Identifying the Causal Effects of Remedial Writing on Collegiate Performance

Using Placement Exam Grader Toughness

Melissa Famulari*

Department of Economics
University of California – San Diego
La Jolla, CA 92093-0508
mfamulari@ucsd.edu

August 2, 2012

Abstract

The University of California requires students with poor writing skills to pass a remedial writing class. The Analytic Writing Placement exam (AWP) is a key determinant of student placement in remedial writing. UCSD's freshman class of 2000 who had to take the AWP and who have English as their first language are equally likely to be assigned to a hard grader, but getting a hard grader significantly increases the likelihood of being placed in remedial writing. Standard errors are large on the IV estimates of the effect of remedial placement on college outcomes and so I focus on two groups where grader toughness strongly affects the likelihood of placement: students with SAT I Writing scores closer to the median and students with lower SAT I Math scores. For both groups, remedial placement significantly reduces the likelihood of graduating from UCSD in six years. For students with writing skills at the margin, remedial placement also increases the probability of dropping out in 5 years and decreases the probability of graduating in 5 years. For students with lower math skills, remedial placement leads to fewer UCSD classes in the first year and though there is no evidence of higher cumulative GPAs, there is some evidence that grades in the required college writing class are higher. Though standard errors are large, I cautiously conclude that there are unlikely to be large positive effects of writing remediation for UCSD students who were required to take remedial writing because they got a tough grader.

*I gratefully thank Kate Antonovics, Julian Betts, Julie Cullen, Gordon Dahl, Nora Gordon, Gordon Hanson and seminar participants at UCSD and UCR for helpful discussions and suggestions. I thank UCSD's former Associate Vice-Chancellor of Undergraduate Education, Mark Appelbaum, and Administrative Analyst Sandy Beattie for their help with the data for this project and George Hanson, UCSD's Basic Writing Director for many useful discussions and providing me with the UC-AWP and Exit exams taken by UCSD's freshmen class of 2000.

1. Introduction

According to the National Center for Education Statistics (NCES, 2010), the six-year graduation rate for freshman entering public four-year colleges and universities in 2000 was only 54.8%. One possible culprit for the low graduation rate is that many colleges and universities admit students they believe are not prepared to take some of the first classes offered by the institution. These students typically must pass a remedial (also called basic, entry level, or developmental) class before they are allowed to take college-level classes in the area(s) where the student is unprepared. In fall 2000, 20% of freshman in public four-year colleges and universities took at least one remedial class in math, writing, or reading (NCES, 2003). A recent analysis of college and university transcripts of students who were 12th graders in 1992 finds that remedial classes in math, reading and writing classes have some of the highest percentages of withdrawals, no-credit repeats, failures and other penalty grades (Adelman, 2004).[1] A key question is whether remedial classes do prepare students to take the institution's classes and so lead to improved educational and labor market outcomes. As states seek to expand access to higher education to a wider range of students, such as through Texas' 1997 "Top 10% Rule" and the University of California's 1999 "Eligibility in a Local Context" and 2011 adoption of holistic review of applications, the effectiveness of remedial classes in higher education will likely be increasingly important.

The key difficulty with determining the effectiveness of remedial programs is that only students with low skills take remedial classes. Differences in graduation rates, time to degree, grades, and other collegiate outcomes across students taking and not taking remedial classes are likely attributable to the lower skills of remediated students and say little about how low skilled students would have done in college in the absence of the remedial class. Several recent papers attempt to resolve this endogeneity issue by using a small randomized trial at the margin of placement (Aiken, West, Schwalm, Carroll, and Hsuing (1998)), instrumental variables (Bettinger and Long (2009)), and regression discontinuity techniques (Butcher, McEwan and Taylor (2009), Calcagno and Long (2008), Leake and Lesik (2007), Lesik (2006, 2007, 2008), Martorell and McFarlin (2011), and Moss and Yeaton (2006)). Bettinger and Long generally find positive effects of both math and English (reading and writing

[1] To increase graduation rates, one of the Department of Education's recommendations is that states not simply fund higher education institutions on enrollments but also base finding on performance indicators such as the number and rate of successful transitions from remedial to college-level coursework and to timely graduation (College Completion Toolkit, March 2011, p.6).

combined) remediation for students at the margin of placement in Ohio. Aiken et. al. find that a remedial writing class did not improve performance on an essay exam but did increase scores on the Test of Standard Written English for Arizona State freshman at the margin of placement. The three large regression discontinuity studies, Butcher, McEwan and Taylor for Wellesley students, Calcagno and Long for Florida community college students and Martorell and McFarlin for Texas students, generally find no effects of remedial classes in quantitative reasoning (Butcher, McEwan and Taylor) or of remedial classes in either reading or in math (both Calcagno and Long and Martorell and McFarlin) on college outcomes. Martorell and McFarlin's is the only study to examine the labor market effects of college remediation, and they find no effect of math or reading remediation on labor market earnings. Note that all of these empirical techniques only allow estimation of the effects of remedial classes for students at the placement margin. Thus, there is no consensus on whether remedial classes improve collegiate performance for students at the margin of placement, there is little evidence on the effects of remedial classes on labor market performance, and there is no evidence of the causal effects of remedial placement for students far below the placement margin.

This paper uses the random assignment of student's essays to graders with different tendencies to fail students to identify the causal effects of writing remediation. This empirical strategy follows that of Kling (2006) and Doyle (2007, 2008). Kling uses the average prison sentence of judges randomly assigned to a criminal case as an instrument to identify the causal effects of incarceration length on the employment and earnings. Doyle uses the fraction of children placed in foster care by randomly assigned caseworkers as an instrument to identify the causal effects of foster care on child outcomes and adult crime. The University of California's Analytical Writing Placement exam (AWP), is a key determinant of whether a University of California (UC) student is required to take remedial writing. The AWP has few grade bins and so just-passing students are quite different from just-failing students, invalidating the regression discontinuity approach to estimating causal effects.[2] However, grader standards likely influence student's essay exam scores. If there are tough and easy graders of the AWP, then, for students whose writing abilities are near the passing threshold, getting a hard grader will place the student into remedial writing classes while getting an easy grader will exempt the student from remedial writing classes. With

[2] Martorell and McFarlin (2011) report being unable to use the regression discontinuity approach to estimate the causal effects of writing remediation due to coarse grade bins for the writing placement exam in Texas.

random assignment of graders to students, it is possible to estimate the causal effects of remedial classes for students whose writing skills are at the margin of placement. As noted by Doyle (2007, 2008) this may be a useful policy margin because, in this context, it informs the UC about whether to encourage AWP graders to score exams more like easy graders or hard graders.

I use data on the University of California, San Diego's (UCSD) freshman class of 2000 to examine the effects of a remedial writing requirement on college outcomes. There is no remediation in either reading or mathematics at the UC, but there is an entry-level writing requirement and just over one quarter (25.6%) of UCSD's freshman class of 2000 were required to take a remedial writing class. I focus on 461 students who had to take the AWP exam, whose first essay exam grader is observed at least 5 times in the data (so I can calculate the grader's failure rate) and who do not have English as a second language (since all ESL students fail the writing exam, grader toughness does not influence ESL student's placement). I cannot reject the hypothesis that getting a hard first grader is unrelated to observed student characteristics, which supports the requirement of random assignment of graders. Further, getting a hard first grader significantly increases the likelihood of remedial placement. However, the first grader's failure rate is a weak instrument in the full sample and so I focus on two groups where grader toughness strongly influences the likelihood of remedial placement. For students with SAT I Verbal scores closer to the median results indicate that remedial placement increases the probability of dropping out in 5 years by 34.3% and decreases the probability of graduating in 5 years by 44.3% and in 6 years by 48.2%. For students with lower SAT I Math scores, remedial placement decreases the probability of graduating in 6 years by 30.8%. Though these point estimates indicate large negative effects of remedial placement, standard errors on the IV estimates are very large for both samples. Looking at the upper bound of the 95% confidence interval, the effect of remedial placement on the likelihood of graduating in six years is -3.2% for students whose verbal skills are close to the median but +5.5% for students with lower math skills. For students whose verbal skills are close to the median, the upper bound of the 95% confidence interval for the effect of remedial writing placement on the five year graduation rate is +7.0% and on the five year dropout rate is -5.7%. I cautiously conclude that there are unlikely to be large positive effects of writing remediation for UCSD students who were required to take remedial writing because they got a tough grader.

As a robustness check, I use the fact that two graders score each student's AWP exam and sometimes these graders disagree. In such cases, a third, more experienced grader determines remedial placement. The potential for ability bias remains among students where the first two graders disagree because third graders place students with lower observed writing ability (and so likely lower unobserved ability) into remedial classes. However, differences in observed ability across students placed and not placed in remediation is much smaller than in the full sample and so I expect ability bias to be smaller among students whose first two graders disagreed on placement. Among students whose first two graders disagreed on placement, those placed in remedial classes by a third grader have lower first year grades in UCSD classes and are less likely to graduate in five years.

Given the IV estimates based on grader toughness and estimates for students where initial graders disagreed on placement, I conclude that, for UCSD students who just barely failed the exam, the UC remedial writing requirement is unlikely to have large positive effects on the academic outcomes. The paper proceeds as follows. Section two discusses the UC Entry Level Writing Requirement and what determines placement in remedial writing. Section three discusses the grader failure rate and the conditions under which it is a valid instrument for taking a remedial writing class. Section four describes the student data and the AWP grader data used in the analysis. Section five presents the empirical results and several robustness checks. Section six concludes and points out directions for future research.

2. The University of California's Entry-Level Writing Requirement (ELWR)

Called the Entry-Level Writing Requirement (ELWR), all UC students must demonstrate a minimal proficiency in English composition.

> *"…Each student must be able to understand and to respond adequately to written material typical of reading assignments in freshman courses. This ability must be demonstrated in student writing that communicates effectively to University faculty"*
>
> Academic Senate Regulation 600

Students may not take the two required university-level writing classes until they fulfill the ELWR. If a UC student does not fulfill the ELWR within her first year, she cannot enroll in any class at the university.[3] In 2000, a high school student could meet the ELWR with a minimum score of 680 on the SAT II-Writing Test, a 3 on either Advanced Placement composition test, a 5 on the International Baccalaureate Higher Level English A exam, or an 8 on the AWP Exam. Prior to matriculation, high school students can also demonstrate writing proficiency by passing with a "C" grade or better a college writing class. Once a student matriculates to a specific UC campus, then she must fulfill the ELWR as specified by the campus. At UCSD, for example, students must pass a remedial writing class with at least a C grade and pass an exit exam similar to the AWP exam.

3. Empirical Approach

For student *i*, let academic outcome, Y_i, be a function of observable student characteristics, X_i, and remedial writing, R_i:

$$(1) \quad Y_i = X_i\beta + \alpha_i R_i + \varepsilon_i$$

One problem for estimating (1) is that R_i is likely correlated with ε_i. Omitted variables, such as taking less rigorous classes in high school, may both increase the likelihood of remediation and lead to lower grades in college, which will bias downward α_i. Doyle (2007, 2008) identifies selection bias, a correlation between α_i and R_i, as a second problem for estimating (1). For example, if AWP graders fail the students they think will benefit the most from taking remedial writing classes, this selection will bias upwards estimates of α_i. To address these issues, let Z_i be a variable that influences placement in remediation but not academic outcomes:

$$(2) \quad R_i = 1 \; if \; (Z_i\gamma + X_i\delta + u_i > 0)$$

[3] At some UCs there is an exception for students who are identified as non-native English speakers. UCSD, for example, places non-native speakers into an English as a Second Language class (ESL). They have three quarters to pass the ESL class and then an additional three quarters to pass the remedial writing class.

A plausible candidate for Z_i is the toughness of the grader randomly assigned to score the exam that determines placement in remedial writing. Following Doyle (2007, 2008), consider the case where there are only two types of graders: easy and hard. The difference in student outcomes across these two graders would estimate the average effect of remedial writing across students where the toughness of their grader determined placement in the remedial class. Let Z_i=1 if the student is graded by a hard grader and Z_i=0 if the student is graded by an easy grader. The Instrumental Variables (IV) approach identifies the local average treatment effect (LATE) as:

$$(3) \qquad a^{LATE} = \frac{E(Y \mid Z=1, X) - E(Y \mid Z=0, X)}{P(R=1 \mid Z=1, X) - P(R=1 \mid Z=0, X)}$$

This approach estimates the effect of remedial writing for students whose writing abilities are so close to the placement threshold that easy and hard graders will disagree on whether the student passed the exam. This approach cannot estimate the effect of remediation far from the placement margin because both hard and easy graders will grade the student the same and so P(R=1|Z=1, X)= P(R=1|Z=0, X)). Doyle (2007, 2008) identifies four conditions that must hold to interpret (3) as a LATE: (i) E(R|Z=z) is a nontrivial function of z (grader toughness is an important determinant of remedial writing placement), (ii) $\gamma > 0$ (getting a tough grader increases the likelihood of remedial writing placement) , (iii) Z and u_i are independent (there is no omitted variable that determines remedial placement and is correlated with getting a tough grader) and (iv) Z and ε_i are independent (grader toughness does not directly influence academic outcomes at UCSD). Since graders of the placement exam do not teach at UCSD, condition (iv) likely holds. I test for conditions (i)-(iii) below.

Following Doyle (2007, 2008), now consider the case where grader toughness is a continuous variable. Doyle points out that estimates will be marginal treatment effects and identifies the marginal treatment effect as:

$$(4) \qquad \alpha^{MTE} = \frac{\partial E(Y \mid (P(R=1) \mid Z=z))}{\partial P(R=1 \mid Z=z)}$$

Thus, I will estimate the effect of the remedial writing requirement on academic outcomes as we move from easier to harder graders.

4. Data

4.1. Student Data

For each of the 3,117 freshman who entered UCSD in 2000, I have six years of data on every UCSD class taken and grade earned. I also know each student's high school GPA, race, sex, parent's income and education, and scores on SAT I, SAT II, AP English tests, and the AWP. Over a quarter of UCSD's freshman did not test out of the ELWR and, of these, 82% took remedial classes while at UCSD, 13% took a remedial class prior to matriculation and 5% did not meet the ELWR.[4] Seven students took a remedial class even though they had fulfilled the writing requirement. Given there are students who took remedial writing when they had fulfilled the ELWR and students who did not take remedial writing when they had not fulfill the ELWR, the estimates in this paper are interpretable as the effect of the remedial writing *requirement* for students at the margin of placement, as we go from lenient to strict graders.

4.2. Analytic Writing Placement Exam (AWP) Scores and Grader IDs

In 2000, students could meet the ELWR with a minimum score of 680 on the SAT II-Writing Test, a 3 on either AP exam in English, a 5 on the International Baccalaureate Higher Level English A exam, or an 8 on the AWP. Apparently, the AWP is the easiest exam because students who fail the AWP exam are unlikely to fulfill the ELWR by passing one of the other exams: 84.2% of the students who just barely fail the AWP exam are required to take remedial writing.[5] Therefore, AWP grader toughness is likely to be an important determinant of being required to take a remedial class.

[4] Most of the students who did not meet the ELWR according to UCSD records have a transfer English class that may have fulfilled the ELWR.

[5] SAT II Writing is the hardest test to pass: only 6.2% of the students who score just below the passing threshold are unable to fulfill the ELWR by passing some other test. Among students who just fail the AP exams, 46.9% and 39.2% for the AP English Literature and Composition and AP English Language and Composition respectively are unable to fulfill the writing requirement with another exam score.

The UC requires students who have not met the ELWR to take the AWP exam in May, prior to matriculation. Students who cannot take the May exam take the AWP exam on campus before the start of classes. Two randomly assigned graders independently read the each student's essay and score it from a low of one to a high of six. The student fulfills the ELWR if the sum of the two graders scores is 8 or more and does not meet the requirement if the sum of the two graders scores is a 6 or less. A third, more experienced grader scores the essay if the student is scored a 4 (just passing) and a 3 (just failing) by the first two graders or if the first two graders' scores are two or more apart. In these cases, the student's grade is twice the third grader's score. Figure 1 illustrates how grader scores map into the student's final grade.

UCSD has an AWP score for 1815 freshman and has the physical exam for 1747 students.[6] On the physical exam is the date the student took the test, each grader's score, and, sometimes, the grader's ID. For the 1,572 students who took the May AWP exam, 682 of the exams have the grader's ID and score for at least one grader. For the 175 students who took the AWP at UCSD, I have all grader scores and IDs. It is for the 857 (682+175) students who have at least one grader ID recorded on their AWP exam that I am potentially able to calculate the grader failure rates described below. In the following section, I explore how the missing May grader IDs may bias the grader failure rate and IV estimates.

5. Results

5.1. First Grader Failure Rate

The instrumental variable is the failure rate of the first grader assigned to score the student's exam.[7] Suppose there are *n* students. Let FR_{ij} be the failure rate of grader *j*, excluding grader *j*'s decision on student *i*'s exam, for student *i:*

[6] It is unclear why there are 68 AWP exams missing. The likelihood an AWP exam is missing is significantly higher if the student is Asian or Hispanic, has a mother who is a college graduate and has a lower high school GPA. However, the likelihood the exam is missing is not significantly related to AWP scores, SAT I Verbal scores, SAT II English scores, fraction female, parent's income, or whether the student is out of state. It is not obvious how the 68 missing exams will bias the estimated effects of remedial placement.

[7] If I had data on all grader IDs, I would use the failure rates of both graders as instruments. However, I observe first grader IDs more frequently than second grader IDs and, with one exception, only observe the second grader's ID if I observe the first grader's ID. Since I am more likely to see both graders IDs when the graders disagreed on placement, there is an induced negative correlation between the first and second grader failure rates. Therefore, I use the failure rate of the first grader as the

$$(5) \qquad FR_{ij} = \frac{\sum_{i=1}^{n} G_j F_{ij} - G_j F_{ij}}{\sum_{i=1}^{n} G_j - 1}$$

Where G_j =1 for grader *j* and 0 otherwise and F_{ij}=1 if grader *j* fails student *i* and is zero otherwise.[8]

While AWP grader assignment is likely to be random, the *recording* of AWP grader IDs may not be random. First, I am more likely to see grader IDs when the grader finds evidence that the student is a non-native English speaker. In these data, graders always fail students whose essays show evidence that English is the student's second language (ESL). Since grader toughness is not a valid instrument for students with ESL, because grader toughness does not influence placement in remedial writing, I exclude the 250 students identified as ESL.

I am more likely to observe grader IDs when one grader passes and the other grader fails the student which creates two issues for estimation.[9] First, this may introduce a positive correlation between grader toughness and student ability even though the pairing of hard and easy graders is random. The key is that hard graders pass high ability students and I will tend to observe hard graders paired with even harder graders who fail the student. Similarly, easy graders fail low ability students and I will tend to observe easy graders paired with even easier graders who pass the student. As shown in Figure 1, this means that the identity of hard graders will available for students with better than average characteristics while the identity of easy graders will be available for students with worse than average characteristics. A positive correlation between grader toughness and student ability will bias the IV estimates towards finding positive effects of remediation. A second issue is that a third grader determines placement when the first two graders disagree. Getting a tough (easy) first grader may simply increase

---

instrumental variable. To reduce sampling error, I base the calculation of the first grader's failure rate on the exams the grader scores as either a first or a second grader on other student's exams. In the robustness checks, I present the results when basing the failure rate just from exams where the grader is a first grader.

[8] The first grader assigned a "3/4" to nine students tested at UCSD and so I cannot determine whether the grader passed or failed the student. Since I do not use the student's own score when calculating her grader's failure rate, I can still use these students in the analysis.

[9] For exams taken in May, runners (people who take bundles of 20 exams from one room of graders to another room of graders) record grader IDs. Runners always record grader IDs on exams that go to a third reader (346 exams) or when graders identify the student as a non-native speaker (170 exams). It is unclear why runners do not record grader IDs. A possible story is that runners record the grader's ID on the top exam in the stack of 20 and whenever an exam is taken out of the stack of 20 (whenever the exam must go to a 3rd grader or the student is identified as ESL). About 30% of the exams go to a 3rd grader or are ESL so this leaves about 14 exams and so I would expect 7% (1/14) of the remaining exams to have ID's under this hypothesis. However, nearly 16% of the exams that do not go to a 3rd grader and the student is a native speaker have an ID.

the chance of getting a third grader who passes (fails) the student, which could violate the monotonicity requirement (condition (ii) $\gamma > 0$, i.e., increases in grader toughness increases the likelihood of remedial placement). I test for both of these possibilities below.

A key question for this study is the minimum number of exams to use to derive to a meaningful estimate of the first grader's failure rate. More exams per grader reduces the sample variance of the failure rate estimate but also excludes many students and graders from the analysis. In my preferred IV specification, I require at least 5 exams per grader which results in a sample 461 students and 81 graders.[10] In this sample, the average number of times I observe the student's first grader (as either a first or a second grader on other student's exam) is 15 with a standard deviation of 17.4. The distribution is highly skewed: for the median student, their first grader is identified on 8 other exams. Table 1 reports sample means for all freshmen with an AWP exam who are not identified as ESL and for the IV subsample (students whose first grader is observed 5 times in the data). Student demographic characteristics, family background, high school GPA, and admissions test scores are quite similar across the two samples of students. However, placement in remedial writing is 6.7 percentage points higher for students in the IV subsample than for all (non-ESL) students who took the AWP exam. A reason for the higher remedial placement is that IV subsample is 21.5 percentage points less likely to take the exam in May and May graders have significantly lower failure rates than UCSD graders. Since students taking the AWP at UCSD are significantly more likely to be men, white, come from higher income families and have parents who are college graduates, the higher failure rates of UCSD graders is likely because they are tougher graders than May graders.

5.2. Are Students and Graders Randomly Paired?

To test the hypothesis that grader failure rates are unrelated to student characteristics, I regress first-grader failure rates on student demographics, family background and test scores.[11] I estimate student-level regressions and cluster the standard errors at the grader level to allow for correlation in errors across students graded by the same

[10] In the robustness checks, I test the sensitivity of the results to requiring more students per grader.

[11] Demographic characteristics are dummy variables for female, Asian, Hispanic, and "other" race. Family background variables are parent's income and a dummy variable for whether mom is a college graduate and a dummy variable for whether dad is a college graduate. Test score data are scores on SAT I Math, SAT I Verbal, SAT II Math, SAT II Writing, AP Literature and Composition, AP Language and Composition, and high school GPA. Missing dummy variables for each variable are also included so that all students could be included in the analysis.

individual.[12] Table 2 presents these results. There are few student characteristics that are significant in the regression: getting a tough grader is significantly positively related to SAT I Verbal scores, missing SAT I scores, and missing SAT II Math scores and getting a tough grader is significantly negatively related to parent's income and SAT II Math scores. The previous section discussed the potential for a positive correlation between student quality and the grader failure rates. However, there is not much evidence of a systematic bias in these data. Furthermore, F-tests do not reject the hypothesis that all regression coefficients are jointly equal to zero. Thus, I cannot reject the hypothesis that grader toughness is unrelated to observed student characteristics in this sample.

5.3. Does Getting a Tough Grader Increase the Likelihood of Remedial Placement?

The correlation between the first grader's failure rate and remedial placement must be strong for instrumental variables estimates to be an improvement over OLS estimates of the effect of remedial writing. Students required to take remedial writing did have significantly harder graders, with graders who were 6% more likely on average to give failing grades, than the graders of students not placed in remedial writing. Epanechnikov kernels in Figure 2 illustrate the distribution of first grader failure rates across students who were required and not required to take remedial writing. A Kolmogorov-Smirnov test rejects the hypothesis that the distribution of grader failure rates is the same across the two groups. I examine the relationship between grader failure rates and remedial placement more formally by estimating the following linear probability model,

$$(6) \qquad R_{ij}=a_0+bX_i+dF_{ij}+e_{ij}$$

Where R_{ij} is 1 if remediation is required for student *i* assigned to grader *j*, F_{ij} is the failure rate for grader *j*, (student *i*'s first AWP grader), X_i is a vector of student *i*'s characteristics and e_{ij} is an error term that is correlated across graders. I cluster the standard errors by grader to address this correlation. Table 3 reports the results of this regression. The estimated effect of the grader's failure rate is significantly positive and the estimated coefficient does not change much (from .273 to .300) depending upon whether I condition on X since, as shown in the previous

[12] A key question is how best to account for grader-level effects in the regressions. In the robustness checks below, I estimate grader-average regressions (so that I have as many observations as graders).

section, the grader's failure rate is unrelated to student characteristics. Getting a grader that is 20% more likely to fail students, about a one standard deviation increase in grader toughness, increases the likelihood a student must take remedial writing by 5.5 to 6 percentage points.

If instruments are weak then the TSLS t-statistic on the endogenous variable in the outcome regressions (here, the effect of remedial placement on academic outcomes) will reject the null hypothesis too often. Stock and Yogo (2005) define weak instruments to be those where the actual size of a 5% hypothesis test is too big for the researcher to tolerate and report critical values for the first stage F-statistic. In the case of one endogenous variable with one instrumental variable, the first stage F-statistic must exceed 16.38, 8.96 and 6.66 respectively to reject the null that a 5% hypothesis test will reject more than 10%, 15%, or 20% of the time. From Table 3, the t-statistic on the grader's failure rate is 2.703 and so the first stage F-statistic is 7.68. I can reject the hypothesis that the actual size of the 5% TSLS t-test in the regressions below exceeds 20% but not that it exceeds 15%. Thus, there is some evidence that the grader's failure rate is a weak instrument for placement in remedial writing.

5.4. Remedial Writing Placement and UCSD Outcomes

5.4.1 Full Sample

The outcomes I examine are the probability that a student drops out of UCSD in 2 or 5 years, the probability that a student graduates in 4, 5 or 6 years, the total number of credit hours (excluding remedial coursework), and the cumulative GPA (excluding remedial coursework) of students who graduate within 6 years. I also examine several first year outcomes including first year GPA (excluding remedial coursework), the total number of UCSD credit hours taken in the first year (excluding remedial coursework), and the total number of class hours taken in the first year (including remedial coursework). By focusing on the first year, I can mitigate the selection effects from students dropping out. Since remedial placement may affect class choice, GPA comparisons reflect both differences in courses taken and course competency. To address this issue, I create a standardized first year GPA, using the grades earned by other freshman from the class of 2000 taking the same class in the same quarter. Unfortunately, I do not know the grades assigned to all students taking a class. However, the student's classmates are likely to be other freshman since I am just looking at classes taken in the first year.

Finally, all students must pass two quarters of college-level writing to graduate from the UC. Remedial writing most directly affects performance in the required college writing classes. Unfortunately, one sixth of the students were graded pass/no pass only in their college writing classes. This is because there are six colleges at UCSD, the college writing classes are unique to each college and one of the colleges graded the writing classes as pass/no pass only. I exclude students from this college from the grades in college writing regressions. I focus on the grade earned the first time the student takes a college-level writing class and so exclude from the analysis 10 students who withdrew from the class the first time they attempted college writing. Thus, I examine the effect of the remedial writing placement on college writing grades conditional on having earned a letter grade in the first college-level writing class.

Table 4 reports OLS and IV estimates of the effect of being required to take a remedial writing class. In all regressions, I cluster the standard errors at the grader level. I condition on test scores (SAT I Math and Verbal, SAT II Math and English, the two AP English Tests), high school GPA, sex, race/ethnicity, parent's income, whether mom is a college graduate, and whether dad is a college graduate. I include missing dummy variables for each independent variable so that all students are included in all regressions. Column 2 of Table 4 reports OLS estimates for both samples. Students required to take remedial classes are 7.0 percentage points more likely to drop out in the first 2 years, are 9.9 percentage points less likely to graduate in 5 years and take 9.6 fewer credits at UCSD. Both first year GPA and the cumulative GPA of graduates are significantly lower for those placed in remedial classes, but the estimated effects of .08 to .10 lower GPAs are very small. Standardized first year GPA is not significantly different across those remediated and not, though the point estimate is negative. Finally, students requiring remediation take significantly fewer UCSD classes in their first year, but there is no significant difference in course load once taking into account remedial coursework. Even after conditioning on an extensive set of observed characteristics, OLS estimates indicate that students placed in remedial writing take longer to graduate and have slightly lower grades than those not placed in remedial classes.

As noted earlier, there may be either omitted variable bias (unobserved differences across students required and not required to take remedial writing which would bias downward OLS estimates) or selection bias (the students most likely to benefit from remedial writing are placed in remedial classes which would bias upwards OLS

estimates). The IV estimates reported in column (3) are the marginal effects of remedial placement as we go from easy to hard AWP exam graders for students near the placement margin. The IV point estimates of the effects of remedial placement are even more negative than the OLS estimates which suggests selection bias dominates omitted variables bias in the OLS estimates.[13] However, the standard errors on the IV estimates are large and the estimated effects are not significantly different from zero nor are they significantly different from the OLS estimates. Given the first stage F-statistic suggests that grader failure rates are a weak instrument, in the next section I look for groups of students where grader toughness strongly affects remedial placement.

5.4.2 Student Groups where Grader Toughness Strongly Affects Remedial Placement.

To reduce the standard errors of the IV estimates, I search for groups of students where the first grader's propensity to fail strongly influences the likelihood of placement in remedial writing. I focus on groups that are predetermined when the student takes the AWP exam so that selecting the sample on these variables will not bias the results but may increase the explanatory power of the first stage regression and so reduce the standard errors of the IV estimates. I examine the likelihood of remedial placement separately for men, women, Asians, and non-Asians. I also examine the likelihood of remedial placement separately for students with ability levels closer to the margin, where I expect the grader's failure rate to be more important in determining remedial placement. I use SAT I Math and Verbal scores as well as high school GPA to define groups of students with ability levels closer to the remedial writing margin. Since it is not clear what scores or grades put students closer to the margin of placement, I searched over every combination in the top and bottom quartiles of the distribution in the IV sample. As shown in Table 5, grader failure rates particularly influence remedial writing placement for men, students whose SAT I verbal scores are between the 10^{th} and 90^{th} percentiles and students whose SAT I Math scores are between 5^{th} and 75^{th} percentiles. For men, I can no longer reject the null hypothesis that there is no relationship between grader failure rates and student demographic characteristics and results are generally consistent with the hypothesis that

[13] Doyle (2007, 2008) uses caseworker propensities to place children in foster care as an instrument. He also finds IV estimates are substantially more negative than OLS estimates which suggests that selection bias swamps omitted variables bias. Kling (2006) uses judge toughness as an IV for the effects of incarceration length on employment and earnings. Kling finds OLS estimates are more negative than the IV estimates which suggests omitted variable bias dominates selection bias.

hard graders score better students.[14] I re-estimate the OLS and IV effects of remedial placement for students whose verbal or math skills put them at the margin of placement, where grader failure rates strongly influence the likelihood of taking a remedial class, and report results in Table 6.

Comparing across Tables 4 and 6, I find little change in the OLS point estimates though standard errors are larger in the subsamples. In contrast, the IV standard errors are smaller (though still large) when estimating the effects of remedial placement for students closer to the placement margin. As shown in Table 6, students placed in remedial writing because they got a tough grader are significantly (at the 10% level) less likely to graduate in six years. The standard errors are large so the 95% confidence interval for the change in the likelihood of graduating in 6 years ranges from -3.2% to -93.2% for students whose verbal skills are close to the placement margin and from +5.5% to -61.1% for students whose math skills are close to the placement margin. In addition, for students whose verbal skills are close to the placement margin, remedial writing placement significantly (at the 10% level) reduces the five year graduation rate (95% confidence interval of 7.0% to -89.0%) and increases the five year dropout rate (95% confidence interval of -5.7% to 74.3%).

There is also some evidence in Table 6 that first year grades (excluding any remedial coursework) improve with placement in remedial writing but there is no evidence that the cumulative GPA of graduates is higher. The IV point estimates of the effect of remedial placement on first year GPA, standardized first year GPA and grades in the first college-level writing class are large and positive. However, only for students with math skills at the placement margin is the effect of remedial placement on grades in the first required college writing class marginally significant at the 11% level (95% confidence interval of -.138 to 1.362). In contrast, the IV estimates of the effect of remedial placement on the cumulative GPA of graduates are small, negative and insignificant. One hypothesis for these different results is that grade gains are concentrated among students who did not graduate in 6 years. Depending on the sample, 17.6% to 18.0% of the students did not graduate in 6 years. To test this hypothesis, I re-estimate the first year grade regressions for graduates only and present the results in Table 6. Point estimates for the effect of remedial writing placement are much smaller and insignificant in the first-year GPA, standardized first

[14] For men, there is a significant positive correlation between grader failure rates and SAT I Verbal scores, having a father who is a college graduate, high school GPA, being Asian and not reporting race. One inconsistency with the hypothesis of a positive correlation between grader toughness and student quality is that there is a significant negative relationship between AP English Literature & Composition scores and grader failure rates for men.

year GPA and grades in the first college level writing class. These results are consistent with the hypothesis that grade gains from remedial placement primarily accrue to students who do not graduate from UCSD in six years.

5.5. Robustness Checks

5.5.1. Robustness Check 1: Results Averaged by Grader

Since the first grader's failure rate only varies at the grader-level, a key empirical question is how to account for grader-level effects in the regressions. To determine how sensitive the previous results are to clustering, I estimate grader-level regressions using weighted (to account for the different numbers of students over which the average is taken) OLS and IV with robust standard errors. Note I can still estimate the first stage for the IV because the fraction of the grader's students placed in remedial writing (the dependent variable) differs from the grader's failure rate (the independent variable) both because students use other exams to place out of remedial writing (such as AP exams) and because a third AWP grader may determine placement. I exclude 39 students who do not report all of the five key continuous ability measures (SAT I Verbal, SAT I Math , SAT II Writing, SAT II Math or high school GPA). This leaves 79 graders for the grader-level regressions.[15] Appendix A reports the results averaging across all students the grader scored as well as averaging across students whose placement is most impacted by grader failure rates (students whose SAT I Verbal scores are between the 10^{th} and 90^{th} percentiles and students whose SAT I Math scores are between the 5^{th} and 75^{th} percentiles)

Point estimates are similar across the grader-average regressions reported in Appendix A and the student-level regressions reported in Tables 5 and 6. The main difference is that OLS estimates are less likely to be significant in the grader-average regressions. The only significant OLS estimate is that students take fewer college-level credits (so excluding remedial coursework), about equal to one course, in their first year. The grader-average IV regressions are similar to the student-level IV regressions. Getting a tough grader significantly (at the 10% significance level) reduces the number of UCSD units taken in the first year (95% confidence interval of -.552 to .020) and reduces six-year completion rates (95% confidence interval of -6.899 to .125 units) for the average

[15] Regressors include SAT I Math and Verbal, SAT II Math and Verbal, HS GPA, and dummy variables for female, Asian, Hispanic, other race, missing race, mom's a college grad, dad's a college grad, missing mom's education and missing dad's education. In these average regressions, I did not condition on three continuous variables that were frequently missing in the data: AP Language and Composition, AP Literature and Composition and parent's income.

student. Remedial placement has the largest negative effects on students whose writing abilities are closer to the placement margin. Getting a tough grader significantly reduces total college-level credits (95% confidence interval of -6.899 to .125 units) and reduces five- and six-year completion rates (95% confidence interval of -.610 to .032 and -.637 to -.020 respectively) while five-year drop out probabilities are higher (95% confidence interval of -.020 to .586). Again, IV estimates are more negative than OLS estimates in these grader-average regressions which again suggests that selection bias (graders place into remediation the students they expect will benefit the most) exceeds omitted variable bias (students with unobserved lower skills are remediated). Therefore, the main results are robust to the method I use to account for grader-level effects.

5.5.2. Robustness Check 2: Requiring More Exams Per Grader

One reason that the first grader's failure rate may be a weak instrument is that I determine some graders' failure rates on only four exams (while I require 5 exams per grader, I do not use the student's own exam when calculating their grader's failure rate). Requiring more exams per grader reduces the sampling error of the failure rate but also reduces the number of students and graders I can use to estimate the effects of remedial placement on collegiate performance. I explore these tradeoffs by requiring at least 6 and at least 7 exams per grader. I cannot reject the hypothesis that there is a no relationship between the student's demographics and the grader's failure rate for students whose math or verbal scores put them closer to the remedial placement threshold so I present results for these groups in Appendix B.

Comparing Table 6 with Appendix B, the estimated effects of remedial placement tend to be smaller and are more likely to be insignificant when requiring more exams per grader. A key exception is that placement in remedial writing significantly increases first year grades, standardized first year grades, and college writing grades when requiring at least 7 exams per grader for students whose math scores are closer to the placement margin. There is no evidence that the cumulative GPA of graduates is higher, however. Since the samples are different (44 of the 240 students have not graduated), I re-estimate the first-year grade regressions for graduates only. In results not presented here, but available by request, point estimates on the three first year grade variables are positive but

much smaller and not significant. These results are consistent with the hypothesis that grade gains from remedial placement primarily accrue to students who do not graduate from UCSD in six years.

5.5.3. Robustness Check 3: Grader Failure Rates Based On Exams Scored as a First Grader Only

There are at least two graders for each student's exam so one student's first grader can be the second grader on another student's exam. In the previous analysis, I based the first grader's failure rate on exams that the grader scored as either a first or a second grader. Since I am more likely to observe the second grader's ID when one grader passes and the other grader fails the student, it is possible that using exams scored as a second grader introduces a bias. There are 38 graders observed 5 or more times as a first grader and it is for this sample that I explore the sensitivity of the results to estimating the first grader's failure rate using exams scored as a first grader only. In Appendix C, I present the IV results when basing the first grader's failure rate on exams scored as a first grader only and, for the same sample of graders and students, when basing the first grader's failure rate on exams scored as either a first or a second grader.

Either way I calculate the grader's failure rate, I typically reject the hypothesis that there is no significant relationship between student demographics and the grader's failure rate: tougher graders are significantly more likely to grade women, Asians, students from lower family incomes and students with higher SAT I Verbal scores. Nevertheless, IV results are fairly similar across the two specifications of the failure rate and consistent with the previous analysis.

5.5.4. Robustness Check 4: Effect of Remedial Placement on Students Whose AWP Graders Disagreed

As a check on the estimates using grader failing tendencies, I examine the effect of being required to take a remedial class among students whose writing abilities are likely to be very similar: students whose first two graders disagreed on placement. In this situation, a third, more experienced, grader decides the student's placement in remedial writing. The potential for omitted ability bias remains: the third grader is more experienced and so remedial placement is likely still correlated with student ability. Nevertheless, ability bias is less likely to be an issue for these students because ability differences across those placed and not placed in remediation are likely to

be much smaller. Again, I will be estimating a local average treatment effect (LATE) for students at the margin of placement because I estimate the effect only for students whose essays were so good that at least one grader thought the student should pass. This empirical approach does not address selection bias, i.e., that students most likely to be helped by remedial coursework are those required to take remedial classes.

Among the students who took the AWP and were not identified as ESL, there are 248 students who earned a 3 (a just failing grade) from one grader and a 4 (a just passing grade) from the other grader. As shown in Columns (1)–(3) of Table 7, students not required to take remedial writing in the full sample have many observed characteristics associated with better collegiate performance: significantly higher SAT I, SAT II and AP test scores, higher high school GPAs, greater likelihood of having parents who graduated from college, and come from families with higher incomes. Columns (4)-(6) in Table 7 compare students required and not required to take a remedial class in the sample of students whose first two graders disagreed on placement in remedial writing. As expected, differences in observed characteristics across those required and not required to take remedial writing are smaller than in the full sample. For example, compared to students not placed in remedial classes, the average SAT I Writing scores of remediated students are 57 points in the full sample but are 37 points lower in the sample where graders disagreed. The hope is that *unobserved* ability differences are also smaller and so there is less downward bias on the estimated effect of remedial placement when graders disagree on placement.

Table 8 reports the OLS estimates of the effect of being required to take remedial writing on performance at UCSD on the full sample (columns 2 and 3) and across students whose AWP graders disagreed on placement (columns 4 and 5). Among all AWP test takers not identified as ESL, remedial placement is associated with uniformly negative effects on student outcomes though conditioning on observed student characteristics (comparing columns 2 and 3) reduces the estimated negative effects. For students whose graders disagreed on placement, I estimate that remedial placement significantly reduces the likelihood of graduating in 5 years (95% confidence interval of -.008 to -.244), reduces first year GPA (95% confidence interval of -.028 to -.314 grade points) and reduces standardized first year GPA (95% confidence interval of -.019 to -.363 grade points). Surprisingly, point estimates in the sample of students whose graders disagreed on placement (column 5) are frequently *more* negative than in the full sample (column 3). This result is surprising since I expected the effects of remedial placement to be

less negative in the sample where graders disagreed because of reduced downward bias due to omitted variables. However, the point estimates for the sample of students whose graders disagreed are not significantly different from the point estimates for the full sample. In future research, it would be very useful to have larger sample sizes to reduce the standard errors of these estimates.

6. Conclusions

There has long been interest in determining whether costly remedial classes improve performance in college and the labor market. Since assignment to remedial classes is typically based on failing a placement exam, regression discontinuity (RD) has been used to estimating causal effects. However, RD is unlikely to be a valid technique for estimating the causal effects of remedial writing classes because there are typically few grade bins so that students just passing are quite different than students just failing the exam. This paper proposes the random assignment of hard and easy exam graders to identify the causal effects for students at the margin of placement.

A UC placement exam, the AWP, largely determines placement in remedial writing. Using data from UCSD's freshman class of 2000, I find that students are equally likely to be assigned a tough AWP grader but getting a tough AWP grader significantly increases the likelihood of placement in remedial writing. While the grader's failure rate is a weak instrument in the full sample, grader toughness strongly influences remedial placement for two subsamples: students with SAT I Verbal scores closer to the median (between the 90^{th} and 10^{th} percentiles) and students with lower SAT I Math scores (between the 75^{th} and 5^{th} percentiles). In both subsamples, students placed in remedial writing because they got a tough AWP grader are significantly less likely to graduate in six years. There is some evidence that remedial writing placement does not just delay graduation but increases the likelihood of dropping out of UCSD within five years: the point estimates are positive for both subgroups, but only statistically significant in the trimmed SAT I Verbal group. There is some evidence of positive effects of remedial placement on first year grades and, in particular, grades in the first college writing class. However, there is no evidence that remedial placement results in higher cumulative GPAs of graduates. Results suggest that first year grade gains may be concentrated among students who do not graduate in 6 years. Though standard errors are large,

I cautiously conclude that it is unlikely that there are large positive effects to requiring a remedial class for UCSD freshman with writing skills close to the placement margin.

UCSD spends around $450,000 annually on the remedial writing program. The results of this study indicate that students who are required to take remedial writing because they received a tough AWP exam grader receive few benefits from these dollars. This suggests that UCSD could lower the placement threshold, or encourage tough AWP graders to grade more like easy AWP graders, with little impact for students at the margin of placement. Alternatively, UCSD could change the remedial program so that it is more effective for students at the margin of placement. A key limitation of the empirical technique used in this study is that I could not estimate the causal effects of remedial placement for students with weaker writing skills or with English as their second language. A randomized trial that randomly excuses students from the remedial writing requirement would identify the causal effects of remedial writing for all students placed in remedial writing.

Bibliography

Adelman, C. (2004). "Principal Indicators of Student Academic Histories in Postsecondary Education, 1972-2000." Washington, DC: U.S. Department of Education, Institute of Education Sciences

Aiken, Leona S., West, Stephen G. Schwalm, David E., Carroll, James L. and Hsuing, Shenghwa. (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.

Baum, C.F., Schaffer, M.E., Stillman, S. (2007). ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression. http://ideas.repec.org/c/boc/bocode/s425401.html

Bettinger, Eric P. and Long, Bridget T. (2007). "Institutional Responses to Reduce Inequalities in College Outcomes: Remedial and Developmental Courses in Higher Education," *Economic Inequality and Higher Education: Access, Persistence and Success*. Stacy Dickert-Conlon and Ross Rubenstein, eds. New York: Russell Sage Foundation Press.

Bettinger, Eric P. and Long, Bridget T. (2009). "Addressing the Needs of Under-prepared College Students: Does College Remediation Work?" *Journal of Human Resources*, 44(3):736-771.

Butcher, Kristin F., McEwan, Patrick J. and Taylor, Corrine H. (2010)."The Effects of Quantitative Skills Training on College Outcomes and Peers," *Economics of Education Review*, 29(2): 187-199.

Calcagno, Juan Carlos and Long, Bridget T. (2008). "The Impact of Postsecondary Remediation Using a Regression Discontinuity Approach: Addressing Endogenous Sorting and Noncompliance," Working Paper 14194, National Bureau of Economic Research.

Doyle, Joseph J. (2007). "Child Protection and Child Outcomes: Measuring the Effects of Foster Care." *American Economic Review*, 97(5): 1583-1610.

Doyle, Joseph J. (2008). "Child Protection and Adult Crime: Using Investigator Assignment to Estimate Causal Effects of Foster Care" *Journal of Political Economy*, 116(4): 746-770

Kling, Jeffrey R. (2006). "Incarceration Length, Employment, and Earnings." *American Economic Review* 96(3): 863-876.

Leake, M. and Lesik, Sally. (2007). " Do Remedial English Programs Impact First Year Success in College? An Illustration of the Regression Discontinuity Design," *International Journal of Research and Method in Education*, Vol 30. No. 1, pp. 89-99.

Lesik, Sally. (2006). "Applying the Regression Discontinuity Design to Infer Causality with Non-Random Assignment." *The Review of Higher Education*, Vol 30. No. 1, pp. 1-19.

_____. (2007). "Do Developmental Mathematics Programs have a Causal Impact on Student Retention? An Application of Discrete-Time Survival and Regression Discontinuity Analysis" *Research in Higher Education*, Vol. 48, No 5, pp. 483-608.

_____. (2008). "Studying the Effectiveness of Programs and Initiatives in Higher Education Using the Regression-Discontinuity Design" J.C. Smart (eds.) Higher Education: Handbook of Theory and Research, Springer Netherlands, pp. 277-297.

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

Moss, Brian G. and Yeaton, William H. (2006). "Shaping Policies Related to Developmental Education: An Evaluation using the Regression-Discontinuity Design." *Educational Evaluation and Policy Analysis*, Vol. 28, No. 3, pp. 215-229.

Stock, J.H. and Yogo, M. (2005). "Testing for Weak Instruments in Linear IV Regression." In D.W.K. Andrews and J.H. Stock (eds.) Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. Cambridge: Cambridge University Press, pp. 80–108.

U.S. Department of Education, National Center for Education Statistics. *Remedial Education at Degree-Granting Post Secondary Institutions in Fall 2000*. NCES 2004-010, by Basmat Parsad and Laurie Lewis. Project officer: Bernard Greene. Washington D.C., 2003.

U.S. Department of Education, National Center for Education Statistics, *Digest of Education Statistics 2009.* NCES 2010-013, by T.D. Snyder and S.A. Dillow, Washington, DC., 2010.

U.S. Department of Education, College Completion Tool Kit, Washington, D.C., 2011.

Table 1: Sample Means for Full Sample and for the Subsample Whose Grader is Identified on at Least 5 Exams

	AWP Test Takers Not Identified as ESL Mean (St. Error)		AWP Test Takers Not Identified as ESL: Student's Grader on 5+ exams Mean (St. Error)	
Remediation Required	.319	(.466)	.386	(.487)
Take UCSD Remedial Class	.293	(.455)	.373	(.484)
Take Other Remedial Class	.025	(.155)	.020	(.139)
Took AWP in May	.903	(.296)	.688	(.464)
Did not take SAT I exams	.028	(.165)	.035	(.183)
SAT I Verbal Score (takers)	580.790	(70.243)	582.331	(68.161)
SAT I Math Score (takers)	639.945	(75.325)	644.629	(74.611)
Did not take SAT II-Writing	.066	(.249)	.061	(.239)
SAT II Writing Score (takers)	575.630	(63.095)	574.781	(62.242)
Did not take SAT II-Math	.069	(.254)	.070	(.255)
SAT II-Math Score (takers)	628.758	(79.967)	636.775	(81.723)
Did not take AP: Lang and Comp	.862	(.391)	.855	(.353)
AP Lang & Comp Score (takers)	2.422	(.809)	2.239	(.698)
Did not take AP: Lit and Comp	.622	(.485)	.678	(.453)
AP Lit & Comp Score (takers)	2.880	(.862)	2.709	(.793)
Missing High School GPA	.005	(.073)	.009	(.093)
High School GPA (reporters)	3.985	(.237)	3.973	(.241)
Fraction Female	.520	(.500)	.508	(.500)
Missing Race	.087	(.282)	.091	(.288)
Faction Asian (reporters)	.407	(.492)	.372	(.484)
Fraction Hispanic (reporters)	.148	(.355)	.134	(.341)
Fraction Other (reporters)	.035	(.174)	.048	(.213)
Fraction White (reporters)	.410	(.492)	.446	(.450)
Missing Mom's Ed	.043	(.203)	.060	(.235)
Fraction Mom Coll Grad (reporters)	.462	(.499)	.493	(.501)
Missing Dad's Ed	.069	(.253)	.065	(.247)
Fraction Dad Coll Grad (reporters)	.552	(.498)	.594	(.492)
Missing Family Inc.	.194	(.395)	.187	(.390)
Family Income (reporters)	73,705	(69,526)	80,385	(77,703)
Sample Size	1,497		461	

Table 2: Are AWP Graders Randomly Observed? First Grader Failure Rates and Student Characteristics

Dependent Variable: Failure Rate of Student's First AWP Grader

OLS regressions with standard errors clustered at the grader level.
(Grader Observed on 5 or more Exams)

	Coefficient (Standard Error)	
SAT I: Math/1000	0.252	(0.287)
SAT I: Verbal/1000	0.474**	(0.184)
SAT II: Writing/1000	0.023	(0.223)
SAT II: Math/1000	-0.446*	(0.255)
High School GPA	0.045	(0.044)
AP Lit & Comp	-0.006	(0.029)
AP Lang & Comp	-0.015	(0.045)
Female	0.029	(0.025)
Asian	0.026	(0.024)
Hispanic	-0.014	(0.043)
Other Race/Ethnic	0.043	(0.080)
Mom College Grad	0.006	(0.034)
Dad College Grad	0.031	(0.028)
Parent's income/1m	-0.301*	(0.161)
Missing SAT I	0.491**	(0.227)
Missing SAT II Writing	0.043	(0.147)
Missing SAT II Math	-0.304*	(0.174)
Missing HS GPA	0.167	(0.242)
Missing AP Lit	-0.100	(0.112)
Missing AP Lang	0.042	(0.075)
Missing Race/Ethnicity	-0.038	(0.035)
Missing Mom Ed	-0.058	(0.047)
Missing Dad Ed	0.028	(0.045)
Missing Parent's Income	-0.032	(0.034)
Constant	0.139	(0.265)
R-Square	0.059	
F-test for Regression	F(24, 80)=1.35	
Prob > F	0.162	
Number of Students	461	
Number of Graders	81	
Mean of Dependent Var.	0.436	
Stand. Dev. Dep. Var.	0.229	

Notes: *=10% significance and **=5% significance.

Table 3: Remedial Writing Placement and Grader Failing Tendencies

Dependent Variable: Dichotomous variable that equals 1 if remediation is required.

Linear probability model with standard errors clustered at the grader level.
(Grader Observed on 5 or more Exams)

	Coefficient (Standard Error)		Coefficient (Standard Error)	
Failure Rate	0.273**	(0.12)	0.300***	(0.11)
SAT I: Math/1000			-0.193	(0.49)
SAT I: Verbal/1000			-1.212***	(0.37)
SAT II: English/1000			-1.648***	(0.32)
SAT II: Math/1000			0.316	(0.49)
AP English Lit & Comp			-0.104**	(0.05)
AP English Lang & Comp			-0.146***	(0.05)
High School GPA			-0.077	(0.11)
Female			0.045	(0.04)
Asian			-0.036	(0.05)
Hispanic			0.060	(0.08)
Other race/ethnic			-0.094	(0.10)
Mom College Grad			-0.034	(0.04)
Dad College Grad			0.014	(0.04)
Parent's income/1m			0.358	(0.34)
Missing SAT I			-0.854**	(0.39)
Missing SAT II			-0.917***	(0.23)
Missing SAT II Math			0.290	(0.33)
No AP EC exam score			-0.010	(0.15)
No AP EL exam score			-0.349**	(0.14)
Missing HS GPA			-0.631	(0.43)
Missing Race/Ethnicity			-0.133*	(0.07)
Missing Mom’s Ed			-0.219*	(0.13)
Missing Dad’s Ed			0.187	(0.11)
Missing Parent's Income			0.145**	(0.07)
Constant	0.267***	(0.05)	2.531***	(0.47)
R-square	0.016		0.271	
Observations	461		461	

Notes: *=10% significance, **=5% significance, and ***=1% significance.

Table 4: The Remedial Writing Requirement and UCSD performance: OLS and IV estimates

Grader Identified on 5 or more exams (81 graders)

Dependent variable	Dependent Variable Mean (St. Dev.) (1)	OLS Remediation Required Coef. (St. Error) (2)	IV Remediation Required Coef. (St. Error) (3)	First-Stage F-statistic (4)	Sample Size (5)
Dropped Out in 2 years	.085 (.279)	.070** (.034)	.224 (.202)	7.68	461
Dropped Out in 5 years	.156 (.363)	.057 (.042)	.410 (.278)	7.68	461
Earned BA or BS in 4 years	.492 (.500)	-.071 (.062)	-.298 (.320)	7.68	461
Earned BA or BS in 5 years	.783 (.413)	-.099** (.047)	-.323 (.257)	7.68	461
Earned BA or BS in 6 years	.818 (.386)	-.064 (.044)	-.480 (.309)	7.68	461
Total Credit Hours: Excluding Remedial	180.527 (49.412)	-9.582* (5.177)	-64.655 (41.567)	7.68	461
Year 1 Credit Hours Excluding Remedial	44.383 (5.564)	-4.351** (.518)	-3.925 (3.045)	7.68	461
Year 1 Credit Hours UCSD+Remedial Hours	46.118	-.019 (.539)	-.617 (3.232)	7.68	461
First Year GPA Excluding Remedial	2.958 (.548)	-.077** (.057)	.121 (.346)	7.68	461
Standardized 1st Yr GPA Excluding Remedial	-.098 (.657)	-.099 (.068)	.219 (.389)	7.68	461
GPA in First College Writing class	3.016 (.472)	-.057 (.051)	.608 (.618)	2.95	319
Cum GPA of Students Graduating in 6 yrs	3.157 (.375)	-.064* (.038)	-.067 (.248)	9.35	377

Notes: Each element of columns (2) and (3) represents a separate OLS or IV regression with the "Remediation Required" coefficient reported. I cluster standard errors at the grader level. Other regressors are SAT I-Math and Verbal, SAT II-Math and Writing, high school GPA, AP Lit & Comp and Lang & Comp, parent's income, dummy variables for female, Asian, Hispanic, mom is a college grad, dad is a college grad, and missing dummy variables for each variable. There are fewer students in the "GPA in First College Writing Class" regression because I exclude students from one of UCSD's six colleges that graded the college writing class pass/no pass only.

*=10% significance, **=5% significance

Table 5: Identifying Student Groups Where the Relationship between the Grader's Failing Tendency and Placement in Remedial Writing is Particularly Strong.

Dependent Variable: Dichotomous variable that equals 1 if remediation is required.

Linear probability model with standard errors clustered at the grader level.
(Graders observed on at least 5 exams)

	# Students	# Graders	Failure Rate Coefficient (Standard Error)	T-statistic
Women	217	73	.280 (.142)	1.97
Men	205	67	.515 (.139)	3.71
Asians	146	67	.173 (.149)	1.16
Non-Asians	276	73	.476 (.152)	3.13
High School GPA between 10^{th} and 95^{th} percentile	374	78	.384 (.125)	3.08
SAT I Verbal between 10^{th} and 90^{th} percentile	331	79	.432 (.112)	3.85
SAT I Math between 5^{th} and 75^{th} percentile	306	77	.497 (.115)	4.30
SAT II Writing between 5^{th} and 100^{th} percentile	388	79	.365 (.116)	3.14
SAT II Math between 5^{th} and 75^{th} percentile	336	77	.455 (.115)	3.96

Notes: When creating these subgroups, I start with 422 students that report all SAT scores and high school GPA. To identify SAT I and high school GPA groups where the grader's failure rate strongly influences placement, I search over every combination (by 5) in the top and bottom quartiles e.g., 95-5, 95-10, 95-15, 95-20, 95-25, 90-5, 90-10, etc. I report the SAT I, SAT II and High School GPA groups with the highest t-statistic on the Grader Failure Rate where I am unable to reject the hypothesis of no significant relationship between the grader's failure rate and student characteristics.

Each failure rate coefficient is from a separate regression that conditions on SAT I-Math, SAT I-Verbal, SAT II Writing, SAT II-Math, high school GPA, AP Lit & Comp, AP Lang & Comp, parent's income, dummy variables for female (except in the male/female groups), Asian (except in the Asian/non-Asian groups), Hispanic, other race/ethnicity, mom is a college grad, dad is a college grad, and dummy variables for not reporting mom's education, dad's education, and parent's income.

Table 6: The Remedial Writing Requirement and UCSD performance for Students with Verbal or Math Skills at the Placement Margin

	Students with SAT I Verbal Between 10th and 90th percentile				Students with SAT I Math Between 5th and 75th percentiles			
Dependent variable	Dependent Variable Mean (SD) (1)	Remediation Required Coef. OLS (2)	Remediation Required Coef. IV (3)	Sample Size (4)	Dependent Variable Mean (SD) (5)	Remediation Required Coef. OLS (6)	Remediation Required Coef. IV (7)	Sample Size (8)
Dropped Out in 2 years	.066 (.249)	.070** (.038)	.175 (.165)	331	.095 (.293)	.100** (.040)	.004 (.159)	306
Dropped Out in 5 years	.139 (.346)	.035 (.050)	.343* (.200)	331	.163 (.370)	.070 (.045)	.186 (.161)	306
Earned BA or BS in 4 years	.520 (.500)	-.023 (.073)	-.197 (.266)	331	.484 (.501)	-.049 (.079)	-.272 (.220)	306
Earned BA or BS in 5 years	.807 (.396)	-.064 (.055)	-.443* (.237)	331	.788 (.410)	-.103** (.052)	-.229 (.194)	306
Earned BA or BS in 6 years	.834 (.373)	-.036 (.051)	-.482** (.225)	331	.820 (.385)	-.076* (.044)	-.308* (.185)	306
Total Credit Hours Excluding Remedial	179.929 (47.220)	-14.692** (6.628)	-43.527 (30.835)	331	175.485 (49.470)	-16.005** (6.680)	-32.883 (23.774)	306
Year 1 Credit Hours: Excluding Remedial	44.707 (5.532)	-4.027** (.646)	-3.644 (2.501)	331	44.227 (5.4467)	-3.893** (.704)	-4.379** (2.132)	306
Year 1 Credit Hours: UCSD+Remedial	46.230 (5.013)	.098 (.650)	-.119 (2.569)	331	46.070 (4.878)	.663 (.691)	-.248 (2.096)	306
Year 1 GPA Excluding Remedial	2.999 (.521)	-.060 (.075)	.234 (.323)	331	2.935 (.546)	-.073 (.084)	.309 (.240)	306
Standardized Year 1 GPA Excluding Remedial	-.048 (.632)	-.098 (.093)	.221 (.380)	331	-.128 (.662)	-.114 (.103)	.327 (.280)	306
GPA in 1st College Writing class	2.960 (.529)	-.085 (.081)	.538 (.478)	234	2.941 (.565)	-.050 (.086)	.612 (.383)	221

Notes: Each element of columns 2, 3, 6 and 7 reports the coefficient and standard error of the "Remediation Required" variable from separate OLS and IV regressions. Each regression conditions on SAT I Math, SAT I Verbal, SAT II Writing, SAT II Math, AP Literature and Composition, AP Language & Composition, high school GPA, parent's income, dummy variables for female, Asian, Hispanic, whether mom is a college grad, whether dad is a college grad, and missing dummies for each variable. In the regressions for grades in UCSD's 1st writing class, I exclude students from one of UCSD's 6 colleges that graded the class pass/no pass only. Finally, *=10% significance, **=5% significance.

Table 6 (continued): The Remedial Writing Requirement and UCSD performance for Students with Verbal or Math Skills at the Placement Margin.

RESULTS FOR STUDENTS WHO GRADUATED IN 6 YEARS:

	Students with SAT I Verbal Between 10th and 90th percentiles				Students with SAT I Math Between 5th and 75th percentiles			
Dependent variable	Dependent Variable Mean (SD) (1)	Remediation Required Coef. OLS (2)	Remediation Required Coef. IV (3)	Sample (4)	Dependent Variable Mean (SD) (5)	Remediation Required Coef. OLS (6)	Remediation Required Coef. IV (7)	Sample (8)
Cumulative GPA at Graduation	3.183 (.369)	-.045 (.048)	-.076 (.218)	276	3.132 (.375)	-.007 (.047)	-.019 (.165)	251
Year 1 GPA Excluding Remedial	3.074 (.434)	-.051 (.066)	-.033 (.275)	276	3.017 (.440)	-.062 (.065)	.107 (.200)	251
Standardized Year 1 GPA Excluding Remedial	.039 (.519)	-.104 (.080)	-.062 (.331)	276	-.032 (.532)	-.101 (.083)	.041 (.250)	251
GPA in 1st College Writing class	3.005 (.468)	-.091 (.071)	.114 (.216)	195	2.990 (.490)	-.066 (.071)	.200 (.207)	185

Notes: Each element of columns 2, 3, 6 and 7 reports the coefficient and standard error of the "Remediation Required" variable from separate OLS and IV regressions. Each regression conditions on SAT I Math, SAT I Verbal, SAT II Writing, SAT II Math, AP Literature and Composition, AP Language and Composition, high school GPA, parent's income, dummy variables for female, Asian, Hispanic, whether mom is a college grad, whether dad is a college grad, and missing dummies for each variable. In the regressions for grades in UCSD's 1st college writing class, I exclude students from one of UCSD's 6 colleges that graded the writing class pass/no pass only. Finally, *=10% significance, **=5% significance.

Table 7: Differences in Means across Students Required and Not Required to take Remedial Writing: Across the Full Sample and Across Students Whose AWP Graders Disagreed on Placement

	AWP Test-Takers Not Identified as ESL (n=1497)			AWP Test-Takers Not Identified as ESL Whose Graders Disagreed (n=248)		
Variable	Remediation NOT Required (St.Dev.)	Remediation Required (St.Dev.)	Difference (St.Error)	Remediation NOT Required (St.Dev.)	Remediation Required (St.Dev.)	Difference (St.Error)
Missing SAT I scores	.026 (.158)	.034 (.181)	-.008 (.009)	.025 (.157)	.039 (.195)	-.014 (.023)
SAT I Verbal (non-missing)	598.570 (62.259)	541.564 (73.260)	57.001** (3.658)	589.231 (64.810)	551.626 (62.678)	37.605** (8.230)
SAT I Math (non-missing)	646.294 (71.864)	627.305 (81.316)	18.990** (4.147)	635.641 (76.928)	632.114 (85.270)	3.527 (10.501)
Missing SAT II-Writing	.060 (.237)	.078 (.268)	-.018 (.014)	.025 (.157)	.070 (.257)	-.045* (.027)
SAT II Writing (non-missing)	591.148 (55.665)	540.862 (65.200)	50.286** (3.334)	578.120 (49.514)	551.345 (61.354)	26.775** (7.265)
Missing SAT II-Math	.065 (.246)	.078 (.268)	-.013 (.014)	.033 (.180)	.086 (.281)	-.053* (.030)
SAT II-Math (non-missing)	632.959 (77.848)	621.056 (84.554)	11.903** (4.535)	628.103 (88.523)	624.786 (86.119)	3.317 (11.442)
Missing AP (36): Lang and Comp	.854 (.354)	.883 (.322)	-.029 (.019)	.817 (.389)	.836 (.372)	-.019 (.048)
AP Lang & Comp (nonmissing)	2.638 (.832)	1.864 (.345)	.773** (.112)	2.409 (.666)	1.952 (.218)	.457** (.153)
Missing AP (37): Lit and Comp	.533 (.499)	.809 (.393)	-.276** (.025)	.500 (.502)	.813 (.392)	-.313** (.057)
AP Lit & Comp (nonmissing)	3.071 (.794)	1.865 (.344)	1.207** (.083)	2.933 (.634)	1.917 (.282)	1.017** (.135)
Missing HS GPA	.008 (.088)	0 (0)	.008** (.004)	.008 (.091)	0 (0)	.008 (.008)
HS GPA (non-missing)	3.994 (.237)	3.961 (.236)	.033** (.013)	3.982 (.232)	3.963 (.247)	.019 (.031)
Fraction Female	.525 (.500)	.507 (.500)	.018 (.027)	.550 (.500)	.539 (.500)	.011 (.064)
Missing Family Inc.	.199 (.400)	.185 (.389)	.014 (.022)	.200 (.402)	.164 (.372)	.036 (.049)
Family Inc/1000 (non-missing)	79.393 (69.850)	62.542 (68.239)	16.851** (4.196)	65.005 (49.535)	57.979 (43.828)	7.026 (6.553)

Table 7 (Continued): Differences in Means across Students who ARE and ARE NOT Required to take Remedial Writing: Across the Full Sample and Across Students Whose AWP Graders Disagreed on Placement

Variable		All Students Not Identified as ESL (n=1497) Remediation NOT Required (St.Dev.)	Remediation Required (St.Dev.)	Difference (St.Error)	Students Not Identified as ESL Whose Graders Disagreed (n=248) Remediation Required (St.Dev.)	Remediation NOT Required (St.Dev.)	Difference (St.Error)
Missing Race		.092	.076	.017	.075	.063	.013
		(.290)	(.265)	(.015)	(.264)	(.243)	(.032)
Among Reporters	Fraction Asian	.398	.439	-.041	.432	.383	.049
		(.490)	(.497)	(.028)	(.498)	(.488)	(.065)
	Fraction Hispanic	.129	.181	-.052**	.198	.208	-.010
		(.335)	(.385)	(.020)	(.400)	(.408)	(.053)
	Fraction Other	.034	.037	-.003	.045	.008	.037*
		(.180)	(.188)	(.010)	(.208)	(.091)	(.021)
	Fraction White	.440	.344	.096**	.324	.400	-.076
		(.497)	(.476)	(.028)	(.470)	(.492)	(.063)
Missing Dad's Ed		.062	.085	-.024*	.017	.094	-.077**
		(.241)	(.280)	(.014)	(.129)	(.293)	(.029)
Among Reporters	Dad HS Drop Out	.076	.146	-.069**	.110	.155	-.045
		(.266)	(.353)	(.017)	(.314)	(.364)	(.044)
	Dad HS Grad	.122	.146	-.023	.127	.129	-.002
		(.328)	(.353)	(.019)	(.335)	(.337)	(.044)
	Dad Some College	.215	.228	-.013	.220	.190	.031
		(.411)	(.420)	(.024)	(.416)	(.394)	(.053)
	Dad Coll Grad	.586	.480	.105**	.542	.526	.017
		(.493)	(.500)	(.028)	(.500)	(.501)	(.065)
Missing Mom's Ed		.044	.042	.002	.042	.070	-.029
		(.206)	(.200)	(.011)	(.201)	(.257)	(.029)
Among Reporters	Mom HS Drop Out	.090	.174	-.084**	.130	.185	-.054
		(.287)	(.380)	(.018)	(.338)	(.390)	(.048)
	Mom HS Grad	.138	.151	-.014	.113	.101	.012
		(.345)	(.359)	(.019)	(.318)	(.302)	(.041)
	Mom Some College	.271	.288	-.017	.304	.336	-.032
		(.445)	(.453)	(.025)	(.462)	(.474)	(.061)
	Mom Coll Grad	.501	.386	.115**	.452	.378	.074
		(.500)	(.487)	(.028)	(.500)	(.487)	(.065)
Sample Size		1019	503		120	128	

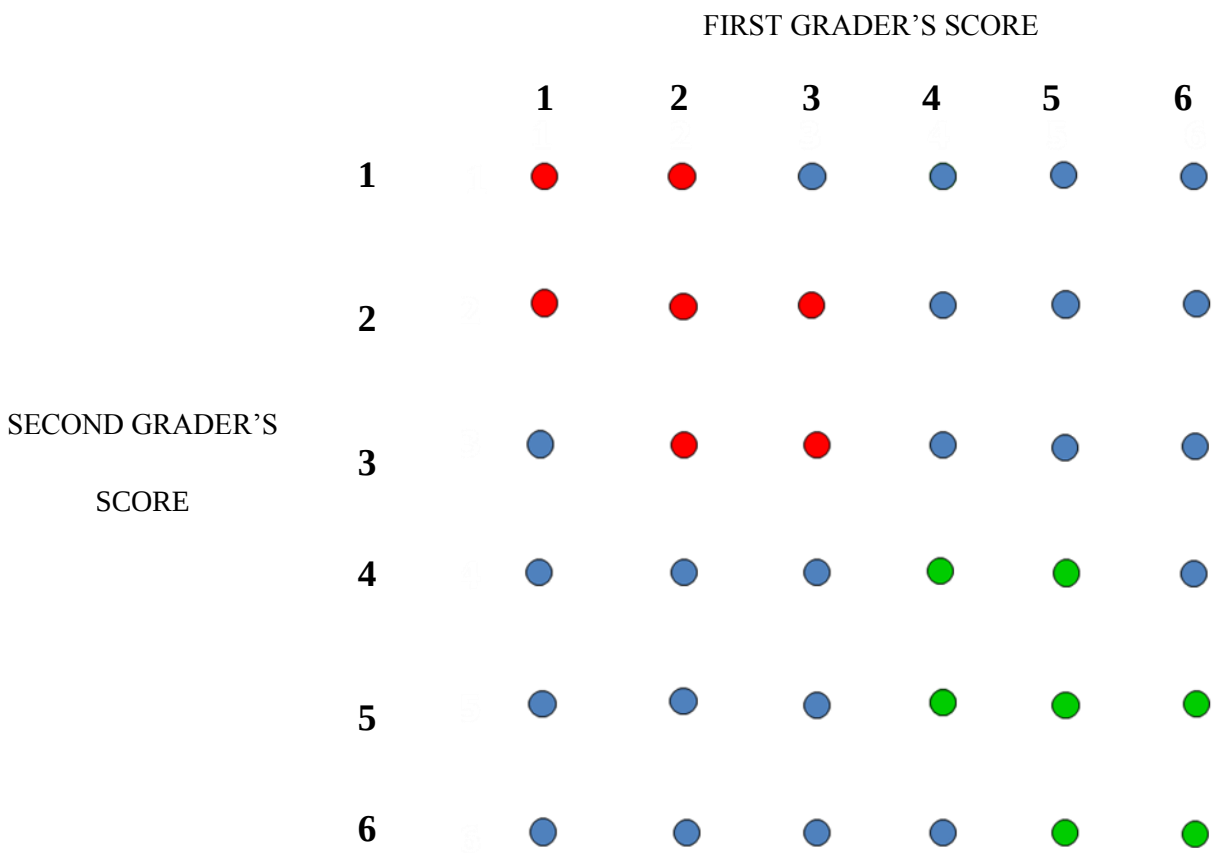
Notes: *=10% significance; **=5% significance

Table 8: Remedial Writing Placement and UCSD performance: All AWP Test-Takers and Test-Takers whose Graders Disagreed on Placement

	ALL AWP TEST TAKERS				TEST-TAKERS WHOSE GRADERS DISAGREED			
Dependent variable	Mean Dependent Variable (1)	Remediation Required Coef. Unconditional (2)	Remediation Required Coef. Conditional (3)	Sample Size (4)	Mean Dependent Variable (5)	Remediation Required Coefficient (6)	Remediation Required Coef. (Conditional) (7)	Sample Size (8)
Dropped Out in 2 years	.075 (.264)	.061** (.015)	.053** (.017)	1497	.081 (.273)	.043 (.035)	.055 (.039)	248
Dropped Out in 5 years	.140 (.347)	.071** (.019)	.049** (.022)	1497	.133 (.340)	.016 (.043)	.029 (.050)	248
Earned BA or BS in 4 years	.495 (.500)	-.122** (.028)	-.069** (.031)	1497	.484 (.501)	-.144** (.063)	-.104 (.075)	248
Earned BA or BS in 5 years	.789 (.408)	-.102** (.022)	-.079** (.026)	1497	.790 (.408)	-.099** (.052)	-.126** (.060)	248
Earned BA or BS in 6 years	.831 (.375)	-.081** (.021)	-.059** (.024)	1497	.827 (.379)	-.061 (.048)	-.063 (.056)	248
Total Credit Hours Excluding Remedial	180.835 (47.060)	-6.261** (2.605)	-6.966** (2.926)	1497	180.696 (47.348)	-.420 (6.028)	-7.732 (6.902)	248
Year 1 Credit Hours: Excluding Remedial	44.542 (5.998)	-5.054** (.306)	-4.188** (.344)	1497	43.617 (5.548)	-5.166** (.625)	-4.882** (.738)	248
Year 1 Credit Hours: UCSD+Remedial	45.963 (5.329)	-.614** (.295)	.085 (.333)	1497	45.827 (4.800)	-.885 (.609)	-.815 (.718)	248
Year 1 GPA: Excluding Remedial	2.924 (.572)	-.221** (.031)	-.108** (.032)	1497	2.914 (.553)	-.173** (.070)	-.171** (.073)	248
Standardized Year 1 GPA Excluding Remedial	-.124 (.679)	-.266** (.037)	-.129** (.038)	1497	-.152 (.659)	-.207** (.083)	-.191** (.088)	248
GPA in 1st College Writing class*	2.940 (.596)	-.237** (.039)	-.119** (.042)	1099	2.938 (.585)	-.149* (.087)	-.163 (.101)	178
Cum GPA of Students Graduating in 6 yrs.	3.147 (.385)	-.178** (.023)	-.083** (.024)	1244	3.126 (.396)	-.135** (.055)	-.095 (.063)	205

Notes: Each element of columns 2, 3, 6 and 7 reports the coefficient and standard error of the "Remediation Required" variable from separate OLS regressions. The conditional regressions in columns (3) and (7) include SAT I-Math, SAT I-Verbal, SAT II Writing, SAT II Math, AP Lit and Comp, AP Lang & Comp, high school GPA, parent's income, dummy variables for female, Asian, Hispanic, whether mom is a college grad, whether dad is a college grad, and missing dummies for each variable. One of UCSD's six colleges graded their writing class P/NP only and so I exclude students from this college in the "GPA in 1st College Writing class" regressions. *=10% significance and **=5% significance.

Figure 1: The Scoring of the University of California's Analytic Writing Placement Exam (AWP)



RED DOTS: Student fails based on first two grader's scores (sum of two scores is 6 or less)

GREEN DOTS: Student passes based on first two grader's scores (sum of two scores is 8 or more)

BLUE DOTS: Student's exam goes to a 3rd grader and student's score is two times the 3rd grader's score.

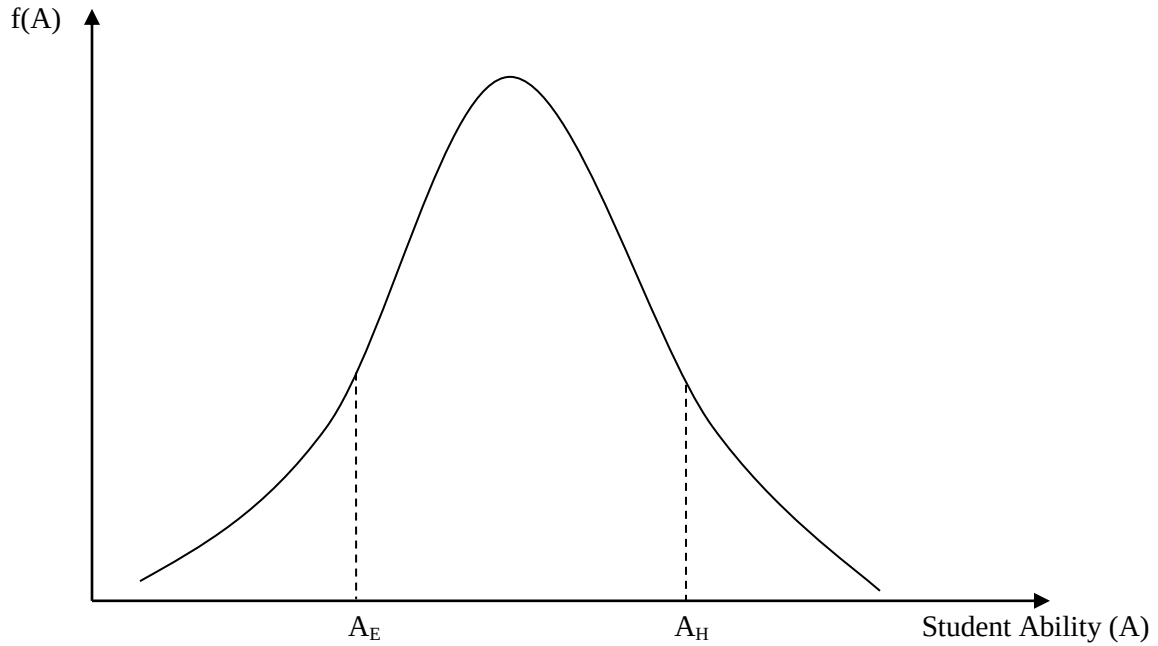
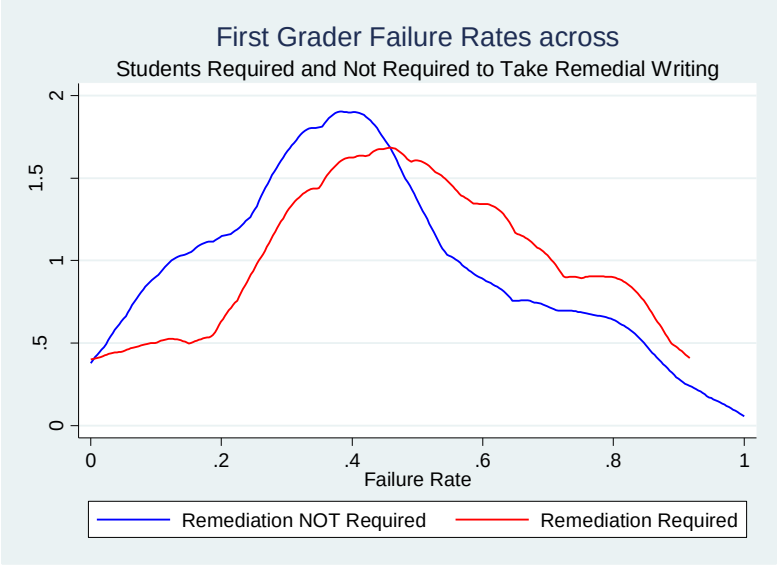


Figure 2: Bias Introduced by Observing Grader IDs only for Students Passed by One Grader and Failed by Another Grader

Let grader *E* be an easy grader who will pass any student whose ability level exceeds A_E. Let grader *H* be a hard grader who will only pass students whose ability level exceeds A_H. Suppose we randomly assign two graders to score each exam but we only observe grader IDs when one grader fails the student and the other grader passes the student. We will observe *H* when she passes a high ability student that an even tougher grader fails which will only happen for students whose ability is above the tough grader threshold ($A_i > A_H$). We will observe *E* when she passes a low ability student that a tougher grader fails which will only happen for students whose ability is above the easy grader's threshold ($A_i > A_E$). So, for students who are passed by our two graders and failed by another grader, the average ability of the tough grader's students is higher than the average ability of the easy grader's students. We will also observe H when she fails a student that an easier grader passes which will only happen for students whose ability is below the tough grader threshold ($A_i < A_H$). We will observe *E* when she fails a low ability student that an even easier grader passes which will only happen for students whose ability is below the easy grader's threshold ($A_i < A_E$). So, for students who are failed by our two graders but passed by another grader, the average ability of the tough grader's students is higher than the average ability of the easy grader's students. As a result, even though grader assignment is random, tough graders will be associated with higher ability students and easy graders will be associated with lower ability students when we observe graders only when they disagree on placement.

FIGURE 3: The Distribution of Grader Failure Rates Across Students Required and Not Required to Take Remedial Writing



NOTES: Failure rates calculated for graders identified at least five times as either a student's first or second grader. Figure for 461 students not identified as having English as a second language by any grader.

Appendix A: Estimating the Effects of Remedial Writing using Grader-Level Averages

Unit of observation: 79 Graders

Dependent variable	Average Over All Students		Average Over Students with SAT Math in 5^{th}-75^{th} percentile		Average Over Students with SAT Verbal in 10^{th}-90^{th} percentile	
	OLS	IV	OLS	IV	OLS	IV
Dropped Out in 2 years	.037 (.052)	.031 (.097)	.067 (.051)	-.007 (.103)	.075 (.049)	.151 (.101)
Dropped Out in 5 years	-.026 (.057)	.198 (.142)	.024 (.068)	.166 (.110)	.009 (.051)	.282* (.154)
Earned BA or BS in 4 years	-.117 (.094)	-.265 (.194)	-.119 (.096)	-.232 (.148)	-.055 (.108)	-.134 (.219)
Earned BA or BS in 5 years	-.026 (.074)	-.249 (.166)	-.069 (.076)	-.164 (.131)	-.046 (.070)	-.289* (.164)
Earned BA or BS in 6 years	-.016 (.060)	-.266* (.146)	-.043 (.068)	-.260** (.108)	-.041 (.057)	-.329** (.157)
Total UCSD Credit Hours	.110 (8.880)	-19.758 (17.867)	-8.346 (8.732)	-26.481* (15.123)	-4.962 (8.206)	-36.748* (19.880)
First Year Credit Hours UCSD Only	-3.633** (1.047)	-3.387** (1.792)	-2.946** (.990)	-4.459** (1.752)	-3.335** (1.037)	-3.760* (2.011)
First Year Credit Hours UCSD + Remedial Classes	.191 (1.079)	.532 (1.800)	1.266 (.973)	-.245 (1.693)	.465 (1.060)	-.041 (2.013))
First Year GPA UCSD classes	.014 (.096)	.071 (.181)	.011 (.105)	.175 (.173)	.015 (.087)	.052 (.191)
First Year Standardized GPA UCSD classes	.070 (.111)	.135 (.197)	.024 (.122)	.163 (.199)	.025 (.105)	.033 (.209)
GPA in First UCSD writing class*	-.113 (.128)	.091 (.254)	-.189 (.146)	.296 (.243)	-.057 (.122)	.245 (.275)
Cum GPA of Students Graduating in 6 yrs	-.065 (.091)	-.026 (.171)	-.054 (.086)	.108 (.148)	-.046 (.085)	-.034 (.158)

Notes: Each element of the table represents a separate regression with the "Remediation Required" coefficient reported. Other regressors are SAT I-Math and Verbal, SAT II-Math and Writing, and high school GPA as well as fraction female, Asian, Hispanic, other race and missing race. I exclude 39 of the 461 students who were incomplete reporters of SAT scores or HS GPA. All regressions are weighted by number of students over which the average is taken and robust standard errors are presented. For all three samples, I reject the hypothesis that there is a significant relationship between average student characteristics and the grader's failure rate. For the IV regressions, the first stage t-statistic is 4.35 for all students and 5.21 for students whose math scores, and 4.53 for students whose verbal scores, put the student closer to the placement margin.
*=10% significance, **=5% significance

Appendix B: Remedial Writing and UCSD performance: Requiring More Exams Per Grader.

	Students with SAT I Verbal between 10th and 90th percentiles				Students with SAT I Math between 5th and 75th percentiles			
	6 exams per grader		7 exams per grader		6 exams per grader		7 exams per grader	
Dependent variable	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dropped Out	.065	.040	.087	.111	.088*	-.126	.109**	-.047
in 2 years	(.041)	(.177)	(.041)	(.159)	(.045)	(.175)	(.047)	(.149)
Dropped Out	.029	.215	.062	.334	.068	.088	.107**	.103
in 5 years	(.054)	(.209)	(.053)	(.216)	(.050)	(.166)	(.051)	(.139)
Earned BA or BS	.011	-.089	-.007	-.261	-.027	-.168	-.044	-.109
in 4 years	(.075)	(.283)	(.080)	(.277)	(.084)	(.214)	(.093)	(.179))
Earned BA or BS	-.059	-.287	-.098*	-.331	-.107*	-.101	-.157**	-.066
in 5 years	(.060)	.240	(.058)	(.240)	(.055)	(.198)	(.056)	(.169)
Earned BA or BS	-.029	-.361	-.063	-.412*	-.075	-.217	-.117**	-.175
in 6 years	(.055)	(.232)	(.053)	(.234)	(.047)	(.188)	(.048)	(.144)
Total UCSD Credit	-14.808**	-21.561	-16.187**	-31.884	-15.187**	-14.532	-17.570	-27.310
Hours	(7.107)	(31.624)	(7.230)	(29.931)	(7.514)	(25.142)	(7.787)	(22.793)
UCSD Credit Hours:	-3.775**	-.997	-3.457**	-1.987	-3.658**	-2.126	-3.568**	-2.500
Year 1	(.662)	(2.577)	(.670)	(2.313)	(.733)	(2.007)	(.781)	(1.930)
UCSD+Remedial	.430	2.340	.722	1.209	.862	1.514	.927	.685
Hours: Year 1	(.654)	(2.570)	(.659)	(2.317)	(.711)	(2.063)	(.766)	(1.965)
First Year GPA	-.075	.058	-.040	.213	-.080	.184	-.043	.369*
UCSD classes	(.079)	(.344)	(.084)	(.351)	(.091)	(.235)	(.097)	(.203)
Standardized 1st Year	-.114	.142	-.078	.284	-.124	.276	-.077	.462*
GPA UCSD classes	(.097)	(.415)	(.105)	(.429)	(.112)	(.284)	(.119)	(.258)
GPA in 1st UCSD	-.084	.584	-.070	.617	-.034	.682*	-.008	.721*
writing class	(.088)	(.531)	(.095)	(.532)	(.090)	(.413)	(.101)	(.376)
Cum GPA of Students	-.047	-.051	-.019	-.083	-.367	.003	.031	.005
Graduating in 6 yrs.	(.050)	(.238)	(.052)	(.225)	(.781)	(.158)	(.052)	(.146)

Notes: Each element reports the coefficient and standard error on the "Remediation Required" variable from separate OLS and IV regressions. The other regressors are SAT I-Math, SAT I-Verbal, SAT II Writing, SAT II Math, AP Lit & Comp, AP Lang & Comp, high school GPA, parent's income, dummy variables for female, Asian, Hispanic, whether mom is a college grad, whether dad is a college grad, and missing dummies for each variable. In the regressions for GPA in 1st UCSD writing class, I exclude students from one of UCSD's six colleges because their writing class is graded pass/no pass only. I also excluded students who withdrew from the class. Finally, *=10% significance, **=5% significant

Appendix C: Illustrating the Sensitivity of IV Estimates to Calculating the First Grader's Failure Rate Using Exams Scored as (1) a First Grader Only or (2) as either a First or Second Grader

Dependent variable	F_{ij} Based on Exams Scored as First Grader Only Coefficient (Standard Error)	F_{ij} Based on Exams Scored as First or Second Grader Coefficient (Standard Error)
I. ALL STUDENTS (339 students / 38 graders)		
Reject Ho: F_{ij} related to student chars?	NO	NO
First Stage F-statistic	11.22	9.46
Dropped Out in 5 years	.197 (.169)	.182 (.196)
Earned BA or BS in 4 years	-.486* (.281)	-.257 (.245)
Earned BA or BS in 5 years	-.305 (.259)	-.190 (.240)
Earned BA or BS in 6 years	-.242 (.188)	-.192 (.203)
First Year Credits UCSD Only	-2.656 (2.286)	-2.292 (2.315)
First Year Credits UCSD+Remedial	.316 (2.382)	.337 (2.459)
First Year GPA, UCSD classes	.050 (.280)	.244 (.295)
First Year Standardized GPA, UCSD classes	.258 (.303)	.437 (.340)
GPA in First UCSD writing class	.296 (.418)	.575 (.535)
Cum GPA of Students Graduating in 6 yrs	-.164 (.239)	-.041 (.224)
II. SAT I MATH SCORES CLOSE TO PLACEMENT MARGIN (BETWEEN 5th AND 75th PERCENTILES) (220 students/ 38 graders)		
Reject Ho: F_{ij} related to student chars?	NO	YES
First Stage F-statistic	22.08	25.23
Dropped Out in 5 years	.150 (.141)	.114 (.139)
Earned BA or BS in 4 years	-.121 (.170)	-.032 (.160)
Earned BA or BS in 5 years	-.175 (.177)	-.076 (.170)
Earned BA or BS in 6 years	-.201 (.153)	-.180 (.146)
First Year Credits UCSD Only	-2.880 (1.756)	-3.441**(1.559)
First Year Credits UCSD+Remedial	.187 (1.790)	-.479 (1.621)
First Year GPA, UCSD classes	.141 (.195)	.401** (.183)
First Year Standardized GPA, UCSD classes	.233 (.240)	.520** (.225)
GPA in First UCSD writing class	.341 (.313)	.640 (.338)
Cum GPA of Students Graduating in 6 yrs	-.067 (.153)	.059 (.139)

Appendix C (CONTINUED): Illustrating the Sensitivity of IV Estimates to Calculating the First Grader's Failure Rate Using Exams Scored as (1) a First Grader Only or (2) as either a First or Second Grader

Dependent variable	Fij Based on Exams Scored as First Grader Only Coefficient (Standard Error)	Fij Based on Exams Scored as First or Second Grader Coefficient (Standard Error)
III. SAT I VERBAL SCORES CLOSE TO THE PLACEMENT MARGIN (BETWEEN 10th AND 90th PERCENTILES) (245 students and 38 graders)		
Reject Ho: Fij related to student chars?	NO	NO
First Stage F-statistic	11.4	8.55
Dropped Out in 5 years	.376* (.206)	.350 (.237)
Earned BA or BS in 4 years	-.581* (.319)	-.354 (.293)
Earned BA or BS in 5 years	-.461* (.275)	-.367 (.281)
Earned BA or BS in 6 years	-.473** (.241)	-.432* (.259)
First Year Credits UCSD Only	-2.436 (2.257)	-.519 (2.637)
First Year Credits UCSD+Remedial	1.332 (2.132)	2.958 (2.456)
First Year GPA, UCSD classes	-.064 (.353)	.252 (.393)
First Year Standardized GPA, UCSD classes	.040 (.399)	.397 (.467)
GPA in First UCSD writing class	.161 (.492)	1.263 (1.006)
Cum GPA of Students Graduating in 6 yrs	-.186 (.278)	-.020 (.256)

Notes: This table explores whether the results are biased because the first grader's failure rate is based on exams the grader scored both as a first and as a second grader on other student's exams. A bias may arise because I am more likely to observe second grader IDs when graders disagree on placement which is likely to occur when student skills are closer to the passing threshold. The sample is 339 students who are not ESL and whose first grader is the first grader for at least four other students. There are 38 graders. Either way I calculate the grader's failure rate, I typically cannot reject the hypothesis that there is a significant relationship between student demographics and the grader's failure rate: tougher graders are associated with significantly higher SAT I Verbal scores (also higher fraction of female, higher fraction of Asian and lower family income depending upon the regression). Though IV estimates are likely biased upwards, the IV estimate are fairly similar whichever way I estimate the grader failure rate. Each element of the table represents a separate IV regression with the "Remediation Required" coefficient reported. The other regressors are SAT I-Math, SAT I-Verbal, SAT II Writing, SAT II Math, AP Lit & Comp, AP Lang & Comp, high school GPA, parent's income, dummy variables for female, Asian, Hispanic, whether mom is a college grad, whether dad is a college grad, and missing dummies for each variable.

*=10% significance, **=5% significance