up to this point, our estimates have all been on the effect of receiving any remediation regardless of subject area compared to not being in remediation at all. To test for any differential effects by remediation subject area, we now analyze the impact of remediation specifically in math or reading. Also note that since the focus now is on remediation in math or reading irrespective of remediation status in any other subject, it is no longer necessary to exclude students who passed the writing portion of the TASP.

Table 6 reports the estimates for selected academic and labor market outcomes that use the math or reading score as the assignment variable in the regression discontinuity analysis. Since the way we define the math (or reading) remediation treatment is not contingent upon remediation status in a different subject, it is possible that the marginal effect of assignment to math remediation on time available to take degree-counting courses could be smaller than the marginal effect of the "any subject remediation" treatment. The results in Table 6 are consistent with this conjecture. Math remediation – the most prevalent form of remediation – has a smaller negative effect on attempted academic credit hours and on the likelihood of completing at least one year in college than what was seen in Table 3. Moreover, the IV point estimates are smaller still and not statistically significant in the narrow-band sample. Similarly, the estimated effects of reading remediation on academic outcomes are small and statistically insignificant. There are several instances where the estimated effect on earnings is negative and statistically significant in either the narrow-band or global-polynomial samples, but never in both specifications. In Appendix Table A2, we report estimates from models that include math and reading remediation in the same equation. This type of model is identified because the discontinuity in the probability of math remediation and the discontinuity in the probability of reading remediation offer two

[26] The narrow-band estimates show significant positive effects on earnings for blacks and Hispanics for two-year college students, but we discount this finding for two reasons. First, in the full sample the point estimates are negative (although not statistically significant). Second, we find negative effects on all academic outcomes for both the full and narrow-band samples, making it highly unlikely that remediation could improve earnings.

instruments. The estimates from this type of model yield similar qualitative and quantitative conclusions to those in the models reported in Table 6.

Finally, we also examined the effect of remediation for writing. The results (not reported) suggest that remediation has large negative effects on most outcomes. However, when the writing score is used as the "running variable", we find evidence inconsistent with a valid regression discontinuity design such as discontinuities in baseline observables at the passing cutoff. As noted in Section 2, we conjecture this is because the writing section is largely a subjective assessment, and the graders know the passing standard. In contrast, the math and reading sections only have multiple choice questions.

6. Discussion

6.1 Why Do We Find No Evidence of Positive Effects of Remediation?

One potential reason we do not find evidence that assignment to remediation benefits students is that the remedial offerings in Texas are ineffective. For instance, because remediation is mandatory for low-scoring students, the incentives to monitor remedial programs and ensure that remedial offerings are of high quality may be low due to the inelastic demand for remedial courses. Alternatively, it may be due to the decisions students make after being assigned a remediation status after receiving the TASP scores. One possibility is that students who barely place out of remediation take easier courses, thereby offsetting any learning advantages imparted by remediation. It is difficult to directly test this claim since our data do not include information on the courses students take. However, we are able to determine college major for students who graduate. With this information we constructed a "difficulty of major" index equal to the average standardized math and reading TASP score in a given major. Although not literally a measure of how challenging a major is, it does reflect whether a student pursues a degree that higher-ability students pursue. Another limitation of this analysis is that it must be done conditional on graduation, which introduces the possibility of selection bias if students just above and below the placement score cutoffs are not comparable on unobservable dimensions.[27] Consequently the estimated effects of remediation on college major should be thought of as suggestive.

With these important caveats in mind, the IV estimates are reported in Appendix Table A3. For the most part the estimated effects are small in magnitude and statistically insignificant.

[27] We find little evidence that remediation affects the likelihood of graduation on average or for most selected subgroups of students, which offers some reassurance that selection bias may not be a serious concern. However, remediation may have a "non-monotonic" effect. That is, it may increase the likelihood for some and decrease it for others, with these effects canceling out in the aggregate. If this is the case, then students just above and below the placement score cutoffs may not be comparable on unobservable dimensions,

If anything, the results suggest that remediation leads to students taking *easier* majors, although we only found statistically significant effects when using average math scores as the outcome, and statistical significance was not robust to changing the sample from the "global polynomial" to the "narrow band" sample, and vice versa. Overall, we find no evidence that remediation leads students to pursue more difficult majors which makes their outcomes appear worse than those of non-remediated students.

A second possibility is that students in remediation simply have to take more total courses to make up for the time not spent in academic courses. This may lead to less time studying, or less time developing a social network, which may be related to success in college (Ellison et al., 2007; Davis, 1991). To examine this possibility, we re-estimated the IV models in Table 3 controlling for total credits enrolled in during the first year (including non-academic and remedial credits). This approach is limited in that total credits are an intermediate outcome and students at the TASP placement cutoff who have the same number of total credits may not be comparable. Nonetheless, comparing the results controlling for total first-year credits to those in Table 3 will offer some suggestion as to whether the additional coursework burden "undoes" beneficial effects of remediation. As seen in Appendix Table A4, the evidence does not support this hypothesis, as the point estimates are close to those from the conventional IV models that do not control for the intermediate credit outcome.

6.2 Comparison to Earlier Research

In contrast to our findings, studies such as Bettinger and Long (2009; henceforth BL) and Jepsen (2006) find positive effects of remediation on some outcomes for certain student groups. There are at least three reasons why our results differ from those of other studies. Most obviously, the effect of remediation could vary across states. It may be that the remediation "treatment" in Texas is less effective than it is in California and Ohio, perhaps because the intensity of remediation varies across states. In results available upon request, we find that for a subsample of students for whom we have information on remedial course-taking, marginal students assigned to remediation take 1.5 and 1.4 remedial courses at two- and four-year colleges, respectively.[28] We see it as plausible that similar patterns exist in other states, although we cannot make firm conclusions since the studies which find positive effects do not report results on this question. However, if remediation is pursued more intensively elsewhere, then

[28] These analyses exclude 35 percent of two-year college students and 5 percent of four-year college students who attend schools that only report information on remediation on an annual basis. When this occurs, the data only identify whether a student was in remediation at any point in the year and not the total number of semesters a student spent in remediation.

that could explain why our results differ. Moreover, there are institutional differences that might matter. States such as Ohio grant more latitude to institutions to set remediation policies, while during our study period Texas had a number of rules that applied statewide. That said, BL note the presence of program features common in Ohio (such as limits on academic course-taking until remediation is complete) that are also present in Texas. At a minimum, our findings suggest there may be important heterogeneity in the effectiveness of remediation.

Second, we look at different types of students. BL's empirical strategy requires limiting the sample to students who took the ACT or SAT. Since non-selective schools often do not use ACT and SAT scores in admissions decisions, we probably look at students of lower average ability. Jepsen's sample is composed of students who were referred to remediation. In contrast, our sample consists of students who took the statewide placement exam and who were seeking an academic degree when starting college.

Third, since our empirical strategies differ, we may not be estimating the same average effect.[29] BL use an IV strategy that captures the effect for students whose remediation status is affected by attending a school with a more (or less) lenient remediation placement policy. Jepsen attempts to estimate the effect of "treatment on the treated," by comparing remediated students to students referred to remediation but who did not participate in remediation. In contrast, our approach identifies the effect of remediation for students whose remediation status is manipulated by passing or failing the placement exam and who score close to the passing cutoff. Although the effect for this group clearly could differ from the average effect other studies attempt to estimate, it should be noted that BL's research design also focuses on "marginal" students who go into remediation based on whether they attend a school with more or less stringent placement policies.

One way to determine whether remediation has different effects in different states is to replicate the approach used in BL, thereby "holding constant" research design and sample restrictions. The details of this analysis can be found in Appendix C. The estimates in Appendix Table 5 are mixed. The estimates produced by the specification closest to BL (that bases the instrument on the closest four-year college) are generally positive and consistent with BL, but the estimates are mainly negative when basing the instrument on the closest college. Further complicating the interpretation is that the first-stage is only a third as large for math and a quarter as large for reading as in Ohio (in the specification closest to BL). This is likely due to

[29] Calcagno and Long's (2008) study, which uses a similar research design to the one we use, also finds little evidence of large long-run benefits of remediation

the statewide TASP policy which limited the relationship between school attended and the likelihood of remediation. Finally, the estimates are almost all statistically insignificant when adjusting for "clustering" on the closest college to the student's home. This adjustment is important since the IV strategy exploits variation across schools in the likelihood of remediation. As a consequence, this replication does not provide clear enough results to shed much light on why the estimates in BL differ from those we report here.

7. Conclusion

This study uses a regression discontinuity design to assess the impact college remediation has on academic and labor market outcomes for a large sample of Texas students. We find little evidence that remediation improves student outcomes. In fact, some of our results suggest a small negative effect on the number of academic credits attempted and the likelihood of completing at least one year of college. The estimated effects on degree completion, labor market earnings and the likelihood of transferring up to a four-year college are generally small in magnitude and statistically insignificant.

These results have several implications. First, they suggest marginal students in Texas receive little benefit from remediation, despite the large financial cost of the program. Second, our findings imply that the passing cutoff is not set at the appropriate level, although we cannot conclude whether it is set too high or too low. If marginal students are actually not in need of remediation, then the passing cutoff should be lowered. On the other hand, the current passing standard might be too low if marginal students are so far behind their peers that remediation cannot make much difference in their outcomes. Finally, they suggest that remediation is an important example of a "second-chance" later-life intervention that does not improve marketable human capital, although some of these estimates are somewhat imprecise.

However, there are some important considerations to bear in mind when evaluating these findings. Chief among these is external validity, as our research design focuses on students at the margin for placement into remediation. That said, a sizable share of remediated students in our sample do score close to the passing cutoff, and the results are similar in two time periods with differing remedial placement cutoffs. Our findings also only pertain to students whose participation in remediation is affected by passing or failing the placement exam. This group is highly relevant for policy but the effects of remediation could be different for students who, for example, seek out remediation regardless of their placement exam scores. Finally, we also only examine a single state (although one that is large and diverse), and remediation might have beneficial effects in states with different student characteristics or remedial policies.

Our results also point to several directions for future research. First, future work should seek to better understand whether and why the effects of remediation differ across regions and student groups (such as those in high-remediation schools). Second, our results do not speak to system-wide benefits of offering remedial education such as preventing college-level courses from being "watered down". Conversely, it might simply serve to draw resources away from other, possibly more valuable, purposes. Future research should evaluate the institutional costs and benefits of offering remediation at the college level.

References

Abraham, Ansley. (1998). Discussant for "Remediation in higher education: Its extent and cost" by David Breneman. In Ravitch, D. (Ed.), *Brookings papers on education policy* (pp. 1-10). Washington, DC: The Brookings Institution.

Abraham, Ansley. (1986). "College-Level Study: What Is It? Variations in College Placement Tests and Standards in the SREB States," *Issues in Higher Education*, Number 22.

Aiken, Leona S., Stephen G. West, David E. Schwalm, James L. Carroll, and Shenghwa Hsiung. 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, Joshua, Daniel Lang and Philip Oreopoulos. 2009. "Incentives and Services for College Achievement: Evidence from a Randomized Trial" *American Economic Journal: Applied Economics*. 1(1): 136-63.

Angrist, Joshua, Guido Imbens and Donald Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association.* 91(434): 444-455.

Bettinger, Eric and Bridget Terry Long. 2009. "Addressing the Needs of Under-Prepared Students in Higher Education: Does College Remediation Work?" forthcoming *Journal of Human Resources.*

Boozer, Michael and Stephen Cacciola. 2001. "Inside the 'Black Box' of Project STAR: Estimation of Peer Effects Using Experimental Data" Yale Economic Growth Center Discussion Paper No. 827.

Bound, John, David Jaeger and Regina Baker. 1995. "Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak." *Journal of the American Statistical Association*, 90: 443-450.

Boylan, Hunter. 2001. "Making the Case for Developmental Education," *Research in Developmental Education*. 12(2): 1-4.

Breneman, David and William Haarlow. 1997. "Remedial Education: Costs and Consequences." *Remediation in Higher Education*. Washington, D.C.: Thomas B. Fordham Foundation.

Calcagno, Juan Carlos and Bridget Long. 2008. "The Impact of Postsecondary Remediation Using a Regression Discontinuity Approach: Addressing Endogenous Sorting and Noncompliance." Mimeo.

Card, David. 1999. "The Causal Effect of Education on Earnings." in *Handbook of Labor Economics*. Orley Ashenfelter and David Card (eds.). Amsterdam: Elsevier.

Carneiro, Pedro and James Heckman. 2003. "Human Capital Policy." in *Inequality in America.* Benjamin Friedman (Ed.). MIT Press.

Cohen, Arthur and Florence Brawer. 2003. "The American Community College." 4th Edition. Jossey-Bass.

Carrell, Scott, Richard Fullerton, Robert Gilchrist and James West. 2007. "Peer and Leadership Effects in Academic and Athletic Performance." Mimeo. Dartmouth College.

Davis, Robert B. 1991. "Social Support Networks and Undergraduate Student Academic Success Related Outcomes: A Comparison of Black Students on Black and White Campuses." In *College in Black and White*. Walter Recharde Allen, Edgar G. Epps, Nesha Hannif, eds. SUNY Press.

Dominitz, Jeff and Charles Manski. 2000. "Using Expectations Data to Study Subjective Income Expectations." *Journal of the American Statistical Association*. 92(439): 855-867

DiNardo, John and David Lee. 2004. "Economic Impacts of New Unionization on Private Sector Employers: 1984-2001." *Quarterly Journal of Economics.* 119(4): 1383-1442.

Dynarski, Susan. 2008. "Building the Stock of College-Educated Labor." *Journal of Human Resources*. 43(3): 576-610.

Ellison, Nicole, Charles Steinfield and Cliff Lampe. 2007. "The Benefits of Facebook `Friends:' Social Capital and College Students' Use of Online Social Network Sites." *Journal of Computer-Mediated Communication*. 12(4).

Graham, Bryan. 2008. "Identifying Social Interactions Through Conditional Variance Restrictions" *Econometrica*. 76(3): 643-660.

Greene, Jay and Greg Foster. 2003. "Public High School Graduation and College Readiness Rates in the United States." Manhattan Institute, Center for Civic Information, Education Working Paper #3, September.

Grubb, Norton. 2001. "From Black Box to Pandora's Box: Evaluating Remedial/Developmental Education." Community College Research Center Brief, No. 11.

Grubb, Norton and Judy Kalman. 1994. "Relearning to Earn: The Role of Remediation in Vocational Education and Job Training." *American Journal of Education*. 103(1): 54-93.

Hahn, J., P. Todd and W. van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design.", *Econometrica*, Vol 69(1): 201-209.

Heckman, James. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1):153-161.

Heckman, James. 2000. "Policies to Foster Human Capital." *Research in Economics*. 54(1): 3-56.

Hoel, Jessica, J. Parker and J. Rivenburg. 2006. "A Test for Classmate Peer Effects in Higher Education." Reed College working paper.

Imbens, Guido and Joshua Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62: 467-475.

Imbens, Guido and Thomas Lemieux. 2007. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142(2): 615-35.

Jacob, Brian and Lars Lefgren. 2004. "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis." *Review of Economics and Statistics* 86(1): 226-244.

Jepsen, Christopher. 2006. "Remedial Education in California's Community Colleges." Mimeo. American Education Research Association, April 2006 conference paper.

Johnson, Rucker. 2007. "Wage and Job Dynamics After Welfare Reform: The Importance of Job Skills." In *Research in Labor Economics*. S.W. Polachek and O. Bargain (Ed.). Elsevier.

Kain, John and Daniel O'Brien. 2000. "High School Outcomes and College Decisions of Texas Public School Students," TSP Working Paper.

Kane, Thomas and Cecilia Rouse. 1995. "Labor Market Returns to Two and Four-Year Colleges," *American Economic Review* 85(June): 600-614.

Kane, Thomas and Cecilia Rouse. 1999. "The Community College: Educating Students at the Margin between College and Work." *Journal of Economic Perspectives* 13(1).

King, Christopher and Deanna Schexnayder. 1999. "The Use of Linked Employer-Employee UI Wage Data: Illustrative Uses in Texas Policy Analysis." Unpublished paper: LBJ School of Public Affairs.

Krueger, Alan. 2003. "Inequaltity, Too Much of a Good Thing" in *Inequality in America*. Benjamin Friedman (Ed.). MIT Press.

Lavy, Victor and Annalia Schlosser. 2005. "Targeted Remedial Education for Underperforming Teenagers: Costs and Benefits." *Journal of Labor Economics* 23(4): 839-874.

Lazarick, L. 1997. "Back to the Basics: Remedial Education" *Community College Journal* 11-15.

Lee, David. 2008. "Randomized experiments from non-random selection in U.S. House elections." *Journal of Econometrics* 142(2): 675-97.

Lee, David. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects", *Review of Economic Studies* 76(3):1071-1102.

Lee, David and David Card. 2008. "Regression Discontinuity Inference with Specification Error." *Journal of Econometrics* 142(2): 655-74..

Lesik, Sally Andrea. 2006. "Applying the Regression Discontinuity Design to Infer Causality with Non-Random Assignment." *Review of Higher Education* 30(1): 1-19.

Manski, Charles. 1989. "Schooling as Experimentation: A Reappraisal of the Postsecondary Dropout Behavior." *Economics of Education Review* 8(4): 343-353.

Matsudaira, Jordan, 2009. "Evaluating the Impact of English Immersion Versus Bilingual Education on Student Achievement", Mimeo. Cornell University.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142(2): 698-714.

McDonough, Linda. 2006. Personal communication. October 24, 2006.

McCrary, Justin and Heather Royer. 2008. "The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Laws Using Exact Date of Birth." Mimeo. University of California - Berkeley.

Moss, Brian G. and William 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). 2007. "Full Time-Equivalent Fall Enrollment in Degree-Granting Institutions, by Control and State or Jurisdiction" http://nces.ed..gov/progams/digest/d07/tables/dt07_212.asp

National Center for Education Statistics (NCES). 2003. Remedial Education at Degree-Granting Postsecondary Institutions in Fall 2000, NCES 2004-010, by Basmat Parsad and Laurie Lewis. Project Officer: Bernard Greene. Washington, DC. http://nces.ed.gov/pubs2004/2004010.pdf

O'Hear, Michael and Ross MacDonald. 1995. "A Critical Review of Research in Developmental Education, Part I." *Journal of Developmental Education* 19(2): 2-6.

Oreopoulos, Philip. 2007. "Do Dropouts Dropout Too Soon? Wealth, Health, and Happiness from Compulsory Schooling." Journal of Public Economics 91:2213-29.

Perin, Dolores. 2002. "The Location of Developmental Education in Community Colleges: A Discussion of the Merits of Mainstreaming vs. Centralization," *Community College Review*. 30(1): 27-44.

Porter, Jack. 2008. "Estimation in the Regression Discontinuity Model." Mimeo. University of Wisconsin – Madison.

Sacerdote, Bruce. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." *Quarterly Journal of Economics* 116(2): 681-704.

Steele, Claude and Joshua Aaronson. 1995. "Stereotype Threat and the Intellectual Test Performance of African Americans." *Journal of Personality and Social Psychology* 69(5): 797-811.

Steinberg, Laurence. 1998. "Commentaries" in *Remediation in Higher Education: A Symposium*. Washington, D.C.: Thomas B. Fordham Foundation.

Stevens, David. 2002. "Employment that is not Covered by State Unemployment." LEHD Technical Working Paper No. TP 2002-16.

Texas Higher Education Coordinating Board (THECB). 1995. *Annual Report on the TASP and the Effectiveness of Remediation.*

Texas Higher Education Coordinating Board (THECB). 1999. *Annual Report on the TASP and the Effectiveness of Remediation.*

Texas Higher Education Coordinating Board (THECB). 2005. *Developmental Education in Texas Higher Education: A Comparison of Policies and Practices.*

Venezia, Andrea, Michael Kirst and Anthony Antonio. *Betraying the College Dream: How Disconnected K-12 and Postsecondary Education Systems Undermine Student Aspirations.* Stanford University Bridge Project Final Report.

Wooldridge, Jeffrey. 2001. *Econometric Analysis of Cross Section and Panel Data.* MIT Press. Cambridge, MA.

Appendix A: Data

Definition of Remediation

We define a student to be in remediation if they were in remediation in either their first or second semester after enrolling. We include remediation in two semesters because some students take the TASP exam after beginning their first semester. The main "treatment" variable is remediation in any subject. We also present results where the "treatment" is defined as remediation for math or remediation for reading. When examining subject-specific remediation, the "treatment" variable does not depend on a student's remediation status in other subject areas. We do not specifically examine remediation for writing because the scores do not lend themselves to a regression discontinuity analysis. An alternative measure of exposure to remediation is the number of remedial courses. Due to data limitations we did not use this as our primary measure of remediation. Section 6.2 discusses results where the "treatment" measure is the number of remedial courses for a subsample of students for whom this information is available).

Definition of Outcomes

Total attempted academic credits: the number of academic credits attempted by a student in the six years after they initially enroll. The THECB data do not include completed credits. The difference between attempted and completed credits is that attempted credits include credits in courses the student receives a failing grade. Note that credits in courses "dropped" before the end of the term are *not* included.

Attempted credits during the first year: the number of academic credits attempted by a student in their first year. This outcome is only defined for students who entered in the fall because some institutions only report annual rather than semester totals.

Transfer Up to a Four-Year School: defined for students initially enrolling in a two-year college as being enrolled in a four-year school in the last semester they are observed enrolled. A student is considered to be enrolled in a four-year school if they have at least 3 credits at a four-year college in that semester.

Transfer Down to a Two-Year School: defined for students initially enrolling in a four-year college as being enrolled in a two-year school in the last semester they are observed enrolled. A student is considered to be enrolled in a two-year school if they have at least 3 credits at a two-year college in that semester.

Years of College Completed: this is based on the highest grade a student completed. For students who completed at least 30 academic credit hours in their final year, the highest grade is defined as the student's highest observed grade, which is included on the THECB student records. Otherwise, it is the highest observed grade minus one.

College Graduation: For two-year college students this is defined as earning a Bachelor's or an Associate's degree within six years. For four-year college students, this is defined as earning a Bachelor's degree.

Earnings: Annual earnings are computed by adding up earnings in each quarter. If an individual does not have earnings in a given quarter, a value of zero is imputed. Earnings in Years 5-7 after entering college are examined in the text, where "Year 1" begins the quarter a student enters college.

Covariates

Our dataset includes basic demographic characteristics such as gender, race/ethnicity and date of birth. Finally, for about two-thirds of our sample, we also have information available from high school records on economically disadvantaged status (generally defined as receipt of free or reduced lunch) as well as the distance of their high school from the college they attend.

Sample

The universe of students used in this study consists of those who enrolled in a Texas public college or university between 1991-92 and 1999-2000 school years, and did so as first-year students. We exclude earlier years because the data from this period do not appear to be complete. Later years are excluded to allow a sufficiently long follow-up period; for each student in our data, we are able to track their academic progress for 6 academic years (strictly speaking, we observe students entering in spring 1999-00 for less than 6 complete academic years). Students were included in the sample if they: (1) were not exempt from the TASP and who took the placement exam, (2) have non-missing data for date of birth and ethnicity, (3) were pursuing academic degrees when first enrolled, and (4) took the placement exam by the end of their first semester.[30] We exclude students who were not pursuing a degree because the TASP requirements do not apply to them, although we find similar results when these students are included.[31] The rationale for the final restriction is that it allows us to focus on remediation taken early on (within the first year) during a student's college career, when it is most likely to have an impact on academic outcomes.[32]

As described in the text, our main analyses also exclude students who passed the writing test. Seventy-five percent and 86 percent of two- and four-year college students, respectively, passed the writing section on the first try. We also exclude students with very low TASP scores (more than 100 scale score points away from passing – about .2% to .3% of the sample) so that values far away from the passing cutoff do not have undue influence on the estimates. This sample has 255,878 two-year college students and 197,502 four-year college students. A drawback of excluding students who fail the writing section is that these students will generally have weaker academic skills and might have the largest benefits of remediation. However, it is important to keep in mind that the analyses which specifically focus on the effect of math or reading remediation include all students regardless of their score on the writing section. This is because assignment to remediation for either math or reading does not depend on the writing score. Thus, the samples used to analyze the effect of math or reading remediation are 14 to 25 percent larger (for two-year and four-year college students, respectively) than the sample used to analyze the effect of remediation in any subject. In the analyses of labor market earnings, we also exclude students who entered college after the fall of 1998. This was done because the TWC data end in the third quarter of 2004, and a six-year follow-up was only possible for students starting in the fall of 1998 (or earlier).

[30] In our study period, there are about 1.18 million entering freshman who were registered as seeking an academic degree (possibly undecided in the case of two-year college students) and who were not listed as exempt from the TASP requirements. We were able to locate TASP test score records for about two-thirds of these students (796,509). Students might not have TASP records because they wound up not seeking an academic degree and therefore were not subject to TASP requirements, dropped out of college before taking the test, were actually exempt from the testing requirement (despite not being identified as such on the first enrollment record), or took an alternative placement exam to the TASP test. About three-quarters of the students with TASP test records took the TASP within their first semester in college.

[31] The THECB records indicate a student's degree-seeking status. For students initially entering a two-year college, we also include "undecided" students since this group includes students who were undecided about pursuing an A.A. or B.A. degree.

[32] Another motivation for the restriction is that performance on the TASP is most strongly related to remedial education for these students; some students who initially took the TASP test in the second semester appear to have been assigned to remediation on the basis of local placement test results (or an alternative to the TASP test), although we cannot confirm that conjecture empirically.

Appendix B: Testing the Continuity of the Distribution of TASP Scores Using McCrary (2008) Test

As an alternative to the results reported in Table 2, we also implemented McCrary's (2008) test of a discontinuous density to examine whether there is evidence of manipulation of TASP scores near the passing cutoff. However, the standard errors McCrary proposes are valid under assumptions which do not appear to hold in our application. The inference framework McCrary proposes assumes a well-behaved distribution that is continuous throughout the support of the running variable (with the possible exception of a discontinuity at the selection cutoff); any deviations from this smooth distribution in a particular sample are due only to sampling error. In contrast, the TASP scale scores are assigned in a way that there are very large differences in the number of observations in adjacent scale score cells, as can be seen in Appendix Figure 5 for the density of scores at 2-year colleges. Thus, the McCrary standard errors will lead to over-rejection of the "no discontinuity" null hypothesis. This can also be seen in Appendix Figure 5. The McCrary 2*SE bands are much tighter than what the actual variability in the distribution of the running variable would seem to imply. In fact, when we implemented McCrary's test using a series of "pseudo" cutoffs, ranging from -5 to 5, we found that the test easily rejected the null hypothesis of no discontinuity in all cases.

At the same time, it is important to note that the point estimates from McCrary's method are close to what we find and report in the paper. The estimated discontinuity he proposes is $\theta = \ln(\lim_{r \to c^+}(f(r)) - \ln(\lim_{r \to c^-}(f(r))$, where *f(r)* is the distribution of the "running variable", *r*, and *c* is the cutoff. Using modified STATA code provided on his website, we obtained point estimates for θ equal to -0.019 (0.009) and -0.048 (0.014) for two- and four-year colleges, respectively. These estimates imply that there are more observations to the *left* of the cutoff (which is the opposite of what would occur if there were manipulation of the test scores due to something like retesting to avoid remediation). The magnitudes of the implied discontinuities are equal to 56 and 76 observations, respectively. These have the same sign and are quite close in magnitude to the estimated discontinuities in the number of observations we report in the paper (306 and 53).

Appendix C: Applying the Research Design of Bettinger and Long (2009) to Texas Data

This appendix describes our effort to use the methodology outlined in Bettinger and Long (2009) to estimate the effect of remediation in Texas. We began by making sample restrictions similar to those used in BL, starting with the complete dataset comprised of students who first enrolled in Texas colleges between the 1991-92 and 1999-00 school years. The sample was limited to students who: (1) were 18-20 years old at the time of first enrollment, (2) were freshman at the time of first enrollment, (3) were full-time students (defined as taking 12 or more credits in the first semester enrolled), (4) took the SAT, (5) took the SAT prior to entering college, (6) had valid zip code for home residence on the SAT record, (7) were enrolled in a four-year college or were planning on transferring up to a four-year college. After making these restrictions, the sample had 215,147 observations with valid math SAT scores and 223,754 observations with valid reading scores and with no missing data on any covariates used in the analysis. Note that the SAT records include questions on family income, a control variable used in BL, only in more recent years. In the analysis, we use dummy variables to control for missing income data. We also produce estimates using data from 1996 (not reported) which are similar to those discussed below.

Note that this sample includes observations that were excluded from our "main analysis" (for instance, no restriction is made for having valid TASP placement scores) and excludes observations that we included (for instance, we made no restrictions based on having valid SAT scores). Due to differences between the data and institutional setting in Ohio and Texas, there are other minor differences between what we were able to do in Texas and what BL did. One is that Texas is an SAT state while Ohio is an ACT state, so we base the sample restrictions and calculation of the instrument using SAT scores. We also do not include some control variables that are on the ACT data but not on the SAT data such as the grade point average in only math or only English courses and type of high school since almost everyone in the sample either attended a public high school or has missing data on this variable.

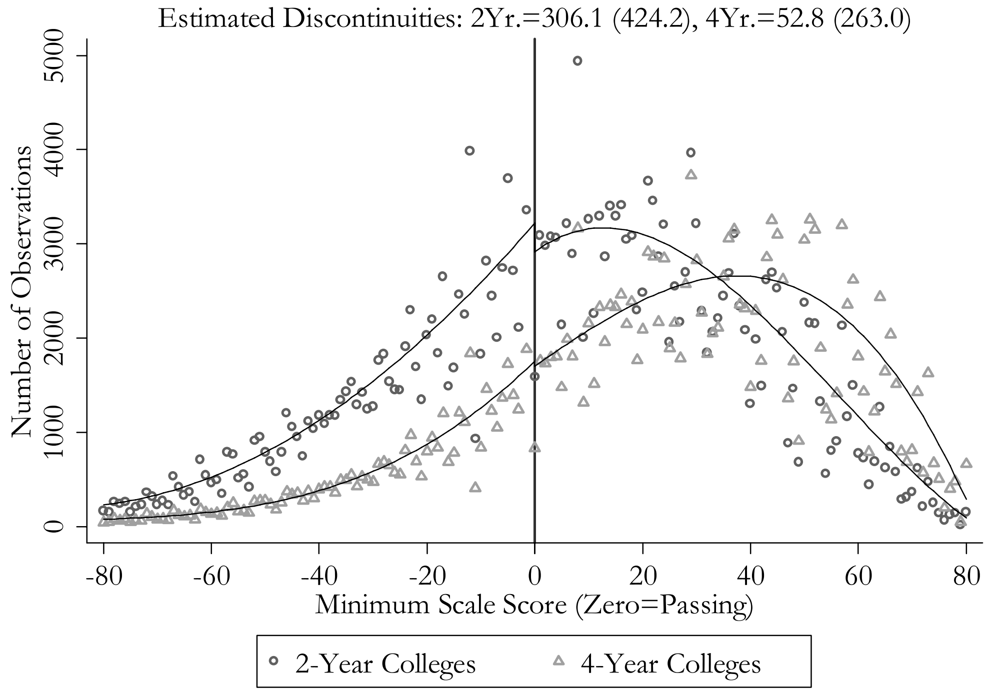
Next, we created the instrumental variable using the approach described in BL. The first step was to estimate a probit model for being in remedial education in math (defined in the same manner as in the main analysis) as a function of a set of school fixed-effects and a set of control variables that closely mirrors those used in BL. Specifically, the additional covariates include a quadratic in the SAT math score, gender, and a complete set of dummy variables for the categorical variables high school GPA, high school class rank, gender, the number of years of math in high school, age at first enrollment, family income, college degree plans, and interactions between high school GPA, SAT math, number of years in math, and college degree plans and the school fixed-effects. A similar model was then estimated for reading remediation, where SAT reading scores replaced SAT math scores and the number of years of high school reading replaced the number of years of high school math.

The instrumental variable was then generated by calculating the probability of being in remediation predicted by the estimated coefficient from the probit model, where the predictions are based on the Euclidean distance between the college and the zip code in which the student lived when they took the SAT. In one specification, we follow BL and base these calculations on the closest four-year college. We also use obtained estimates where the predictions were based on the closest college, irrespective of whether it is a four- or two-year school.

Once the instrument was generated, we estimated the first-stage relationship between remediation and the instrument controlling for SAT verbal and math scores, high school grade point average, gender, number of years enrolled in math in high school, high school class rank, age at first enrollment, family income, college plans in high school, race, academic year indicators and distance from the nearest college, the nearest two-year college and nearest four-year college. As noted in BL, since we are controlling for the main effects of covariates (such as SAT scores) used to predict remediation, the identification of the effect of remediation comes from variation in the conditional probability of being in remediation at the closest across institutions rather than some other student-level factor. The results in Table A5, indicate that the instrument has a much weaker association with remediation than that reported by BL for Ohio. For instance, in the specification where the IV is based on the nearest four-year college, the effect of 10 percentage point increase in the predicted probability of remediation is associated with an increase in the actual probability of remediation of only 0.014 percentage points for math and 0.009 for reading. In contrast for Ohio, the first stage is 0.051 for math and 0.039 for reading. Moreover, the first stage relationship is statistically significant when the standard errors are adjusted for clustering on the nearest college (closest four-year college in columns 1 and 2, closest college in columns 3 and 4). This adjustment is important because the variation in remediation BL exploit is due to differences in the remedial placement policies of the nearest college. The weaker first-stage in Texas is not surprising since the TASP policy established statewide policies governing remedial placements, thereby diminishing the variation in the likelihood of remediation across colleges conditional on test scores.

The estimates in Appendix Table 5 are mixed. The estimates in columns 1 and 2 (where the instrument is based on the closest four-year college) are generally positive and consistent with BL, but the estimates in columns 3 and 4 (when basing the instrument on the closest college) are mainly negative. Additionally, the first-stage is only a third as large for math and a quarter as large for reading as in Ohio in columns 1 and 2, and even smaller in columns 3 and 4. This is likely due to the statewide TASP policy which limited the relationship between school attended and the likelihood of remediation. Finally, the estimates are almost all statistically insignificant when adjusting for "clustering" on the closest college to the student's home. This adjustment is important since the IV strategy exploits variation across schools in the likelihood of remediation. As a consequence, this replication does not provide clear enough results to shed much light on why the estimates in BL differ from those we report here.

Figure 1: Number of Observations by TASP Scale Score



Note: Cell size of cells S=-2 and S=-1 replaced with the average number of observations in these two cells, and is assigned a value of S=-1.5

Figure 2: Estimated Probability of Graduation Conditional on Baseline Covariates by TASP Scale Score

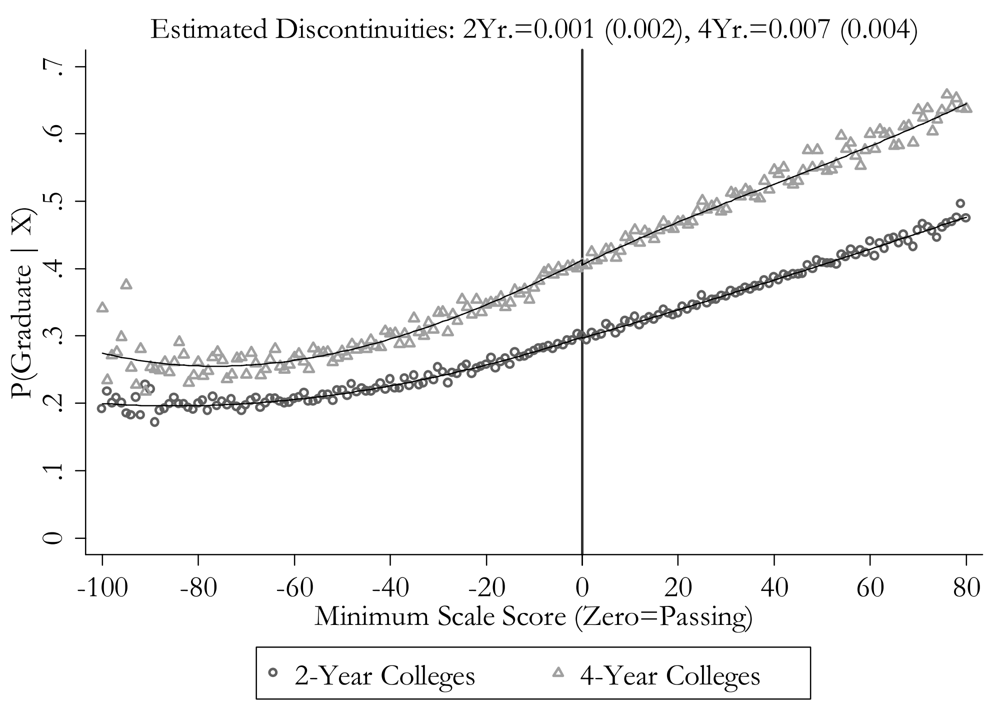


Figure 3: Probability of Remediation in Any Subject by TASP Scale Score

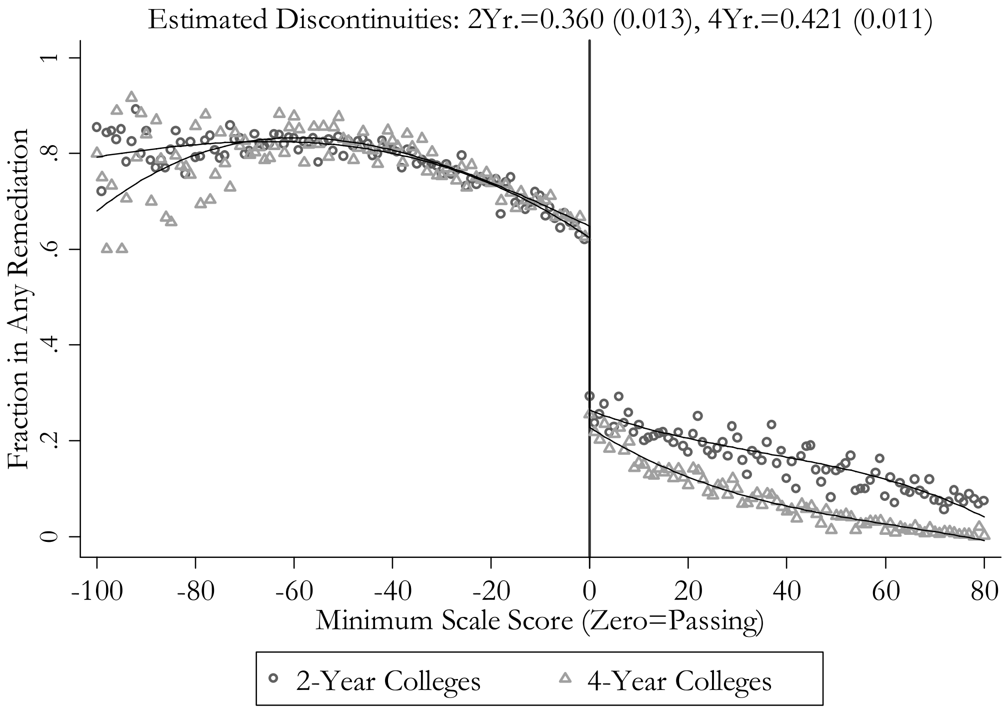


Figure 4: Academic Credits Attempted in First Year by TASP Scale Score

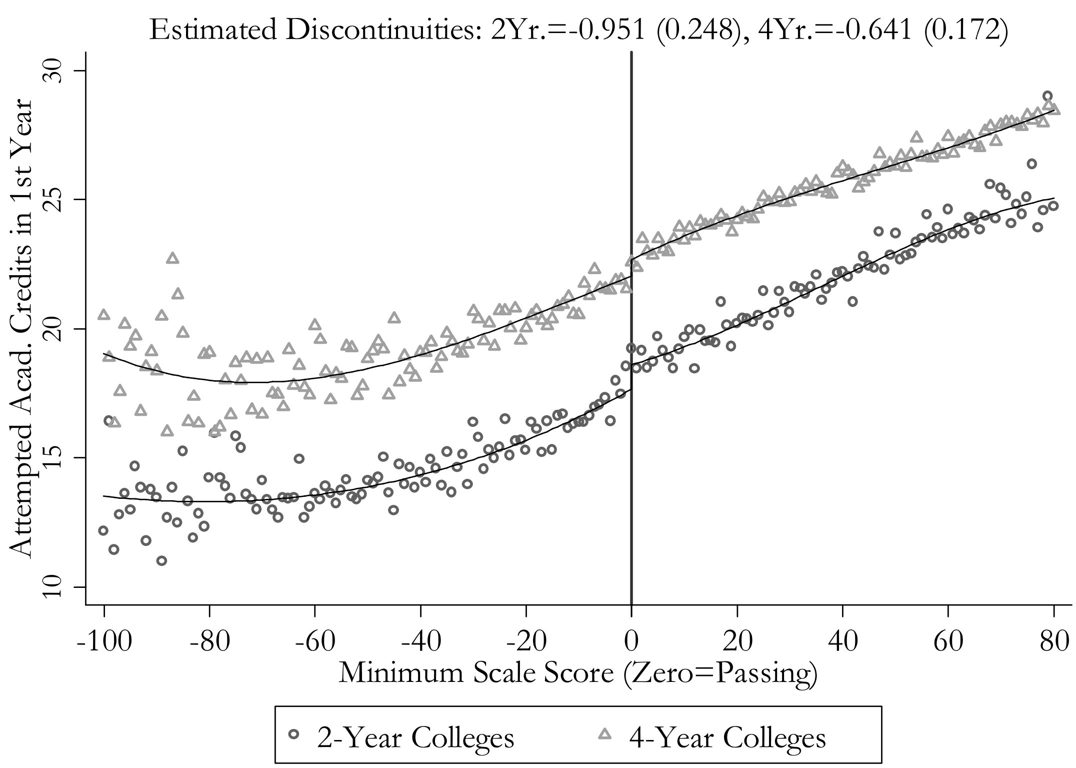


Figure 5: Fraction of Transferring up to 4-Year School (from 2-Year School) and Transferring Down to 2-Year School (from 4-Year School) by TASP Scale Score

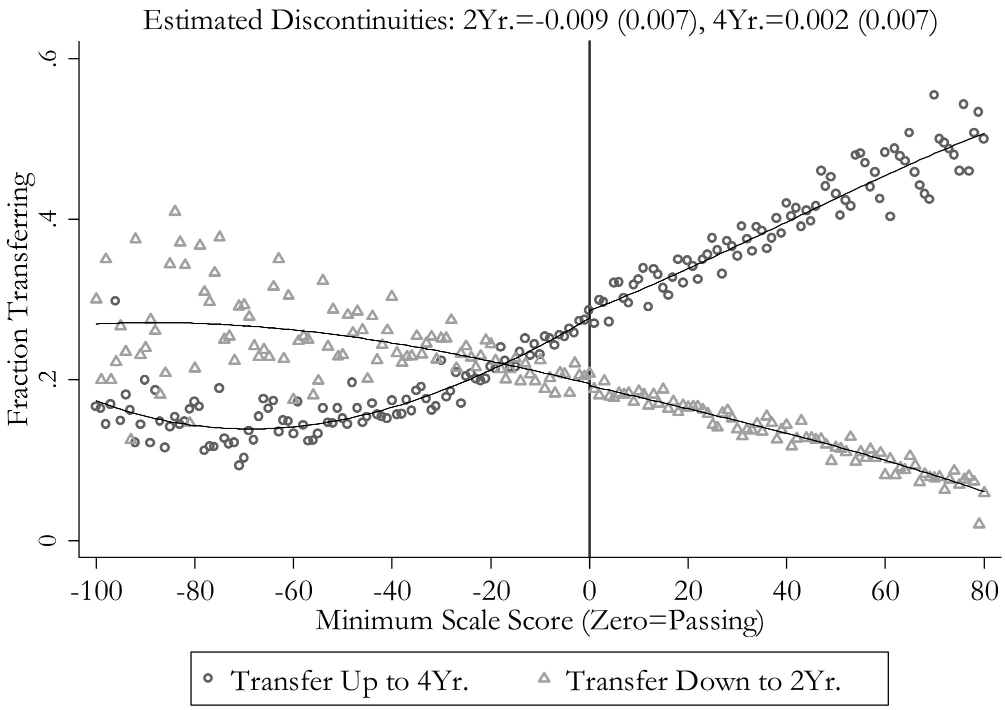


Figure 6: Fraction Completing at Least 1 Year of College by TASP Scale Score

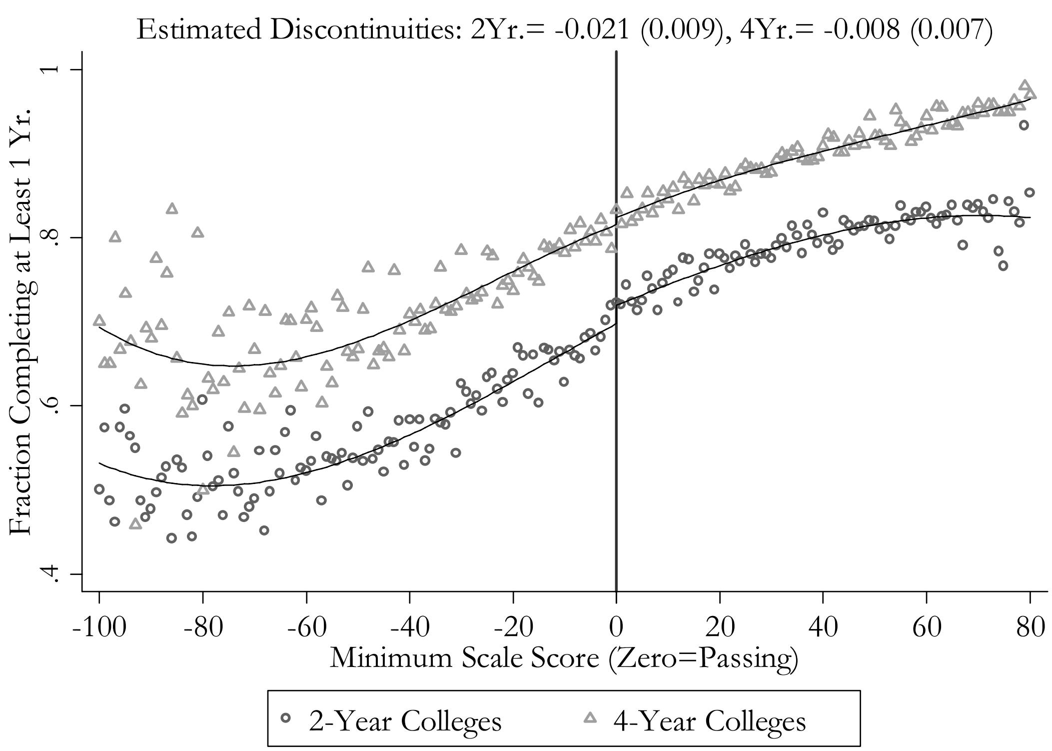


Figure 7: Fraction Graduating within 6 Years by TASP Scale Score

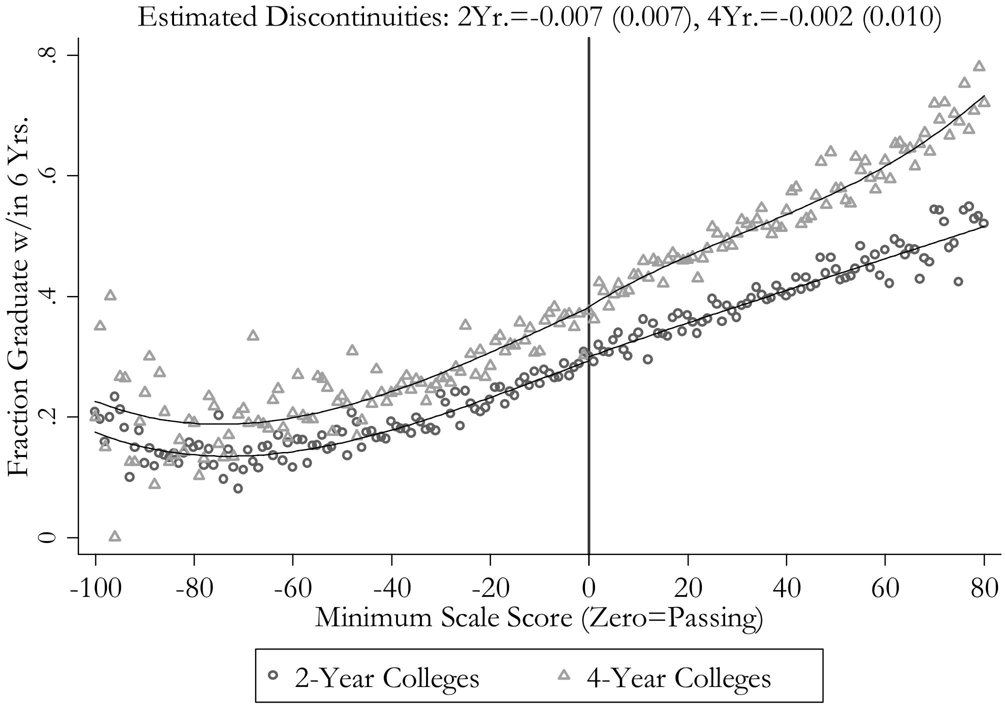
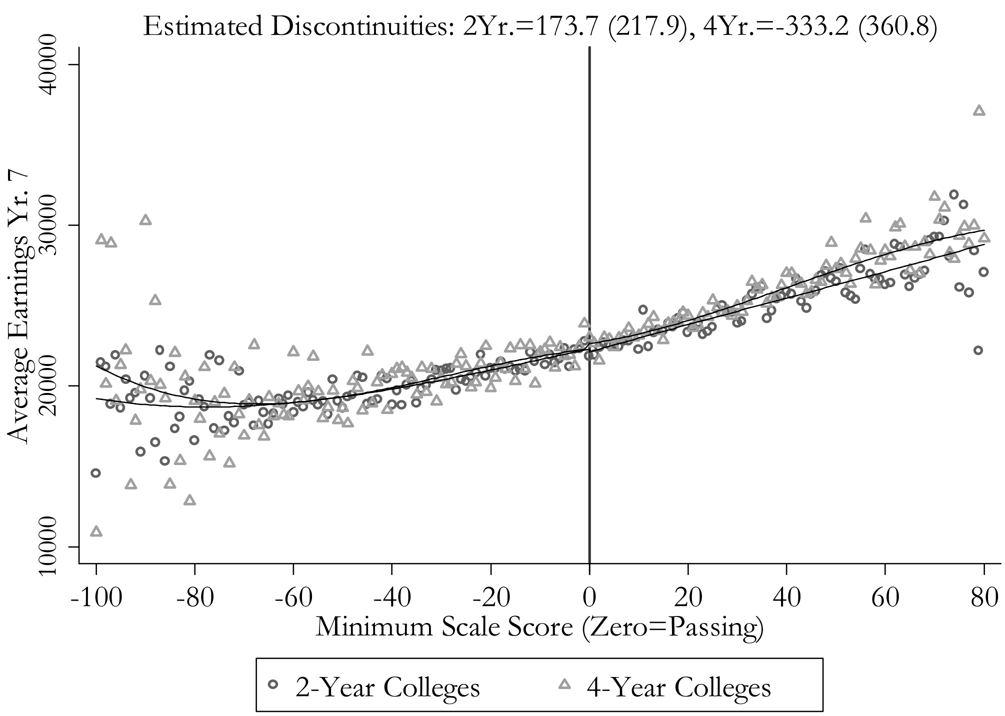
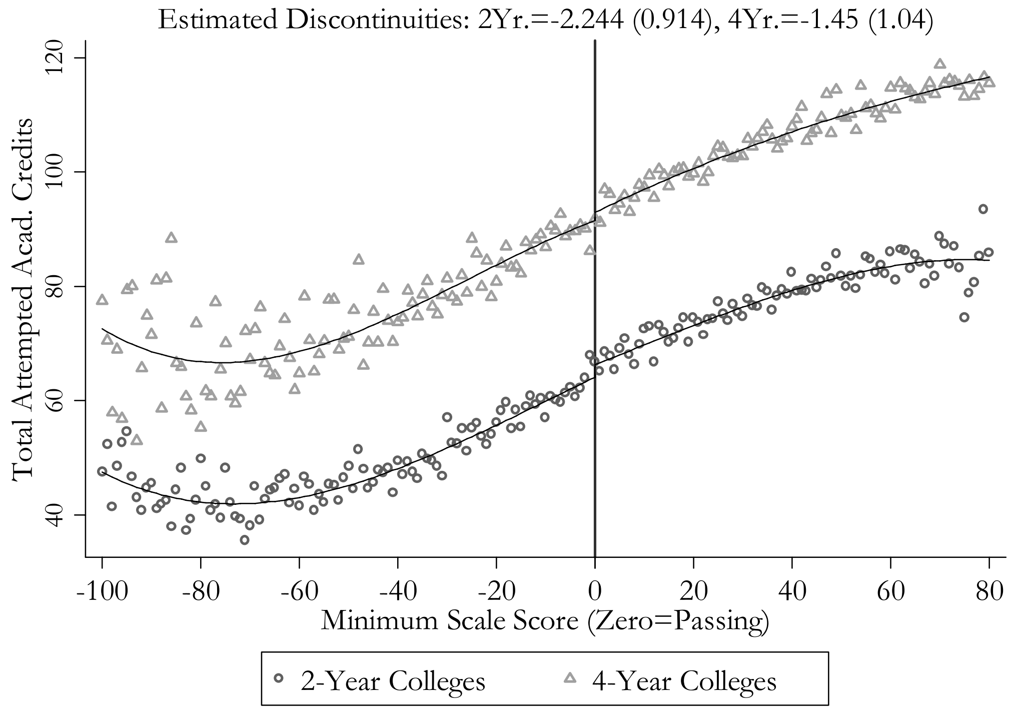


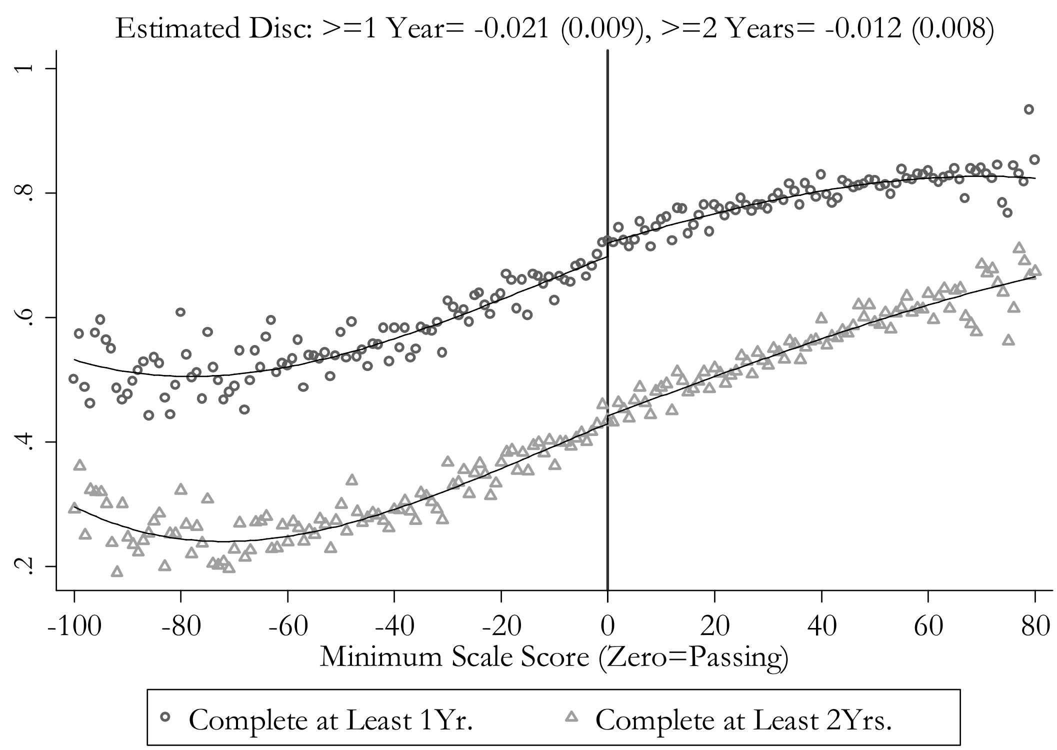
Figure 8: Earnings in Year 7 by TASP Scale Score, Conditional on Positive Earnings



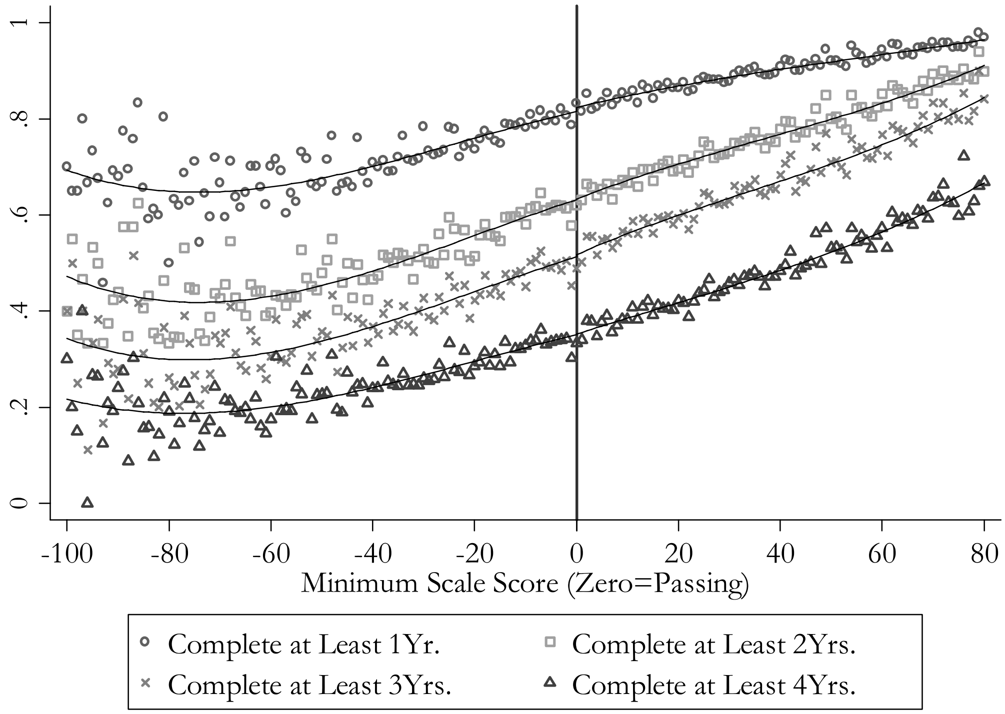
Appendix Figure 1: Total Attempted Academic Credits by TASP Scale Score



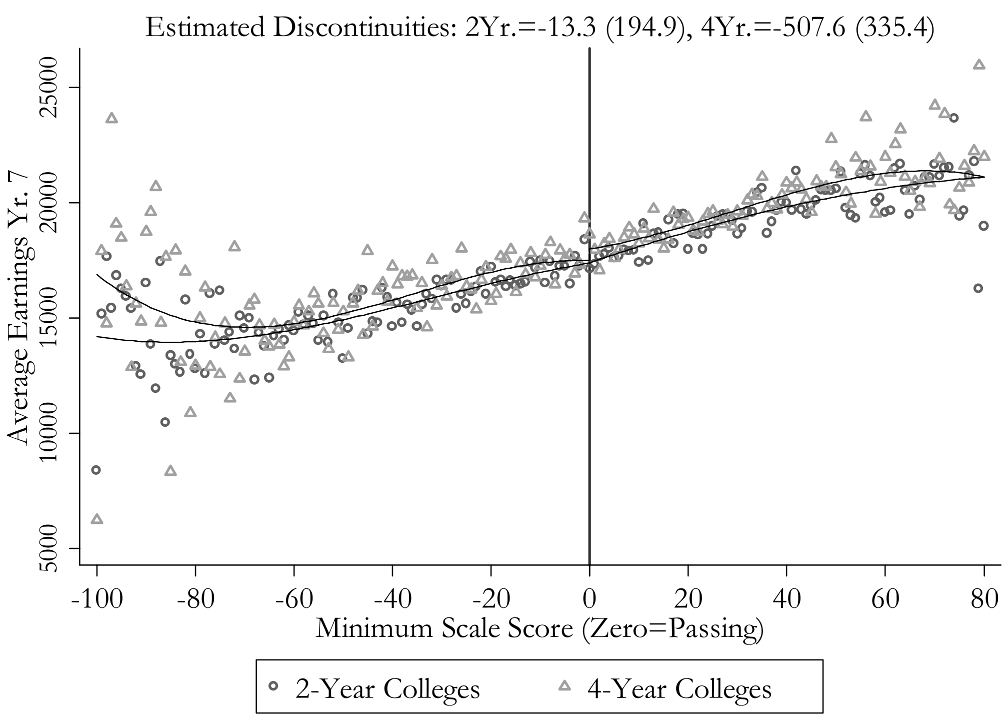
Appendix Figure 2: Years of College Completed by TASP Scale Score, 2-Year Colleges



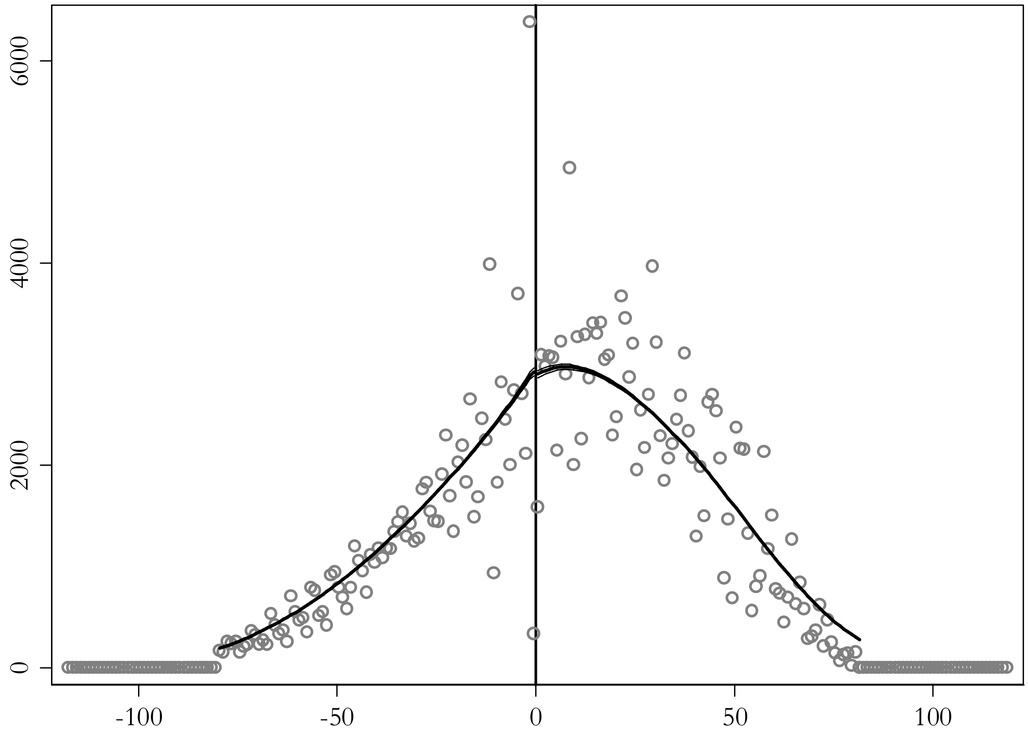
Appendix Figure 3: Years of College Completed by TASP Scale Score, 4-Year Colleges



Appendix Figure 4: Earnings in Year 7 by TASP Scale Score, Includes Observations with Zero Earnings



Appendix Figure 5: Estimated Density of Rescaled TASP Scores, 2-Year Colleges



Notes: Figure shows the estimated density (scaled up by the total number of observations) and the standard error bands produced by the method proposed in McCrary (2008) superimposed over the number of observations in a test score cell. (2-year colleges)

Table 1: Sample Means by Remediation Status

	2-Year Colleges			4-Year Colleges		
	All	Remed.	Non-Remed.	All	Remed.	Non-Remed.
Math Remediation	0.35	0.85	0.00	0.19	0.85	0.00
Reading Remediation	0.11	0.28	0.00	0.06	0.25	0.00
Writing Remediation	0.08	0.20	0.00	0.03	0.12	0.00
Pass TASP	0.60	0.28	0.83	0.79	0.30	0.93
Attempted Academic Credits	18.63	16.07	20.52	24.42	20.44	25.52
in First Year*	(10.32)	(9.49)	(10.50)	(7.59)	(7.81)	(7.15)
Total Attempted Academic	66.82	55.83	74.40	100.40	83.83	105.23
Credits	(48.81)	(44.94)	(49.91)	(47.55)	(51.25)	(45.28)
Transfer Up/Down	0.30	0.20	0.37	0.16	0.22	0.14
Complete at Least 1 Yr.	0.71	0.64	0.76	0.86	0.76	0.89
Complete at Least 2 Yrs.	0.45	0.36	0.52	0.71	0.55	0.76
Complete at Least 3 Yrs.	0.28	0.17	0.35	0.61	0.43	0.66
Complete at Least 4 Yrs.	0.18	0.10	0.23	0.44	0.30	0.48
Graduate within 4 Yrs.	0.17	0.12	0.20	0.16	0.07	0.18
Graduate within 5 Yrs.	0.26	0.18	0.31	0.38	0.22	0.43
Graduate within 6 Yrs.	0.32	0.23	0.38	0.48	0.30	0.53
Positive UI Earnings in Yr 5	0.80	0.79	0.80	0.81	0.81	0.81
Positive UI Earnings in Yr 6	0.79	0.79	0.79	0.80	0.80	0.80
Positive UI Earnings in Yr 7	0.78	0.78	0.78	0.78	0.79	0.77
White	0.68	0.60	0.74	0.62	0.49	0.66
Economically Disadvantaged	0.10	0.12	0.08	0.09	0.16	0.08
Econ. Disadvantaged Missing	0.34	0.39	0.31	0.25	0.29	0.24
Age ≥ 21Years in 1st Semester	0.21	0.27	0.18	0.06	0.11	0.05
Distance from HS < 25 Miles	0.68	0.70	0.67	0.34	0.43	0.31
Distance from HS > 50 Miles	0.15	0.15	0.15	0.55	0.43	0.58
Distance from HS Missing	0.37	0.43	0.34	0.22	0.25	0.21
Receive In-District Tuition	0.58	0.60	0.56			
Enrolled in 1995 or Earlier	0.44	0.40	0.47	0.53	0.37	0.58
Started College Fall Semester	0.63	0.66	0.61	0.86	0.82	0.87
Rescaled Max(Reading,Math)	33.91	22.23	41.97	45.24	24.65	51.25
Score	(26.01)	(27.25)	(21.72)	(24.05)	(25.47)	(19.95)
Rescaled Min(Reading, Math)	6.35	-13.49	20.03	23.21	-9.86	32.86
Score	(33.00)	(30.80)	(26.97)	(31.19)	(28.52)	(24.64)
Sample Size*	255,878	104,405	151,473	197,502	44,617	152,885

* - Analyses that use attempted credits in first year, grade in first college-level math course, and earnings in year 7 require additional sample restrictions and use fewer observations than listed in the Table. "Remediated" and "Non-Remediated" columns refer to remediation in any subject area. See text for additioanl details.

Table 2: Estimated Discontinuities in Baseline Characteristics

	2-Year Colleges		4-Year Colleges	
Test Score Cell Size+	1774.1	3125.3*	944.8	1713.7
	(1175.1)	(1539.4)	(727.1)	(907.5)
Adjusted Test Score Cell Size+	306.1	777.7	52.8	391.7
	(424.2)	(543.3)	(263.0)	(289.1)
Predicted Prob(Graduation | X)	0.001	0.001	0.007	0.004
	(0.002)	(0.003)	(0.004)	(0.004)
Fitted E(Academic Credits | X)	-0.060	0.133	0.315	0.064
	(0.350)	(0.465)	(0.426)	(0.522)
Pred. Prob(Complete ≥ 1 Yr | X)	0.000	0.001	0.001	0.000
	(0.003)	(0.004)	(0.003)	(0.003)
Pred Prob(Complete ≥ 4 Yrs | X)	-0.000	0.000	0.003	0.002
	(0.002)	(0.002)	(0.003)	(0.005)
Fitted E(Earnings in Year 7 | X)	0.559	-8.470	19.910	-59.163
	(74.993)	(101.474)	(137.624)	(138.004)
Start in Fall Semester	0.004	0.018	0.001	-0.005
	(0.013)	(0.012)	(0.010)	(0.013)
Non-Hispanic White	-0.002	-0.003	0.003	0.006
	(0.005)	(0.007)	(0.009)	(0.010)
Age ≥ 21Years in First Semester	0.006	0.003	0.002	0.006
	(0.006)	(0.008)	(0.006)	(0.006)
Max Subject Score	0.174	0.039	-0.337	-0.153
	(0.593)	(0.872)	(0.854)	(1.402)
Economically Disadvantaged	0.001	0.003	-0.009	-0.012
	(0.004)	(0.005)	(0.005)	(0.007)
Missing Economically Disadvantaged	0.004	-0.003	0.012	0.004
	(0.006)	(0.006)	(0.010)	(0.013)
Entered College in 1995 or Earlier	-0.069	-0.075	-0.058	-0.037
	(0.056)	(0.090)	(0.056)	(0.093)
College <25 Miles from HS	-0.015	-0.008	-0.012	-0.011
	(0.008)	(0.012)	(0.008)	(0.009)
College >50 Miles from HS	0.005	0.007	0.006	0.004
	(0.004)	(0.005)	(0.012)	(0.016)
Distance from HS Mising	0.006	-0.004	0.007	0.010
	(0.006)	(0.008)	(0.009)	(0.013)
Receive In-District Tuition	0.000	-0.002		
	(0.007)	(0.008)		
Test Score Range	Global	Narrow Band	Global	Narrow Band
Num. Observations	255,878	59,344	197,502	33,910

Notes: Cell entries are least squares estimates of discontinuity in the conditional expectation of a given covariate (i.e., the coefficient on "fail placement exam"). Statistical significance at 1% and 5% level denoted by **, *, respectively. The running variable is the minimum of the math and reading scores. Standard errors adjusted for clustering at the test score level in parentheses. See text for details concerning specification. +: Estimated discontinuity in the mean cell size obtained from test score cell-level regression weighted by the cell size using the same parametric specification as that used for other covariates.

Table 3: IV Estimates of Effect of Remediation on Academic Outcomes

Dependent Variable	**First Stage** 2-Year Colleges				4-Year Colleges			
Remediation in Any Subject	0.360**	0.360**	0.371**	0.376**	0.421**	0.409**	0.426**	0.425**
	(0.013)	(0.015)	(0.010)	(0.013)	(0.011)	(0.014)	(0.009)	(0.013)
	IV Estimates of Effect of Remediation in Any Subject							
	2-Year Colleges				4-Year Colleges			
Academic Credit Hours:								
Attempted in First Year+	-2.407**	-2.166**	-2.421**	-2.042**	-1.500**	-1.654*	-1.625**	-1.717**
	(0.636)	(0.810)	(0.394)	(0.418)	(0.402)	(0.659)	(0.343)	(0.530)
Total Attempted	-6.239*	-3.208	-6.068**	-3.956	-3.447	-6.906*	-4.475*	-7.588**
	(2.589)	(3.151)	(1.890)	(2.151)	(2.463)	(3.428)	(1.998)	(2.712)
Transferring:								
Up to 4-Yr (from 2 Yr.) or Down to 2-Yr. (from 4 Yr.)	-0.025	-0.008	-0.027	-0.013	0.006	0.037	0.013	0.043**
	(0.021)	(0.024)	(0.018)	(0.020)	(0.017)	(0.020)	(0.016)	(0.016)
College Attainment:								
At Least 1 Year	-0.060*	-0.029	-0.059**	-0.035*	-0.019	-0.035	-0.025	-0.038*
	(0.024)	(0.027)	(0.018)	(0.017)	(0.017)	(0.023)	(0.014)	(0.019)
At Least 2 Years	-0.034	-0.001	-0.034	-0.008	-0.004	-0.049*	-0.017	-0.060**
	(0.023)	(0.025)	(0.018)	(0.019)	(0.022)	(0.024)	(0.019)	(0.018)
At Least 3 Years	-0.025	-0.017	-0.028	-0.023	-0.010	-0.040	-0.028	-0.054
	(0.019)	(0.021)	(0.016)	(0.016)	(0.025)	(0.040)	(0.021)	(0.029)
At Least 4 Years	-0.018	-0.003	-0.018	-0.005	-0.005	-0.031	-0.015	-0.042
	(0.014)	(0.017)	(0.012)	(0.014)	(0.020)	(0.027)	(0.017)	(0.023)
Graduate Within 4 Years	-0.013	-0.002	-0.018	-0.004	0.016	0.009	0.003	-0.002
	(0.014)	(0.015)	(0.012)	(0.013)	(0.018)	(0.025)	(0.014)	(0.019)
Graduate Within 5 Years	-0.012	-0.015	-0.017	-0.020	0.001	-0.008	-0.021	-0.028
	(0.017)	(0.017)	(0.015)	(0.014)	(0.025)	(0.040)	(0.018)	(0.026)
Graduate Within 6 Years	-0.020	-0.024	-0.023	-0.029	-0.004	-0.021	-0.023	-0.040
	(0.020)	(0.022)	(0.016)	(0.017)	(0.024)	(0.039)	(0.020)	(0.028)
Control for Baseline Covariates?	No	N o	Yes	Yes	No	N o	Yes	Yes
Test Score Range	Global	Narrow Band	Global	Narrow Band	Global	Narrow Band	Global	Narrow Band

Notes: Statistical significance at 1% and 5% level denoted by **, *, respectively. Estimates in first row are least squares estimates of discontinuity in probability of being in remediation for any subject. The remainder of the rows show 2SLS IV estimates of the effect of remediation for any subject. Standard errors adjusted for clustering at the test score level in parentheses. The running variable is the minimum of the math and reading scores. Models with covariates include controls for max(math, reading) score, dummies for white, Hispanic, starting in fall semester, being 21 or older in first semester, academic year of first semester, academic year student initially took the TASP, economically disadvantaged, missing economically disadvantaged, receiving in-district tuition, missing data for in-district tuition status, distance from HS <25 miles, distance from HS>50 miles, distance from HS missing, starting college more than 1 semester after initially taking the TASP test. +: Additional sample restrictions made for these outcomes (see text for details).

Table 4: IV Estimates of the Effect of Remediation on Labor Market Outcomes

		All Observations		Restricted to Positive Earnings		Controls for Inverse Mills Ratio	
2-Year Colleges							
Year 5	Mean Earnings	12857.1		16134.9		16134.9	
	Disc. in Prob(Pos Earnings)			-0.003	-0.009		
				(0.005)	(0.007)		
	Effect of Remed. (IV)	-126.5	-697.1	3.2	-349.6	-328.6	-734.7
		(413.6)	(516.7)	(442.3)	(571.4)	(858.4)	(1055.4)
Year 6	Mean Earnings	15658.2		19834.8		19834.8	
	Disc. in Prob(Pos Earnings)			-0.005	-0.015**		
				(0.004)	(0.005)		
	Effect of Remed. (IV)	-416.9	-698.5	-205.1	209.4	-692.9	-850.2
		(487.6)	(602.4)	(544.3)	(628.7)	(829.9)	(3964.9)
Year 7+	Mean Earnings	18191.1		23340.5		23340.5	
	Disc. in Prob(Pos Earnings)			-0.006	-0.012*		
				(0.005)	(0.005)		
	Effect of Remed. (IV)	-86.1	-469.1	431.8	445.3	-212.4	-213.9
		(597.3)	(749.2)	(684.3)	(839.0)	(941.9)	(1782.0)
4-Year Colleges							
Year 5	Mean Earnings	11593.7		14320.7		14320.7	
	Disc. in Prob(Pos Earnings)			0.005	0.010		
				(0.005)	(0.007)		
	Effect of Remed. (IV)	-14.2	466.6	-190.8	281.7	139.5	683.0
		(560.1)	(749.2)	(633.0)	(787.0)	(761.5)	(2021.4)
Year 6	Mean Earnings	16218.0		20377.8		20377.8	
	Disc. in Prob(Pos Earnings)			0.008	0.009		
				(0.007)	(0.010)		
	Effect of Remed. (IV)	-290.5	69.6	-758.7	-326.3	94.9	87.0
		(628.8)	(901.5)	(675.2)	(1009.4)	(1195.1)	(1828.4)
Year 7+	Mean Earnings	19503.9		25093.8		25093.8	
	Disc. in Prob(Pos Earnings)			-0.011	-0.008		
				(0.008)	(0.011)		
	Effect of Remed. (IV)	-1413.9	-858.3	-1001.8	-414.2	-2406.7	-873.5
		(912.8)	(1299.2)	(930.3)	(1384.3)	(1677.9)	(2529.2)
	Test Score Range	Global	Narrow Band	Global	Narrow Band	Global	Narrow Band

Notes: Statistical significance at 1% and 5% level denoted by **, *, respectively. Cell entries are estimated effect of remediation for any subject on labor market outcomes. Year refers to the number of years since first enrolling in college. For a given year, the first row contains mean earnings, the second and third rows contain the estimated discontinuity in the probability of positiv earnings and its standard error, the fourth and fifth rows contains IV estimate of the effect of remediation on earnings and the standard error. Standard errors adjusted for clustering at the test score level in parentheses. The test score used in all models is the minimum of the math and reading scores. All models estimated with covariates listed in the notes to Table 3. +: Year 7 estimates restricted to students enrolled in college in fall of 1997 or earlier.

Table 5: IV Estimates of the Effect of Remediation on Selected Outcomes by Subgroup

Outcomes:	Academic						Earnings in Year 7					
	Complete ≥ 1 Yr. of College		Attempted Acad. Credits		Graduate within 6 Years		Prob(Pos. Earnings)		All Observations		Restricted to Pos. Earnings	
Subgroup	2-Year	4-Year	2-Year	4-Year	2-Year	4-Year	2-Year	4-Year	2-Year	4-Year	2-Year	4-Year
Passing Standard												
Lower (Bef. Sep. 95)	-0.048	0.039	-4.290	2.169	-0.016	-0.030	-0.005	-0.015	-41	-1930	445	-1081
	(0.028)	(0.032)	(2.750)	(4.008)	(0.023)	(0.032)	(0.006)	(0.012)	(870)	(1614)	(984)	(1677)
Higher (After Sep. 95)	-0.061*	-0.053*	-7.158**	-7.398*	-0.030	-0.031	-0.007	-0.004	-428	-702	145	-697
	(0.025)	(0.021)	(2.563)	(2.892)	(0.022)	(0.030)	(0.007)	(0.011)	(743)	(1050)	(841)	(1074)
Institution												
High Remed. School	-0.122**	-0.024	-9.642**	-5.231	-0.074**	-0.031	-0.004	-0.019	1465	-1286	2428	-16
	(0.026)	(0.028)	(2.394)	(3.304)	(0.025)	(0.049)	(0.011)	(0.015)	(1601)	(2104)	(1787)	(2158)
Low Remed. School	-0.051	-0.017	-3.889	-0.290	-0.060	-0.025	-0.003	-0.013	-1450	-5514**	-1806	-6118**
	(0.029)	(0.036)	(3.763)	(4.455)	(0.033)	(0.048)	(0.014)	(0.023)	(1779)	(1947)	(1467)	(2393)
Demographics												
Black or Hispanic	-0.074**	-0.023	-5.797**	-3.884	-0.039*	-0.017	-0.009	-0.014	-1156	-644	-678	196
	(0.022)	(0.024)	(2.217)	(3.195)	(0.020)	(0.028)	(0.010)	(0.011)	(1544)	(880)	(1785)	(1215)
Econ. disadvantaged	-0.033	-0.010	-3.733	-10.032*	-0.047	-0.053	-0.001	0.007	1149	-364	1442	-932
	(0.032)	(0.038)	(3.970)	(4.641)	(0.039)	(0.048)	(0.018)	(0.026)	(2296)	(2286)	(2437)	(2909)
Age ≥ 21 Yrs. when	-0.062	-0.047	-3.994	0.900	-0.012	0.014	-0.003	-0.036	363	-8345*	714	-9733*
started college	(0.037)	(0.069)	(2.488)	(6.492)	(0.028)	(0.056)	(0.012)	(0.030)	(1682)	(3569)	(2141)	(4266)
Male	-0.065**	-0.021	-7.394**	-6.943*	-0.021	-0.048*	-0.007	-0.008	-869	-445	-440	126
	(0.020)	(0.028)	(2.034)	(3.318)	(0.016)	(0.024)	(0.008)	(0.012)	(1141)	(1321)	(1292)	(1677)
Female	-0.056*	-0.029	-5.222*	-2.920	-0.024	-0.010	-0.005	-0.012	338	-1996	898	-1663
	(0.027)	(0.016)	(2.510)	(2.294)	(0.024)	(0.025)	(0.007)	(0.010)	(808)	(1041)	(814)	(928)

Notes: Statistical significance at 1% and 5% level denoted by **, *, respectively. Cell entries are 2SLS estimates of the effect of being in remediation for any subject. Sample limited to students in the indicated subgroup. The test score used in all models is the minimum of the math and reading score. All models estimated with covariates listed in the notes to Table 3. Standard errors adjusted for clustering at the test score level in parentheses. See text for details concerning regression specification.

Table 6: IV Estimates for Selected Outcomes by Remediation Subject Area

	2-Year Colleges				4-Year Colleges			
	Math Remediation		Reading Remediation		Math Remediation		Reading Remediation	
Remediation	0.379**	0.383**	0.284**	0.277**	0.446**	0.460**	0.352**	0.309**
	(0.013)	(0.017)	(0.010)	(0.013)	(0.011)	(0.013)	(0.016)	(0.021)
Total Attempted Acad. Credits	-4.704*	-3.187	-1.159	-0.497	-2.446	-2.698	-2.864	-3.058
	(2.162)	(2.614)	(4.317)	(6.324)	(2.071)	(3.053)	(4.411)	(6.909)
Transfer Up to 4-Yr (from 2 Yr.) or	-0.008	-0.003	-0.020	-0.057	-0.004	0.003	-0.033	-0.010
Down to 2-Yr. (from 4 Yr.)	(0.017)	(0.023)	(0.028)	(0.037)	(0.014)	(0.016)	(0.027)	(0.041)
Complete ≥ 1 Yr. of College	-0.048*	-0.019	0.040	0.081	-0.020	-0.022	-0.017	-0.023
	(0.021)	(0.022)	(0.041)	(0.063)	(0.017)	(0.024)	(0.032)	(0.052)
Graduate Within 6 Years	-0.032	-0.030	-0.007	-0.021	0.015	-0.000	-0.033	-0.013
	(0.017)	(0.021)	(0.031)	(0.043)	(0.017)	(0.025)	(0.031)	(0.055)
Prob(Pos Earnings) in Year 7	-0.001	-0.009*	-0.000	-0.002	-0.012	-0.010	-0.002	0.015
	(0.004)	(0.005)	(0.007)	(0.009)	(0.006)	(0.009)	(0.010)	(0.012)
Earnings in Year 7 (All Obs.)	-700.8	-1279.4**	522.3	1140.2	-86.2	17.6	-2190.2*	-326.5
	(452.5)	(448.8)	(802.4)	(988.0)	(662.4)	(863.5)	(1060.7)	(1681.0)
Earnings in Year 7 (Pos. Earnings)	-931.1	-806.3	553.4	1536.1	683.8	813.7	-2726.9**	-1925.0
	(574.3)	(542.3)	(1077.2)	(1442.9)	(685.9)	(991.9)	(966.3)	(1435.1)
Prob(Pos Earnings) in Year 6	0.000	-0.005	-0.002	-0.000	0.007	0.017	0.005	0.005
	(0.003)	(0.003)	(0.005)	(0.009)	(0.006)	(0.010)	(0.009)	(0.010)
Earnings in Year 6 (All Obs.)	-249.5	34.9	489.1	883.4	336.0	642.2	-1206.6	-598.0
	(467.3)	(531.6)	(720.6)	(881.3)	(480.4)	(738.5)	(806.0)	(1294.9)
Earnings in Year 6 (Pos. Earnings)	-442.0	431.0	809.4	1275.6	68.7	-1.8	-1813.1*	-1144.5
	(543.2)	(590.5)	(762.8)	(921.3)	(577.1)	(944.8)	(826.0)	(1442.0)
Test Score Range	Global	Narrow Band	Global	Narrow Band	Global	Narrow Band	Global	Narrow Band

Notes: **, * Denote statistical significance at 1% and 5%, respectively. Cell entries are IV estimates of the effect of remediation in a subject area. IV estimates use the math test score for the "Math Remediation" panel and reading test scores for the "Reading Remediation" panel. The "treatment" for the IV estimates is remediation in math for the "Math Remediation" panel and remediation in reading for the "Reading Remediation" panel. Standard errors adjusted for clustering at the test score level in parentheses. See text for additional details concerning regression specification.

Table A1: IV Estimates of the Effect of Remediation on Selected Outcomes by Subgroup (Narrow Band)

Outcomes:	Academic						Earnings in Year 7					
	Complete ≥ 1 Yr. of College		Attempted Acad. Credits		Graduate within 6 Years		Prob(Pos. Earnings)		All Observations		Restricted to Pos. Earnings	
Subgroup	2-Year	4-Year	2-Year	4-Year	2-Year	4-Year	2-Year	4-Year	2-Year	4-Year	2-Year	4-Year
Passing Standard												
Lower (Bef. Sep. 95)	-0.006	0.077	-0.053	5.872	-0.023	0.027	-0.014**	-0.013	-979	-162	103	1331
	(0.021)	(0.053)	(2.298)	(5.118)	(0.024)	(0.053)	(0.005)	(0.016)	(1253)	(2650)	(1444)	(2648)
Higher (After Sep. 95)	-0.046	-0.074**	-5.559*	-11.683**	-0.029	-0.061	-0.009	-0.003	75	-1251	779	-1445
	(0.024)	(0.023)	(2.615)	(3.341)	(0.020)	(0.033)	(0.006)	(0.011)	(650)	(1057)	(636)	(1257)
Institution												
High Remed. School	-0.113**	0.006	-5.631	-11.258*	-0.073*	-0.100	-0.004	-0.030	309	-3482	827	-1620
	(0.025)	(0.044)	(3.079)	(4.561)	(0.033)	(0.067)	(0.012)	(0.018)	(1746)	(2629)	(1903)	(2912)
Low Remed. School	-0.023	-0.058	-0.351	-3.163	0.005	-0.036	0.011	0.025	1904	-2835	1557	-5131
	(0.034)	(0.031)	(4.372)	(4.440)	(0.042)	(0.079)	(0.018)	(0.033)	(2432)	(2655)	(1973)	(3727)
Demographics												
Black or Hispanic	-0.056*	-0.043	-5.448*	-9.513**	-0.072**	-0.036	-0.007	-0.009	3232*	-351	4495**	257
	(0.023)	(0.035)	(2.573)	(3.645)	(0.025)	(0.034)	(0.010)	(0.015)	(1492)	(1274)	(1742)	(1791)
Econ. disadvantaged	-0.027	-0.016	-3.384	-15.984**	-0.010	-0.062	-0.004	0.031	2268	4505	2804	3239
	(0.035)	(0.047)	(4.787)	(6.023)	(0.047)	(0.059)	(0.020)	(0.037)	(2484)	(2320)	(2504)	(2881)
Age ≥ 21 Yrs. when started college	-0.007	-0.002	-0.258	10.428	0.008	0.044	-0.005	-0.039	82	-8521	642	-10283
	(0.045)	(0.107)	(2.828)	(10.651)	(0.037)	(0.085)	(0.014)	(0.032)	(2273)	(4485)	(2380)	(7069)
Male	-0.062*	-0.072	-6.509**	-14.665**	-0.030	-0.070*	-0.001	-0.012	642	2162	944	3969
	(0.025)	(0.040)	(2.412)	(4.842)	(0.016)	(0.030)	(0.008)	(0.015)	(1396)	(1749)	(1558)	(2155)
Female	-0.020	-0.021	-2.375	-3.918	-0.031	-0.028	-0.018*	-0.005	-945	-2479	386	-2772*
	(0.030)	(0.018)	(3.135)	(2.844)	(0.026)	(0.031)	(0.008)	(0.013)	(1005)	(1279)	(1095)	(1153)

Notes: **, * Denote statistical significance at 1% and 5%, respectively. Cell entries are 2SLS estimates of the effect of being in remediation for at least one subject. Sample limited to students in the indicated subgroup. The test score used in all models is the minimum of the math and reading score (these estimates are not adjusted for other covariates). Standard errors adjusted for clustering at the test score level in parentheses. See text for details concerning regression specification.

Table A2: IV Estimates of Math and Reading Remediation on Selected Outcomes, Math and Reading Effects Estimated in Single Equation

Outcomes:	2-Year Colleges		4-Year Colleges	
	Math	Reading	Math	Reading
Total Attempted Acad. Credits	-4.566*	-1.403	-2.411	-3.224
	(2.079)	(4.428)	(2.101)	(4.598)
Transfer Up to 4-Yr (from 2 Yr.) or	-0.008	-0.018	-0.005	-0.036
Down to 2-Yr. (from 4 Yr.)	(0.016)	(0.030)	(0.013)	(0.029)
Complete ≥ 1 Year of College	-0.044*	0.038	-0.020	-0.022
	(0.021)	(0.040)	(0.017)	(0.033)
Graduate Within 6 Years	-0.032	-0.008	0.013	-0.029
	(0.018)	(0.032)	(0.018)	(0.033)
Prob(Pos Earnings) in Year 7	0.002	0.001	-0.030	-0.006
	(0.013)	(0.026)	(0.019)	(0.028)
Earnings in Year 7 (All Obs.)	-593.9	587.4	-115.5	-2043.1
	(417.7)	(749.7)	(785.5)	(1053.8)
Earnings in Year 7 (Pos. Earnings)	-889.1	633.5	596.2	-2465.7**
	(519.3)	(1054.2)	(802.8)	(923.9)
Prob(Pos Earnings) in Year 6	0.003	-0.008	0.017	0.015
	(0.011)	(0.022)	(0.012)	(0.024)
Earnings in Year 6 (All Obs.)	-177.7	539.9	338.6	-998.6
	(491.5)	(687.1)	(520.7)	(768.9)
Earnings in Year 6 (Pos. Earnings)	-385.2	910.1	28.5	-1545.6
	(594.7)	(749.5)	(637.4)	(879.1)

Notes: Cell entries are IV estimates of the effect of remediation in math and reading in a single equation. Both math and reading IV estimates are based on use of their respective test scores. Standard errors are adjusted for multi-clustering (Cameron, Gelbach and Miller, 2006) at the test score level in parentheses. See text for additional details concerning regression specification. All models are specified as global polynomials.

Table A3: IV Estimates of Remediation on Difficulty Index for Major Field of Study, Condtional on Graduating Within 6 Years

Outcomes:	**Math Index**				**Reading Index**			
Samples:	2-Year Colleges		4-Year Colleges		2-Year Colleges		4-Year Colleges	
Global Polynomial								
Remediation	-0.045	-0.060	-0.065	-0.085	-0.012	0.002	-0.011	-0.009
	(0.073)	(0.073)	(0.050)	(0.050)	(0.081)	(0.067)	(0.057)	(0.055)
Math Remediation	0.012	0.003	-0.057	-0.111*	0.000	0.041	-0.083	-0.060
	(0.061)	(0.059)	(0.053)	(0.054)	(0.082)	(0.065)	(0.046)	(0.044)
Reading Remediation	-0.001	-0.063	0.145	0.149	0.094	-0.009	0.125	0.135
	(0.105)	(0.089)	(0.132)	(0.141)	(0.101)	(0.097)	(0.080)	(0.083)
Narrow Band Sample								
Remediation	-0.143	-0.130	-0.124	-0.118*	-0.044	-0.022	0.005	0.011
	(0.083)	(0.080)	(0.069)	(0.059)	(0.097)	(0.074)	(0.083)	(0.073)
Math Remediation	-0.061	-0.067	-0.045	-0.077	-0.037	-0.010	-0.013	-0.004
	(0.063)	(0.060)	(0.081)	(0.070)	(0.106)	(0.088)	(0.066)	(0.056)
Reading Remediation	-0.250	-0.343**	0.041	0.071	-0.121	-0.236	-0.026	-0.038
	(0.152)	(0.097)	(0.245)	(0.216)	(0.142)	(0.147)	(0.115)	(0.125)
Controls for Baseline Covariates?	N	Y	N	Y	N	Y	N	Y

Notes: **, * Denote statistical significance at 1% and 5%, respectively. Cell entries are IV estimates of the effect of remediation, math remediation, and reading remediation on indices for difficulty of major field of study. The math and reading index are z-scores (M=0, SD=1) constructed using 2-digit U.S. Department of Education Classification of Instructional Programs (CIP). The indices were generated using the average standardized math TASP score in a CIP cell, and the other using the average standardized reading TASP score. Standard errors are adjusted for clustering at the test score level in parentheses. See text for additional details concerning regression specification.

Table A4: IV Estimates of Remediation, Controlling for First-Year Credits

	2-Year Colleges		4-Year Colleges	
Academic Credit Hours:				
Total Attempted	-7.115**	-7.382**	-3.772*	-6.780**
	(1.581)	(1.891)	(1.870)	(1.892)
Transferring:				
Up to 4-Yr (from 2 Yr.) or	-0.026	-0.030	0.018	0.044*
Down to 2-Yr. (from 4 Yr.)	(0.018)	(0.022)	(0.017)	(0.019)
College Attainment:				
At Least 1 Year	-0.078**	-0.075**	-0.026*	-0.041**
	(0.013)	(0.014)	(0.011)	(0.012)
At Least 2 Years	-0.031	-0.035	-0.015	-0.056**
	(0.016)	(0.021)	(0.019)	(0.016)
At Least 3 Years	-0.027	-0.036	-0.023	-0.042
	(0.015)	(0.018)	(0.020)	(0.024)
At Least 4 Years	-0.009	-0.020	-0.003	-0.038
	(0.012)	(0.016)	(0.016)	(0.021)
Graduate Within 4 Years	-0.020	-0.011	-0.004	-0.008
	(0.012)	(0.011)	(0.017)	(0.025)
Graduate Within 5 Years	-0.021	-0.027	-0.027	-0.030
	(0.014)	(0.016)	(0.018)	(0.023)
Graduate Within 6 Years	-0.028	-0.036	-0.025	-0.038
	(0.019)	(0.024)	(0.022)	(0.027)
Test Score Range	Global	Narrow Band	Global	Narrow Band

Notes: **, * Denote statistical significance at 1% and 5%, respectively. Estimates show 2SLS IV estimates of the effect of remediation on a given outcome, controlling for total attempted first-year credits. Standard errors are adjusted for clustering at the test score level in parentheses. See notes on Table 3 for additional information on specifications.

Table A5: IV Estimates from Replication of Bettinger and Long (2009) Using Texas Data

	Based on Closest 4-Yr College		Based on Closest College	
	Math	Reading	Math	Reading
First Stage	0.140*	0.089	0.060**	0.040**
	(0.071)	(0.056)	(0.015)	(0.011)
	{0.002}	{0.003}	{0.002}	{0.002}
Academic Credits in First Year	-4.513*	-6.078*	-4.205	-11.935
	(2.101)	(2.592)	(3.229)	(6.344)
	{0.436}	{0.931}	{0.691}	{1.445}
Total Attempted Acad. Credits	6.457	8.849	-32.594*	-74.743*
	(16.432)	(28.033)	(15.165)	(32.840)
	{2.543}	{5.234}	{4.209}	{8.589}
Transfer Up to 4 Yr. or Remain	-0.028	-0.098	0.004	-0.068
at 4 Yr. School	(0.090)	(0.183)	(0.110)	(0.180)
	{0.023}	{0.049}	{0.035}	{0.070}
Complete at Least 1 Yr.	0.080	0.195	-0.198	-0.348
	(0.102)	(0.194)	(0.114)	(0.231)
	{0.018}	{0.039}	{0.030}	{0.062}
Complete at Least 2 Yrs.	0.170	0.325	-0.267	-0.397
	(0.156)	(0.274)	(0.196)	(0.367)
	{0.024}	{0.051}	{0.039}	{0.080}
Complete at Least 3 Yrs.	0.114	0.183	-0.269	-0.465
	(0.119)	(0.191)	(0.222)	(0.420)
	{0.026}	{0.054}	{0.042}	{0.085}
Complete at Least 4 Yrs.	0.094	-0.152	-0.071	-0.309
	(0.079)	(0.149)	(0.130)	(0.187)
	{0.027}	{0.057}	{0.045}	{0.088}
Graduate within 4 Years	0.072	0.232**	-0.290	-0.298
	(0.054)	(0.086)	(0.162)	(0.313)
	{0.022}	{0.047}	{0.035}	{0.066}
Graduate within 5 Years	0.150	0.263	-0.443	-0.659
	(0.113)	(0.186)	(0.268)	(0.505)
	{0.027}	{0.057}	{0.044}	{0.087}
Graduate within 6 Years	0.143	0.217	-0.418	-0.697
	(0.128)	(0.210)	(0.259)	(0.499)
	{0.027}	{0.057}	{0.045}	{0.090}

Note: First stage estimates are the coefficient on the predicted probability of remediaiton at the nearest 4 year college. Remaining entries are IV estimates of the effect of math or reading remediation. Additional controls for SAT verbal and math scores, high school GPA, gender, number of years of HS math, HS class rank, age at first enrollment, family income, college plans in high school, race, academic year indicators, distance from the nearest college, and dummy variables for missing family income (family income from SAT records unavailable prior to 1996). Standard errors adjusted for clustering by the nearest 4 year college in parentheses and unadjusted standard errors in brackets. **, * indicate statistical significance at 1% and 5% levels, respectively (based on cluster-adjusted standard errors)

Manuscript Notes:

June 14, 2013

Publication Information:

Martorell, Paco and Isaac McFarlin Jr. 2011. "Help or Hindrance? The Effects of College Remediation on Academic and Labor Market Outcomes," *The Review of Economics & Statistics*, 93(2): 436-454.

The following tables have been updated due to inconsistent fonts:

- Table 5: IV Estimates of the Effect of Remediation on Selected Outcomes by Subgroup
- Table 6: IV Estimates for Selected Outcomes by Remediation Subject Area