Grades and the Economics Major at a Women's College*

Patrick J. McEwan
Wellesley College

Sheridan Rogers
Bates White Economic Consulting

Akila Weerapana
Wellesley College

May 2019

Abstract: This paper uses a regression-discontinuity design to compare the major choices of women who barely received higher or lower letter grades in two introductory economics courses at a women's college. Students just above letter-grade cutoffs have markedly higher probabilities of majoring in economics, although they are otherwise similar to students just below cutoffs. Grade sensitivity cannot be explained by majority-male courses (since classrooms are single-sex) or the presence of male instructors (since grade sensitivity is not moderated by instructor gender). We discuss alternate causal mechanisms, including preferences and behavioral evidence that men and women respond differently to negative feedback. We conclude with a discussion of low-cost interventions that might diminish grade sensitivity and increase the number of high-ability women majoring in economics.

* We are grateful to Kristin Butcher, Phil Levine, Eni Mustafaraj, Seth Neumuller, Kyung Park, Kartini Shastry, Olga Shurchkov, and Brian Tjaden for their comments and assistance. We are also grateful to many faculty members for sharing their grading spreadsheets.

1. Introduction

The college attainment of women exceeds that of men, but there are persistent gender gaps in the choice of undergraduate majors (Goldin, Katz, and Kuziemko, 2006; Altonji, Blom, and Meghir, 2012; Gemici and Wiswall, 2014; Goldin, 2015; Bayer and Rouse, 2016). Women are much less likely to choose mathematically-intensive fields of study such as economics, computer science, mathematics, and the physical sciences (see Figure 1), though a similar gap does not exist in the biological sciences. A voluminous literature has explored the causes of these gaps, largely motivated by the fact that women are less likely to major in fields with potentially high labor market returns. Indeed, gender differences in fields of study may account for over half of the male-female wage gap in the United States (Brown and Corcoran, 1997; Altonji, Blom, and Meghir, 2012).

An important strand of this literature has empirically assessed whether major choice is affected by gender differences in objective ability. It finds that gender gaps in major choice are large and persistent, even after controlling for objective measures of pre-college math skills and high-school preparation (Turner and Bowen, 1999; Weinberger, 2001; Riegle-Crumb et al., 2012). Moreover, the gender gaps in mathematically-intensive majors are large even among colleges with the highest math SAT scores (see Figure 1). Rather than objective ability, the literature has noted that subjective *beliefs* about ability determine major choices, and that beliefs evolve during college (Arcidiacono, 2004; Arcidiacono, Hotz, and Kang, 2012; Stinebrickner and Stinebrickner, 2014).[1] For example, Stinebrickner and Stinebrickner (2014) show that

[1] The literature has also assessed whether gender gaps are explained by differences in preferences for non-pecuniary attributes of majors or careers, or by differences in expected labor market returns, perhaps due to discrimination (e.g., Zafar, 2013). For detailed reviews of the wider literature, see Altonji, Blom, and Meghir (2012) and Altonji, Arcidiacono, and Maurel (2016). This paper will consider the indirect effects of major preferences on major choice, insofar as preferences influence the degree to which choices are influenced by lower grades.

students begin college with significant over-optimism about their chances of completing a science degree, and subsequently adjust downward their self-perceptions of science ability.

Students may learn about their abilities (and update beliefs accordingly) as they take introductory courses.[2] A major concern in the literature is that women respond differently than men to negative signals of ability, such as lower grades. In observational studies, women who receive lower grades in introductory economics courses are much less likely to major in economics than men with similar grades (Horvath, Beaudin, and Wright, 1992; Rask and Tiefenthaler, 2008; Goldin, 2015). Relative grades across fields also have a stronger influence on the choice of women to persist in physical science majors (Ost, 2010) and male-dominated STEM majors (Kugler, Tinsley, and Ukhaneva, 2017). Even so, the partial correlation between grades and major choice might not have a causal interpretation. In particular, it may reflect the effects of correlated but unobserved determinants of major choice, such as skills or preferences.

We identify the pure signaling effect of grades using a regression discontinuity design that compares the major choices of students whose final course averages place them just above or below letter-grade cutoffs (also see Owen, 2010; Main and Ost, 2014).[3] We use 10 years of data from Wellesley College, a highly-selective liberal arts college that only admits women.[4] Students

[2] Related papers report causal estimates of how educational experiences during college—including instructor and classmate gender relative to one's own gender—influence major choice (Bettinger and Long, 2005; Hoffman and Oreopoulos, 2009; Carrell, Page, and West, 2010; Booth, Cardona-Sosa, and Nolen, 2013; Zölitz and Feld, n.d.). For example, Carrell et al. (2010) find that female students with high math ability at the U.S. Air Force Academy are significantly more likely to perform well in math and science classes (and to major in these fields) when they are randomly assigned to a female instructor. One interpretation of the reduced-form effects is that female instructors influence students' beliefs about their abilities in math and science by offering positive role models.

[3] Two pioneering papers found mixed results that may be partly explained by sample sizes and wide confidence intervals. Near the A/B cutoff, Owen (2010) found that women with higher grades in an introductory economics course have an higher probability (by 0.16) of majoring in economics, although men do not. The effects at the B/C cutoff were not statistically distinguishable from zero, although the implied 95% confidence interval includes effects as large as 14 percentage points for women. Near the A/B cutoff, Main and Ost (2014) found that effects for both men and women were not statistically distinguishable from zero. However, the confidence interval is consistent with effects as large as 20 percentage points for women.

[4] Rogers (2017) describes early results from this research project, limited to the sample of students who took Economics 101.

enroll in two sequentially-taught economics courses. Economics 101 is taken by 37% of all students and Economics 102 is taken by 26%. In Economics 101, students just above letter-grade cutoffs are 12 to 13 percentage points more likely to major in economics than students just below cutoffs. In Economics 102, they are 23 to 29 percentage points more likely. Both represent gains of more than 40% over baseline probabilities. There is no evidence that instructors purposely manipulate the values of final scores or cutoffs, since students just above and below cutoffs have similar pre-college math skills and preferences, on average. We cannot reject the null hypothesis that effects are similar across higher letter-grade cutoffs (e.g., A-/B+) and lower ones (e.g., B+/B).

Grade sensitivity is not just an artifact of economics, or even of mathematically-intensive majors. Using a similar research design in computer science, we find that grades in an introductory course have similarly large—albeit less precisely estimated—effects on the decision to major in that field. And, in earlier research, we found that an exogenous drop in grades in the non-economics social sciences—caused by a unique anti-grade-inflation policy at Wellesley College—led to decreased choice of those majors and increased choice of majors that already had lower grades, including economics and the sciences (Butcher, McEwan, and Weerapana, 2014).

There are varied explanations for the grade sensitivity of women, although the literature is not conclusive. The unique data and setting of Wellesley College facilitate direct tests of some causal mechanisms. One hypothesis is that lower grades cause women to view themselves through the lens of a negative stereotype about their performance in a mathematically-intensive field (e.g., "I'm not an Economics person"). If this is the case, then grade-induced stereotype threat should reduce the probability of taking the next economics course, or at least impair students' performance in that course.[5] And yet, we find that students just above or below letter-grade cutoffs in 101 are similarly likely to take 102, and even receive similar grades in 102. Other research has speculated that women are more grade sensitive when they are in majority-male classrooms or when they are taught by male rather than female instructors (e.g., Ost, 2010; Kugler, Tinsley, and Ukhaneva, 2017). However, women are sensitive to grades despite single-sex classrooms in all fields of study. Moreover, we do not find that effects are larger when students are taught by male instructors.

There is better evidence that women's grade sensitivity is explained by weaker preferences for economics-related courses and careers. Our point estimates suggest that women are less responsive to grades when they express a pre-college preference for economics. However, women are far less likely than men to express a pre-college preference for economics (Goldin, 2015), which partly explains their relatively heightened sensitivity to introductory grades. If women's major preferences are based on imperfect information, then a simple and low-cost intervention would be to provide students—before and during college—with more accurate information about the content of the economics major and its career options. Several field experiments evaluate such interventions among college women, but it is too early to tell whether information alone can increase the number of women majoring in economics (Avilova and Goldin, 2018; Li, 2018; Bayer, Bhanot, and Lozano, 2019).[6]

Even if women and men have similar major preferences, research in behavioral economics suggests a final reason why women are more sensitive to grades. Experiments find that women

[5] The experimental literature on stereotype threat, reviewed below, has consistently found that the mathematics performance of female college students declines when they are spurred by cue to view themselves through the lens of a negative stereotype. These include simple demographic questions, as well as the presence of males in the testing conditions (e.g., Inzlicht and Ben-Zeev, 2000; Danaher and Crandall, 2008).

[6] Li (2018) finds that an informational interventions significantly increases the number of high-ability women majoring in economics, but the average effect is entwined with that of a related intervention that targets encouragement towards women with high exam grades.

who receive negative feedback are more likely to attribute it to low ability, while men are more likely to attribute it to luck, unless they already evaluated themselves as below-average (Shastry, Shurchkov, and Xia, 2018). Similarly, lower grades are a noisy signal of ability, and perhaps more likely to dissuade women—even students with high abilities—from majoring in economics. A practical implication is that women's major choices may be influenced by the targeted encouragement of high-ability women to take additional courses (e.g., Li, 2018).

2. Background on Wellesley College

Seventeen percent of Wellesley College students receive a degree in economics, making it the most popular major (see Table 1). The major requires nine courses. Three introductory courses include 101 (Introductory Microeconomics), 102 (Introductory Macroeconomics), and 103 (Probability and Statistics). Three intermediate courses include 201 (Intermediate Microeconomics), 202 (Intermediate Macroeconomics), and 203 (Econometrics). Finally, 200- and 300-level electives provide intermediate and advanced overviews of fields such as development, health, trade, finance, and the like.

Introductory courses are prerequisites to intermediate ones, and intermediate courses to advanced ones. Furthermore, the college sequences its introductory courses so that students have to complete Economics 101 (or use AP or IB credits) before they can enroll in Economics 102. Because of this, a large majority of students take 101 and 102 in the first four semesters. Between Fall 2004 and Spring 2014, 37% of students completed Economics 101 (see Table 1). Of these, 41% did so in the first semester, 30% in the second, 15% in the third, and 6% in the fourth. Twenty-six percent of all students completed Economics 102. Of these, 1.5% did so in the first semester, 40% in the second, 28% in the third, and 16% in the fourth.

Students must declare a major during the fourth semester or, if they intend to study abroad, during the third. (They also have the option to declare a minor, which consists of three introductory courses and two electives.) Given the sequencing and timing of introductory courses, students are likely to use their grades in Economics 101 and 102 as a way of learning about major-specific abilities.

Figure 2 reports a histogram of Economics 101 grades, pooling across multiple years of data (the distribution is similar in Economics 102). The modal letter grade is a B, while the next most common grades are B+, A-, and B-. In short, there is little grade compression at the top of the grading scale. The top-right panel of Figure 2 shows that the unconditional probability of majoring in Economics is much higher among "A" students (about 0.5) than students with lower grades. We aim to assess whether this correlation has a causal interpretation, although the descriptive data in two bottom panels provide reason for caution (and a motive for using exogenous variation near letter-grade cutoffs). Students with higher grades are also more likely to have higher math SAT scores, as well as a preference for economics listed on their college application.

3. Empirical Strategy and Data

A. Identification and Estimation

As a starting point, one might regress a dependent variable indicating the economics major on indicators of 101 letter grades, as well as controls for the pre-college variables. The empirical dilemma is that letter grades are plausibly correlated with omitted variables that influence major choice, such as unobserved abilities or preferences. Thus, it is not feasible to identify the pure signaling effect of introductory grades without strong assumptions about unobserved variables.

Instead, we implement a regression-discontinuity design. Instructors calculate a final percentage score as the weighted average of exam scores and other assessments. Define S_{ijk} as the final score of student *i,* taught by instructor *j* in academic term *k*. Instructors order students by values of S_{ijk} and choose letter-grade cutoffs. For the moment, let us focus on a single cutoff: instructors define z_{jk} and then allocate students to receive either a B+ or an A-.

Pooling across many instructors and terms—each with different values of z_{jk}—we can estimate a local linear regression among students with values of S_{ijk} "close" to z_{jk} (we later revisit methods used to define this bandwidth):

(1) $$M_{ijk} = \alpha + \beta G_{ijk} + \theta(S_{ijk} - z_{jk}) + \mu(G_{ijk} \times (S_{ijk} - z_{jk})) + \delta_{jk} + \varepsilon_{ijk},$$

where M_{ijk} is a dummy variable indicating that student *i* graduated with an economics major, having taken 101 with instructor *j* in academic term *k*. $G_{ijk} = 1\{(S_{ijk} - z_{jk}) \geq 0\}$ is equal to one if a student's score in 101 exceeds the A-/B+ cutoff, and zero otherwise. The discontinuity is sharp, rather than fuzzy, unless instructors subjectively assign a B+ to at least one student above the cutoff (or an A- to a student below it). The local linear specification controls for a piecewise spline in S_{ijk}, allowing the slope to vary around the cutoff. The inclusion of term-by-instructor fixed effects (δ_{jk}) ensures that comparisons near cutoffs are between students who took the course in the same term with the same instructor.[7]

β is the magnitude of the discontinuity in the dependent variable at the cutoffs. It can be interpreted as the local average treatment effect of receiving an A- (rather than B+) on the probability of majoring in economics—as long as the relationship between S_{ijk} and the potential outcomes of students is continuous near the cutoffs (Hahn, Todd, and van der Klaauw, 2001; Lee

[7] Sometimes instructors teach more than one section of an introductory course in the same semester. However, they apply the same assessments, and determine letter-grade cutoffs in the pooled sample of students.

and Lemieux, 2010). This seems plausible, as long as instructors do not manipulate S_{ijk} or z_{jk} in order to ensure that a particular student receives a higher or lower grade. In the presence of manipulation, a letter-grade "effect" might be due to the unobserved qualities of students near cutoffs.

To rule out manipulation, we assess whether there are discontinuities in a rich set of pre-college variables (see Table 1), which might be consistent with instructor manipulation of S_{ijk} or z_{jk}. We also report estimates that drop a small number of observations near letter-grade cutoffs. These observations are the most plausibly contaminated if instructors precisely manipulated S_{ijk} or z_{jk} to ensure that a particular student received a particular grade (for a related application, see McEwan and Shapiro, 2008).

B. Pooled Sample and Specification

Between Fall 2004 and Spring 2014, 2,815 students completed Economics 101, and 1,908 students completed Economics 102 (see Table 1). We gathered 68 grading spreadsheets from 101 instructors, and 48 from 102 instructors. The spreadsheets include 58% and 56%, respectively, of 101 and 102 students.[8] The spreadsheets do not report exact cutoffs, and so we imputed them. Define z_{jk}^{s} as the grade cutoff s for instructor j in term k, where s=1 is the A/A- cutoff, s=2 is the A-/B+ cutoff, and so on. We impute each z_{jk}^{s} as the midpoint between two values of S_{ijk}: that of the highest A- (or B+, B, etc.) in a term-by-instructor group and the next highest value. For example, suppose that the highest A- within a term-by-instructor group is 92.2%. If the next highest value within the cell is 92.4%, then $z_{jk}^{1} = 92.3$. We then assign each student to a unique term-by-instructor-by-cutoff group, also indexed by s. Note that students might be assigned to

[8] Not all spreadsheets were available, primarily due to missing electronic records and faculty departures.

two groups: an A-, for example, could be assigned to s=1 or s=2. Instead, we uniquely allocate an A- student to s=1 (or s=2) if her score is higher (or lower) than $(z_{jk}^1 + z_{jk}^2)/2$, the midpoint of the A/A- and A-/B+ cutoffs. We follow a similar procedure to assign students with other letter grades to mutually-exclusive term-by-instructor-by-cutoff cells, indicated by fixed effects (δ_{jk}^s).

This facilitates a regression in the 101 sample (and, separately, in the 102 sample) that pools across instructors, terms, and letter-grade cutoffs:

(2) $$M_{ijk}^s = \alpha + \beta G_{ijk}^s + \theta(S_{ijk}^s - z_{jk}^s) + \mu(G_{ijk}^s \times (S_{ijk}^s - z_{jk}^s)) + \delta_{jk}^s + \varepsilon_{ijk}^s,$$

where M_{ijk}^s is the major outcome of student i near cutoff s, and taught by instructor j in term k. The pooled specification maximizes the statistical power of the design, but assumes that effects are homogeneous across letter-grade cutoffs. As Table 1 describes, the 101 and 102 samples are fairly well distributed across the letter-grade cutoffs. For example, 18% of students are in s=1 (A/A-), 18% in s=2 (A-/B+), 30% in s=3 (B+/B), and 20% in s=4 (B/B-). In additional specifications, therefore, we estimate effects at each letter-grade cutoff, and formally test for homogeneity of effects.

C. Pre-College Variables

Panel A of Table 1 describes seven variables that measure pre-college attributes of students. These include SAT and ACT scores, a quantitative reasoning assessment taken just before first-semester classes, a student's self-declared preference for the economics major (at the time of application), a binary proxy of income (based on students' receipt of financial aid), and an admissions score that reflects a global assessment of the student's college application. Not surprisingly, students who actually take an introductory course are more likely than all students to have declared a preference for economics. However, they are similarly likely to receive financial aid, and their test scores are slightly, but not markedly, higher.

In the following section, we will analyze the non-random subset of 101 and 102 students for which we were able to obtain grading spreadsheets; we refer to these as regression-discontinuity (RD) samples. Panel A of Table 1 shows that students in the RD samples are substantively similar to the larger samples of students that completed either 101 or 102.

4. Regression-Discontinuity Results

A. Visual Evidence

Figures 3 and 4 illustrate the main results in the 101 and 102 samples, respectively. In the panels of each figure, the circles are unweighted means of residuals taken within bins of one-third of a percentage point of the assignment variable ($S_{ijk}^{s} - z_{jk}^{s}$). The residuals are from regressions of the y-axis variable on fixed effects (δ_{jk}^{s}) that indicate mutually-exclusive combinations of term-by-instructor-by-cutoff. (We do so because the "local" randomized assignment to letter grades occurred within the groups indicated by the fixed effects.) The solid lines are fitted values from linear regressions—estimated on either side of the cutoff—in which the residuals are regressed on the assignment variable, applying triangular weights that decrease in proportion to observations' distances from the cutoff. In all panels, the bandwidth is four percentage points, which trims outliers from the full regression-discontinuity (RD) samples described in Table 1.

The upper-left panels in Figures 3 and 4 are "first-stage" estimates, in which we verify that students' grade points in introductory courses are sharply determined by their final scores relative to cutoffs. Indeed, students with values of S_{ijk}^{s} just above the cutoffs had 101 grades that were 0.32 higher, and 102 grades that were 0.31 higher (i.e., a third of a letter grade). In subsequent analyses, we proceed as if the regression-discontinuity design is sharp rather than fuzzy. On the other hand, the upper-right panels show no evidence of discontinuities in students' pre-college math skills—specifically, the quantitative reasoning assessment taken just before the first semester.[9]

The lower-right panels clearly illustrate that the probability of majoring in economics jumps discontinuously in the vicinity of letter-grade cutoffs. In the 101 sample, it increases by 11 percentage points, an increase of 37% over the proportion of majors to the left of cutoffs. In the 102 sample, it increases by 24 percentage points, an increase of 57% over the proportion of majors to the left of cutoffs. The 101 and 102 samples are subject to very slight attrition, since 5% of the 101 sample and 2% of the 102 sample did not graduate from Wellesley College. To increase the sample size, therefore, we analyze a dummy variable indicating whether students took Economics 201 (Intermediate Microeconomics). It is a nearly perfect proxy for majoring, with a correlation of 0.97, since few students undertake advanced courses without the promise of a degree credential. The magnitudes of the discontinuities—in the lower-left panels of the figures—are similar to those of the major variable.

B. Main Estimates

Table 2 reports point estimates and standard errors for the RD sample of Economics 101 students. The estimates in the first two columns apply the same bandwidth (4 percentage points) as Figure 3. The first column simply reports estimates of equation (3), while the second includes controls for baseline variables. Columns 3 and 4 report estimates (also with controls) in smaller

---

[9] The point estimates for the quantitative reasoning score (listed below each figure's panel) are both less than 6% of a standard deviation of the dependent variable.

samples defined by a data-driven bandwidth selector (Calonico et al., forthcoming). The bandwidth selectors in both columns are mean-squared-error (MSE) optimal, although the selector in the first column omits a regularization term, leading to slightly larger bandwidths than otherwise. Concretely, we apply the rdbwselect package in Stata (Calonico et al., 2017; forthcoming) to the RD samples described in Table 1, allowing MSE-optimal bandwidths to vary on either side of the cutoff. For computational simplicity, we apply the bandwidth selector to the residuals analyzed in Figures 3 and 4 (that is, residuals after controlling for δ_{jk}^{s}), although the bandwidth selector explicitly controls for the pre-college variables described in panel A of Table 1.

As in Figure 3, the discontinuities in the first row are consistent with a sharp regression-discontinuity design: students just above grade cutoffs have 101 grades about one-third of a letter grade higher. We also report a simple placebo test. If students' grades in 101 reflected a component of good or bad luck—at least near the cutoff—then grades in other courses taken in the same semester should be similar near the 101 cutoff. The second row confirms that is the case, with very small and statistically insignificant effects on the grade point average in non-Economics courses.

The grade in Economics 101 had no effect on the probability of taking Economics 102, the next introductory course in the sequence. Given the apparent absence of grade-induced selection into 102, we further estimate the effect of 101 grades on the eventual grade point in 102 (we shall later discuss mechanisms for such an effect). As it happens, the effects on the 102 grade are only slightly positive, and not statistically distinguishable from zero.

Lastly, the point estimates suggest that the probability of majoring in Economics increases by a stable 12 to 13 percentage points (an increase of 40% to 43%). The estimate in the final column—within the narrowest bandwidth—is not significant at conventional levels. However, there are uniformly large and statistically significant effects on the major proxy (having taken Economics 201). We also calculated robust p-values for all the estimates (Calonico, Cattaneo, and Titiunik, 2014; Calonico et al., forthcoming), and these mirror conclusions from conventionally-estimated standard errors and p-values.[10] In contrast to the Economics major, the 101 grade does not have an effect on the probability of minoring.

Table 3 reports similar estimates in the RD sample of Economics 102. As above, the discontinuity in the 102 grade is consistent with a sharp, rather than fuzzy design. There are large and statistically significant effects on probability of majoring in Economics (23 to 29 percentage points). The point estimates are double those from the 101 sample, although the effects are only modestly larger when expressed as percentage increases (53% to 67%), given the larger proportion of 102 students that ultimately major in Economics. The effect on the probability of taking Economics 201 is substantially similar. As in the 101 sample, there are no effects of 102 grades on the probability of minoring in Economics.

Our discontinuity estimates show that the major choices of women are responsive to the letter grade in economics courses (also see Owen, 2010). Indeed, the discontinuity estimates are broadly consistent with observational studies using grade variation that is less plausibly exogenous (Horvath, Beaudin, and Wright, 1992; Rask and Tiefenthaler, 2008; Ost, 2010; Goldin, 2015). This suggests a final exercise in our data. We re-estimated the major regressions in Tables 2 and 3, including all controls (e.g., pre-college covariates) except for continuous measures of the assignment variable (S_{ijk}^{s}). By excluding the latter variables, we also use grade variation "away" from letter-grade cutoffs that is potentially confounded with omitted

[10] For the economics major, the robust p-values for the estimates in the third and fourth columns are, respectively, 0.04 and 0.16. For Economics 201, the robust p-values are 0.01 and 0.02.

determinants of major choice. Even so, the observational estimates suggest that effects on major choice are large and statistically significant (but not larger than the comparable discontinuity estimates).[11] We cautiously conclude that the observational literature probably does not overstate the effects of letter grades on major choices (and may do the opposite).

C. Internal Validity

The internal validity of the estimates might be threatened if instructors manipulated final scores (S_{ijk}^{s}) or letter-grade cutoffs (z_{jk}^{s}). The placebo tests in Tables 2 and 3 did not suggest this is the case, since students' grades in classes other than 101 and 102 were similar near the cutoffs. Table A1 further tests for discontinuities in seven pre-college variables (including the quantitative reasoning score already illustrated in Figures 3 and 4). Despite several coefficients at the margin of statistical significance, there is no consistent evidence—across bandwidths or variables—of discontinuities in pre-college variables. Consistent with this, the inclusion of covariates did not substantively affect the point estimates in Tables 2 and 3, although it led to slight reductions in most standard errors.

We next assessed robustness of the main results to dropping observations very close to the cutoffs, since the grades corresponding to these observations are most likely to be precisely manipulated by instructors. As a benchmark specification, we used a bandwidth of four percentage points and included controls to maximize precision. In Table 2, the benchmark estimate (standard error) was 0.134 (0.049) for the economics major variable. In Table 3, it was 0.226 (0.055). We dropped observations within 0.25 percentage points of the cutoff, reducing the

[11] For 101, the coefficient (standard error) is 0.124 (0.026); for 102, it is 0.118 (0.032).

101 sample by 7% and the 102 sample by 6%, and re-estimated the regressions. In both cases, the estimates and standard errors were consistent with the benchmarks.[12]

D. Is Grade Sensitivity Limited to Economics Courses?

We present two pieces of complementary evidence that grade sensitivity at Wellesley College extends beyond the field of economics. First, we gathered data from an introductory course in computer science and applied a similar research design. Between Fall 2004 and Spring 2014, 718 students completed Computer Science 111, and we obtained grading spreadsheets for 395 students; this constitutes the RD sample (see Table A2). Table A3 reports estimates of equation (2), using the same bandwidth selection procedures as before. The first-stage estimates confirm that the CS 111 grade point is about one-third of a letter grade higher for students just above the cutoffs (i.e., a sharp discontinuity). Across all specifications, the point estimates imply that the probability of majoring in computer science increases by at least 12 percentage points when students receive a higher grade. Since 22% of students to the left of the cutoff ultimately choose the major, this represents a 55% increase, on average. The estimates cannot be statistically distinguished from zero because of the comparatively small samples, but the point estimates are large and comparable to the more precise estimates for economics.

Second, an earlier paper used a different source of exogenous variation in grades to estimate effects on major choice (Butcher, McEwan, and Weerapana, 2014). At most colleges and universities, there is grade compression at the top of the letter grade scale, and it is particularly common in the humanities and the non-economics social sciences (Sabot and Wakeman-Linn, 1991; Rojstaczer and Healy, 2010; Butcher, McEwan, and Weerapana, 2014). In Fall 2004, Wellesley College capped the mean grade at 3.33 (B+) in 100- and 200-level classes with at least

[12] In the 101 sample, they were 0.155 (0.055). In the 102 sample, they were 0.178 (0.063).

10 students, although it was only binding on higher-grading departments in the humanities and the non-economics social sciences.

Grades immediately fell in these departments—so as to barely comply with the cap—although they still remained modestly higher than the sciences and economics. Butcher, McEwan, and Weerapana (2014) compared major choices—before and after the policy change—in a treatment group of newly-complying departments and a comparison group of always-complying departments. Majors decreased in treated departments, relative to the comparison group. In particular, students shifted major elections from the non-economics social sciences (e.g., political science and psychology), and towards economics and the sciences.

E. Heterogeneity

In Table 4, we explore whether the effects are heterogenous across the different cutoffs (i.e., whether the difference between receiving a B+ or a B has less of an impact on major choice than the difference between an A- and a B+).[13] As expected, students just above each cutoff have grades about one-third of a letter grade higher, and we cannot reject the null that the coefficients are equal. Moreover, the effects on economics major choice—both the direct and proxy measures—are uniformly positive for both 101 (panel A) and 102 (panel B). They vary in magnitude, but not in a way suggesting an obvious pattern, and we cannot reject the null hypothesis of equality.

[13] For each moderating variable (e.g., five categories of letter-grade cutoffs), we calculate dummy variables indicating the mutually-exclusive and exhaustive set of categories (Table 4 excludes a few observations corresponding to the sixth and least common cutoff at C+/C). We fully interact the dummies with G_{ijk}^{s} in equation (2) and report these coefficients in the tables. In addition, we fully interact the dummies with the continuous measures of the assignment variable in equation (2), so that estimated slopes are allowed to vary by each category of the moderator. All regressions include term-by-instructor-by-cutoff fixed effects and uninteracted controls for pre-college variables.Table 4 reports effects at each of five letter-grade cutoffs, for both 101 (panel A) and 102 (panel B).

Table 5 assesses whether the effects are moderated by several variables, including pre-college math skills and major preferences, a household income proxy, and instructor gender in the introductory course. In brief, there is no evidence that effects on major choice are larger for students whose quantitative reasoning scores are above or below the median.[14] The point estimates of effects on major choice are larger for the subset of students that did not declare a preference for economics on their college application, particularly for 102. However, we cannot reject equality of the effects across the two groups at conventional levels of statistical significance. The proxy of household income is whether students did not receive any financial aid (about one-third of the sample), versus some financial aid. Table 5 shows that major effects are larger—in a practical and statistical sense—among students with at least some financial aid, but only in the 101 sample.

Lastly, Table 5 assesses whether effects are moderated by the gender of 101 and 102 instructors. In top colleges and universities, less than one-fifth of economics faculty are female, including tenured, tenure-track, and adjunct (CSWEP, 2017). At Wellesley College, the economics faculty is gender-balanced, as is the instruction in 101 and 102 sections. The point estimates in Table 5 suggest that the major choices of female students are more sensitive to grades when they are taught by female instead of male instructors, although we cannot reject equality of the effects.

5. Why Are Women (More) Sensitive to Grades?

A. Stereotype Threat, Stereotype Vulnerability, and Perceived "Fit"

The literature has speculated about causal mechanisms that could explain why women's major choices are particularly sensitive to grades, both in an absolute sense and relative to men.

[14] We prefer this pre-college measure of math scores because it is available for all students, unlike SAT scores.

First, social psychologists have repeatedly shown that the mathematical performance of female college students is affected by stereotype threat (for a review, see Good, Aronson, and Harder, 2008). Stereotype threat occurs when situational cues—often subtle ones—lead students to view themselves through the lens of a negative or positive stereotype, which in turn affects their performance. Female performance on math tasks declines when tests are characterized as sensitive to gender differences (Spencer, Steele, and Quinn, 1999), when demographic questions are posed before rather than after a test (Danaher and Crandall, 2008), and when women complete tasks in the presence of men (Inzlicht and Ben-Zeev, 2000; Sekaquaptewa and Thompson, 2003). One might hypothesize is that lower grades in introductory economics courses cause women to view themselves through the lens of a negative stereotype about female performance in a mathematically-intensive field (e.g., "I'm not really an Economics person").[15] However, recall that the letter-grade cue is received upon *completion* of 101. If stereotype threat mediates the ultimate effect on major choice, then one would anticipate that performance would decline in the next economics course, or at least that students would be less likely to take it. But, on the contrary, we did not find that 101 grades affected take-up or performance in 102 (see Table 2).[16]

Second, Ost (2010) speculates that female students in the physical science classes—typically a minority in coeducational settings—are particularly vulnerable to stereotypes about female performance in mathematically-intensive fields. He further notes stereotype-vulnerable students have been shown to have unstable academic self-concept and efficacy, which is associated with "increased sensitivity to performance feedback, both positive and negative" (Aronson and

[15] Owen (2010) provides a related discussion when interpreting her results.
[16] If 101 grades had affected 102 performance, then stereotype threat is not necessarily the causal mechanism. Another is that students use the 101 letter grade to guide allocation of effort in 102 (e.g., working harder to compensate for a lower-than-anticipated grade).

Inzlicht, 2004, p. 834, cited in Ost, 2010). Coeducational settings might therefore be expected to increase grade sensitivity when women are in majority-male courses (e.g., physics and chemistry), but not in gender-balanced courses (e.g., biology). Kugler, Tinsley, and Ukhaneva (2017) describe a different model with similar implications, in which majority-male courses cause women, particularly those with low grades, to infer that they are a poor fit for the major. Men in majority-male courses infer that they are a good fit.

The single-sex classrooms at Wellesley College might be expected to attenuate grade sensitivity, since they plausibly reduce vulnerability to negative stereotypes about female performance (Ost, 2010), and offer better signals of fit (Kugler, Tinsley, and Ukaneva, 2017). In fact, we find that women's major choices are quite sensitive to grades even at a women's college. It is still possible that women are stereotype vulnerable (or infer poor fit) in the presence of *any* male in the classroom, including course instructors. Approximately half of students take 101 and 102 with a male instructor, and one might expect these students to be relatively more sensitive to grades. Yet, Table 5 reported that grade sensitivity was statistically similar in 101 and 102 sections taught by males and females, and point estimates suggested that sensitivity was actually higher with female instructors.

Kugler, Tinsley, and Ukhaneva (2017) extend their model to assume that stereotypically male fields of study (e.g., STEM majors) send a further signal to women about fit, regardless of whether classrooms are majority-male. It may provide an additional explanation as to why women are sensitive to grades in a single-sex setting. We cannot directly assess this explanation, but it seems inconsistent with evidence that women's major choices are also affected grades in non-STEM fields (Butcher, McEwan, and Weerapana, 2014).

B. Preferences and Behavioral Explanations

Rask and Tiefenthaler (2008) hypothesize that men are less sensitive to economics grades because they have more entrenched preferences for economics coursework or the types of careers facilitated by it. Indeed, Zafar (2013) argues that women and men at a selective university are markedly different in their workplace preferences, with males caring more about pecuniary outcomes; in their setting, preferences explain most of the gender gaps in major choice. At a selective, coeducational liberal arts college, Goldin (2015) reports that men are twice as likely as women to express a preference for economics before starting classes. The point estimates in Table 5 suggested that the major choices of women with an expressed preference for economics on their college applications are less responsive to grades than women without such a preference (but with the caveat that the estimates are not statistically different). In this view, at least part of the gender difference in grade sensitivity is not caused by gender per se. Rather, gender differences in pre-college preferences mediate grade sensitivity.

Finally, there is evidence from behavioral economics that men and women respond differently to noisy feedback (Shastry, Shurchkov, and Xia, 2018). A letter grade is a particular form of such feedback: while it communicates major-specific abilities to students, it also contains an element of luck. (Indeed, our regression-discontinuity design explicitly compared “lucky” and “unlucky” students in the vicinity of letter grade cutoffs.) Shastry, Shurchkov, and Xia (2018) conducted a series of lab experiments to explore how men and women respond differently to positive or negative feedback that reflects an ability component as well as luck. Women were more likely to attribute negative feedback to a lack of ability, regardless of their actual ability. On the other hand, men were more likely to attribute negative feedback to luck (unless it confirmed a negative self-evaluation of ability). The implication is that that men who receive a lower grade in an introductory course are more likely than women to choose and persist in a major.

6. Conclusions

This paper used a regression-discontinuity design to identify the causal effect of introductory course grades on major choice at a women's college. We compared the choices of women on either side of letter-grade cutoffs. Women just above the cutoffs in two introductory courses were far more likely to major in economics (increasing baseline probabilities by more than 40%). There were no discontinuities in a variety of pre-college variables, including math skills and major preferences.

We evaluated several causal mechanisms that might explain the robust effects of grades on women's major choice, both in this paper and the wider literature (Horvath, Beaudin, and Wright, 1992; Rask and Tiefenthaler, 2008; Owen, 2010; Goldin, 2015). One plausible explanation is related to preferences. Our estimates show that a pre-college preference for economics tends to make women less grade sensitive. Women are less likely than men to declare a preference for economics at the time of application (Goldin, 2015). Therefore, full-sample estimates that compare men and women will likely reveal more grade sensitivity among women (as in Rask and Tiefenthaler, 2008). Even when men and women share similar preferences, evidence from behavioral economics suggests that women are more likely to attribute negative signals to ability, while men are more likely to attribute them to luck (Shastry, Shurchkov, and Xia, 2018).

The evidence on grade sensitivity suggests two avenues for research and policy. The first is to develop a greater understanding of how preferences for mathematically-intensive majors and careers are shaped before and during college. To the extent that preferences reflect a (partial) misunderstanding of the actual content of majors and associated careers, then simple and low-cost informational interventions could increase the preference of women for economics (and diminish grade sensitivity as a byproduct). For example, students might misperceive the economics major as a pre-business track that exclusively prepares students for careers in finance and banking (while being unaware of the broader menu of options in public policy, research, and other careers).

Bayer, Bhanot, and Lozano (2019) conducted a randomized experiment in which students were sent emails with an encouragement to take economics courses and additional information about the types of research conducted by economists. It had modest effects on the likelihood of taking additional courses. Avilova and Goldin (2018) describe an ongoing experiment in which college students are provided with better information about the economics major and the myriad careers that it may facilitate. Li (2018) evaluated a treatment that provided information about economics-related careers and class grade distributions, as well as targeted encouragement of women with above-median scores in economics. The combined intervention had large effects on the probability that high-ability women chose a major in economics, although separate contribution of career information (versus other interventions) could not be separately identified.

The second avenue is to develop and test interventions that might diminish the sensitivity of women to grades, especially high-ability women who receive an “unlucky” grade. For example, instructors could provide students with better information on how narrowly they missed a higher letter grade; more nuanced evaluations of students’ accomplishments and likelihood of succeeding in future coursework; or simply targeted nudges of encouragement to promising students (e.g., Li, 2018).

References

Altonji, Joseph G., Peter Arcidiacono, and Arnaud Maurel. 2016. "The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects." Edited by Eric Hanushek, Stephen Machin, and Ludger Woessmann. In *Handbook of the Economics of Education* (vol. 5): 305–396.

Altonji, Joseph G., Erica Blom, and Costas Meghir. 2012. "Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers." *Annual Review of Economics, 4:* 185–223.

Arcidiacono, Peter. 2004. "Ability Sorting and the Returns to College Major." *Journal of Econometrics, 121:* 343–375.

Arcidiacono, Peter, V. Joseph Hotz, and Songman Kang. 2012. "Modeling College Major Choices Using Elicited Measures of Expectations and Counterfactuals." *Journal of Econometrics, 166:* 3–16.

Aronson, Joshua, and Thomas Dee. 2011. "Stereotype Threat in the Real Word." In Michael Inzlicht and Toni Schmader (Eds.), *Stereotype Threat: Theory, Process, and Application.* Oxford: Oxford University Press.

Avilova, Tatyana, and Claudia Goldin. 2018. "What Can UWE Do for Economics?" *AEA Papers and Proceedings, 108:* 186–190.

Bayer, Amanda, and Cecilia Elena Rouse. 2016. "Diversity in the Economics Profession: A New Attack on an Old Problem." *Journal of Economic Perspectives, 30:* 221–242.

Bayer, Amanda, Syon O. Bhanot, and Fernando Lozano. 2019. "Does Simple Information Provision Lead to More Diverse Classrooms? Evidence from a Field Experiment on Undergraduate Economics." *AEA Papers and Proceedings, 109:* 110–114.

Bettinger, Eric P., and Bridget T. Long. 2005. "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students." *AEA Papers and Proceedings, 95:* 152–157.

Booth, A. L., L. Cardona-Sosa, and P. Nolen. 2013. "Do Single-Sex Classes Affect Achievement? An Experiment in a Coeducational University." Unpublished manuscript, University of Essex.

Brown, Charles, and Mary Corcoran. 1997. "Sex-Based Differences in School Content and the Male-Female Wage Gap." Journal of Labor Economics, 15: 431–465.

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

Butcher, Kristin F., Patrick J. McEwan, and Akila Weerapana. 2014. "The Effects of an Anti-Grade-Inflation Policy at Wellesley College." *Journal of Economic Perspectives, 28:* 189–204.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik. Forthcoming. "Regression Discontinuity Designs Using Covariates." *Review of Economics and Statistics.*

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik. 2017. "rdrobust: Software for Regression-Discontinuity Designs." *The Stata Journal, 17:* 372–404.

Calonico, Sebastian, Matias D. Cattaneo, and Rocío Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Design." *Econometrica, 82:* 2295–2326.

Carrell, S., M. E. Page, and J. E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *Quarterly Journal of Economics:* 1101–1144.

Committee on the Status of Women in the Economics Profession (CSWEP). 2017. "The 2016 Report on the Status of Women in the Economics Profession." Downloaded July 27, 2016 from: https://www.aeaweb.org/content/file?id=3643.

Danaher, K., and C. S. Crandall. 2008. "Stereotype Threat in Applied Settings Re-Examined." *Journal of Applied Social Psychology, 38:* 1639–1655.

Gemici, A., and M. Wiswall. 2014. "Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors." *International Economic Review, 55:* 23–56.

Goldin, Claudia, L. F. Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap. *Journal of Economic Perspectives, 20:* 133–156.

Goldin, Claudia. 2015. "Gender and the Undergraduate Economics Major: Notes on the Undergraduate Economics Major at a Highly Selective Liberal Arts College." Unpublished manuscript.

Good, Catherine, Joshua Aronson, and Jayne Ann Harder. 2008. "Problems in the Pipeline: Stereotype Threat and Women's Achievement in High-Level Math Courses." *Journal of Applied Developmental Psychology, 29:* 17–28.

Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica, 69*(1): 201–209.

Hoffman, F., and P. Oreopoulos. 2009. "A Professor Like Me: The Influence of Instructor Gender on College Achievement." *Journal of Human Resources, 44:* 479–494.

Horvath, Jane, Barbara Q. Beaudin, and Sheila P. Wright. 1992. "Persisting in the Introductory Economics Course: An Exploration of Gender Differences." *Journal of Economic Education, 23:* 101–108.

Inzlicht, Michael, and Talia Ben-Zeev. 2000. "A Threatening Intellectual Environment: Why Females Are Susceptible to Experiencing Problem-Solving Deficits in the Presence of Males." *Psychological Science, 11:* 365–371.

Kugler, Adriana D., Catherine H. Tinsley, and Olga Ukhaneva. 2017. "Choice of Majors: Are Women Really Different from Men?" NBER Working Paper No. 23735.

Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature, 48:* 281–355.

Li, Hsueh-Hsiang. 2018. "Do Mentoring, Information, and Nudge Reduce the Gender Gap in Economics Majors?" *Economics of Education Review, 64:* 165–183.

Main, Joyce, and Ben Ost. 2014. "The Impact of Letter Grades on Student Effort, Course Selection, and Major Choice: A Regression-Discontinuity Analysis. *Journal of Economic Education, 45:* 1–10.

McEwan, Patrick J., and Joseph S. Shapiro. 2008. "The Benefits of Delayed Primary School Enrollment: Discontinuity Estimates Using Exact Birth Dates." *Journal of Human Resources, 43:* 1–29.

Ost, Ben. 2010. "The Role of Peers and Grades in Determining Major Persistence in the Sciences." *Economics of Education Review, 29:* 923–934.

Owen, Ann L. 2010. "Grades, Gender, and Encouragement: A Regression-Discontinuity Analysis." *Journal of Economic Education, 41:* 217–234.

Rask, Kevin, and Jill Tiefenthaler. 2008. "The Role of Grade Sensitivity in Explaining the Gender Imbalance in Undergraduate Economics." *Economics of Education Review, 27:* 676–687.

Riegle-Crumb, Catherine, Barbara King, Eric Grodsky, and Chandra Muller. 2012. "The More Things Change, the More They Stay the Same? Prior Achievement Fails to Explain Gender Inequality in Entry Into STEM College Majors Over Time." *American Educational Research Journal, 49:* 1048–1073.

Rogers, Sheridan. 2017. "A- or B+: The Signaling Effect of the Letter Grade." Unpublished B.A. thesis, Wellesley College.

Rojstaczer, Stuart, and Christopher Healy. 2010. "Grading in American Colleges and Universities." *Teachers College Record,* March 4.

Sabot, Richard, and John Wakeman-Linn. 1991. "Grade Inflation and Course Choice." *Journal of Economic Perspectives, 5:* 159–170.

Sekaquaptewa, Denise, and Mischa Thompson. 2003. "Solo Status, Stereotype Threat, and Performance Expectancies: Their Effects on Women's Performance." *Journal of Experimental Social Psychology, 39:* 68–74.

Shastry, G. Kartini, Olga Shurchkov, and Lingjun Xia. 2018. "Luck or Skill: How Women and Men Respond to Noisy Feedback." Unpublished manuscript, Wellesley College.

Spencer, Steven J., Claude M. Steele, and Diane M. Quinn. 1999. "Stereotype Threat and Women's Math Performance." *Journal of Experimental Social Psychology, 35:* 4–28.

Stinebrickner, Ralph, and Todd R. Stinebrickner. 2014. "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout." *Review of Economic Studies, 81:* 426–472.

Turner, S. E., and W. G. Bowen. 1999. "Choice of Major: The Changing (Unchanging) Gender Gap." *Industrial and Labor Relations Review, 52:* 289–313.

Weinberger, Catherine J. 2001. "Is Teaching More Girls More Math the Key to Higher Wages?" In Mary C. King, ed., *Squaring Up: Policy Strategies to Raise Women's Incomes in the U.S.* Ann Arbor: University of Michigan Press.

Zafar, B. 2013. "College Major Choice and the Gender Gap. *Journal of Human Resources, 48:* 545–595.

Zölitz, U., and J. Feld. (n.d.). "The Effect of Peer Gender on Major Choice and Occupational Segregation." Unpublished manuscript.

Table 1: Descriptive statistics

	All students			Ever took Economics 101		Ever took Economics 102	
				All	RD sample	All	RD sample
	Mean	S.D.	N	Mean	Mean	Mean	Mean
Panel A: Pre-college variables							
Preference for Economics major (1/0)	0.09	(0.29)	7611	0.15	0.16	0.19	0.22
Quantitative reasoning score	13.0	(2.7)	7611	13.2	13.5	13.3	13.6
ACT	29.8	(2.9)	2534	30.0	30.5	30.1	30.5
SAT (Mathematics)	681.5	(67.2)	6751	691.2	701.3	696.2	703.7
SAT (Verbal)	689.9	(68.9)	6752	691.0	694.4	688.6	693.4
Receives no financial aid (1/0)	0.37	(0.48)	7611	0.39	0.39	0.38	0.38
Admissions score	4.17	(0.42)	7602	4.14	4.12	4.11	4.09
Panel B: College variables							
Ever took Economics 101 (1/0)	0.37	(0.48)	7611	1.00	1.00	0.90	0.92
Grade point in Economics 101	3.13	(0.61)	2849	3.13	3.23	3.22	3.31
G.P.A. in non-Economics courses (semester of 101)	3.32	(0.44)	2878	3.32	3.38	3.33	3.40
Ever took Economics 102 (1/0)	0.26	(0.44)	7611	0.64	0.67	1.00	1.00
Grade point in Economics 102	3.15	(0.60)	1991	3.15	3.22	3.16	3.28
G.P.A in non-Economics courses (semester of 102)	3.34	(0.43)	2000	3.34	3.38	3.34	3.41
Ever took Economics 201 (1/0)	0.15	(0.36)	7611	0.33	0.37	0.47	0.51
Economics minor (1/0)	0.06	(0.24)	7084	0.13	0.12	0.18	0.18
Economics major (1/0)	0.17	(0.38)	7084	0.32	0.37	0.45	0.49
Did not graduate (1/0)	0.07	(0.25)	7611	0.06	0.05	0.03	0.02
Panel C: Letter-grade cutoff samples							
A/A- (1/0)					0.18		0.21
A-/B+ (1/0)					0.18		0.25
B+/B (1/0)					0.30		0.28
B/B- (1/0)					0.20		0.17
B-/C+ (1/0)					0.11		0.04
C+/C (1/0)					0.03		0.05
Maximum number of observations	7611			2815	1623	1908	1072

Notes: "All students" include traditionally-aged, non-transfer students who enrolled in any Wellesley College course between Fall 2004 and Spring 2014. "Ever took" samples include the subset of students who took and received a grade in Economics 101 (or 102). "RD samples" are further limited to students with complete data from instructor grading spreadsheets in either 101 or 102.

Table 2: Student outcomes at letter-grade cutoffs in Economics 101

		Bandwidth for local linear regressions			
		4 percentage points		MSE-optimal (1)	MSE-optimal (2)
Grade point in 101 (3.04)		0.312***	0.311***	0.322***	0.340***
		(0.008)	(0.008)	(0.011)	(0.012)
	N	1,466	1,466	1,248	916
G.P.A. in non-Economics courses (semester of 101) (3.33)		0.026	0.018	-0.026	0.005
		(0.035)	(0.035)	(0.044)	(0.051)
	N	1,463	1,463	1,292	1,048
Completed Economics 102 (0.64)		0.005	0.012	-0.056	-0.074
		(0.046)	(0.046)	(0.062)	(0.070)
	N	1,466	1,466	1,102	958
Grade point in Economics 102 (3.10)		0.073	0.079	0.091	0.105
		(0.055)	(0.055)	(0.057)	(0.081)
	N	975	975	959	701
Minored in Economics (0.13)		-0.044	-0.049	-0.042	-0.029
		(0.036)	(0.035)	(0.037)	(0.057)
	N	1,388	1,388	1,376	988
Completed Economics 201 (0.30)		0.130***	0.144***	0.196***	0.196**
		(0.048)	(0.047)	(0.070)	(0.081)
	N	1,466	1,466	1,216	1,006
Majored in Economics (0.30)		0.120**	0.134***	0.126**	0.125
		(0.051)	(0.049)	(0.051)	(0.080)
	N	1,388	1,388	1,365	993
Controls?		N	Y	Y	Y

Notes: Each cell reports an estimate of β from equation (2), applying triangular weights within the bandwidth specified in the column header (see text for an explanation of MSE-optimal bandwidths). Robust standard errors are in parentheses. *** indicates statistical significance at 1%; ** at 5%; * at 10%. The final row indicates regressions that contain the controls listed in panel A of Table 1, as well as dummy variables indicating missing values. In the first column, the number in parentheses is the mean of the dependent variable to the left of grade cutoffs (within 4 percentage points of the cutoff).

Table 3: Student outcomes at letter-grade cutoffs in Economics 102

		Bandwidth for local linear regressions			
				MSE-optimal	
		4 percentage points		(1)	(2)
Grade point in 102 (3.12)		0.304***	0.305***	0.314***	0.320***
		(0.008)	(0.008)	(0.013)	(0.017)
	N	1,013	1,013	640	485
G.P.A. in non-Economics courses (semester of 102) (3.39)		-0.018	-0.009	-0.038	-0.126**
		(0.038)	(0.037)	(0.052)	(0.064)
	N	1,009	1,009	793	593
Minored in Economics (0.13)		0.005	0.008	-0.020	-0.038
		(0.047)	(0.048)	(0.052)	(0.065)
	N	989	989	922	689
Completed Economics 201 (0.45)		0.236***	0.216***	0.280***	0.304***
		(0.057)	(0.055)	(0.066)	(0.077)
	N	1,013	1,913	843	677
Majored in Economics (0.43)		0.246***	0.226***	0.311***	0.292***
		(0.057)	(0.055)	(0.066)	(0.064)
	N	989	989	856	856
Controls?		N	Y	Y	Y

Notes: Each cell reports an estimate of β from equation (2), applying triangular weights within the bandwidth specified in the column header (see text for an explanation of MSE-optimal bandwidths). Robust standard errors are in parentheses. *** indicates statistical significance at 1%; ** at 5%; * at 10%. The final row indicates regressions that contain the controls listed in panel A of Table 1, as well as dummy variables indicating missing values. In the first column, the number in parentheses is the mean of the dependent variable to the left of grade cutoffs (within 4 percentage points of the cutoff).

Table 4: Heterogeneity by letter-grade cutoff

	Grade point in 101 (panel A) or 102 (panel B)	Minored in Economics	Completed Economics 201	Majored in Economics
Panel A: RD sample; Economics 101				
A/A-	0.326***	-0.092	0.233***	0.250***
	(0.005)	(0.073)	(0.087)	(0.091)
A-/B+	0.322***	-0.069	0.144	0.099
	(0.011)	(0.063)	(0.101)	(0.103)
B+/B	0.303***	-0.033	0.141*	0.126*
	(0.016)	(0.056)	(0.072)	(0.075)
B/B-	0.309***	-0.058	0.065	0.039
	(0.015)	(0.060)	(0.081)	(0.085)
B-/C+	0.305***	-0.000	0.225**	0.354***
	(0.023)	(0.081)	(0.103)	(0.107)
p-value	.52	.92	.64	.15
Panel B: RD sample; Economics 102				
A/A-	0.324***	-0.061	0.133	0.154
	(0.005)	(0.094)	(0.100)	(0.105)
A-/B+	0.338***	0.033	0.135	0.123
	(0.004)	(0.062)	(0.098)	(0.095)
B+/B	0.311***	-0.081	0.252***	0.280***
	(0.010)	(0.073)	(0.085)	(0.087)
B/B-	0.307***	0.215*	0.133	0.142
	(0.025)	(0.116)	(0.121)	(0.130)
B-/C+	0.270***	0.148	0.573***	0.294
	(0.043)	(0.222)	(0.158)	(0.258)
p-value	.05	.22	.13	.74

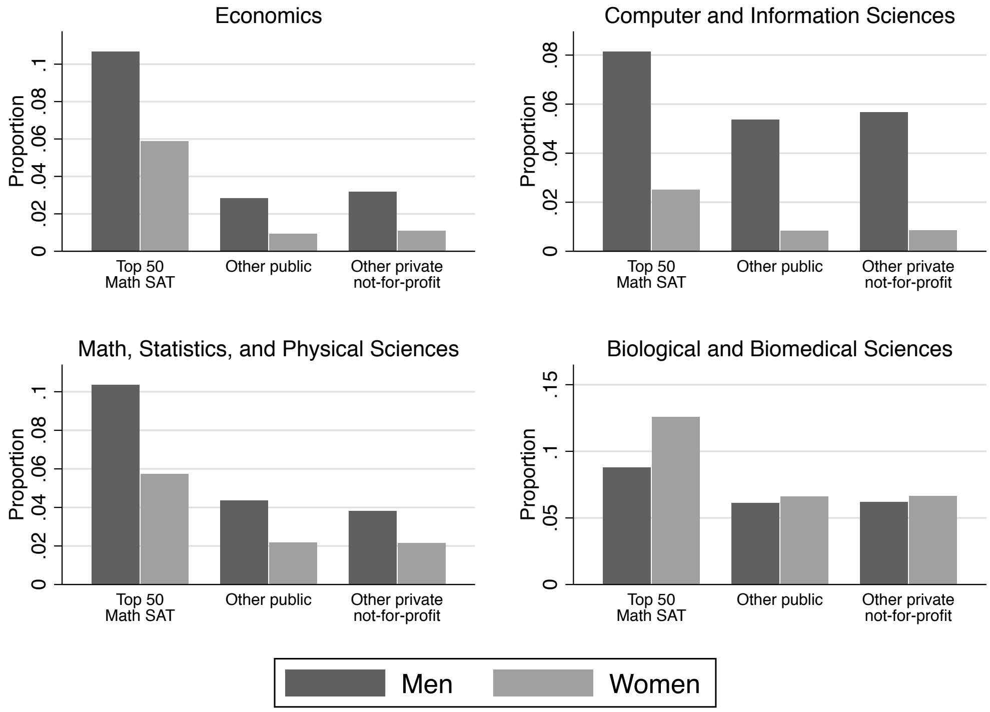
Notes: Each cell reports an estimate of β from equation (2), applying triangular weights and a bandwidth of 4 percentage points. Robust standard errors are in parentheses. *** indicates statistical significance at 1%; ** at 5%; * at 10%. All regressions include the controls listed in panel A of Table 1, as well as dummy variables indicating missing values. The p-values correspond to a test of the null that the coefficients are jointly equal.

Table 5: Heterogeneity by math skills, preferences, income, and instructor gender

	Grade point in 101 (panel A) or 102 (panel B)	Minored in Economics	Completed Economics 201	Majored in Economics
Panel A: RD sample; Economics 101				
Lower QR score	0.314***	-0.036	0.126**	0.133**
	(0.009)	(0.044)	(0.060)	(0.062)
Higher QR score	0.308***	-0.066	0.165***	0.137**
	(0.011)	(0.045)	(0.061)	(0.064)
p-value	.69	.57	.60	.96
Econ preference	0.327***	-0.082	0.053	0.080
	(0.026)	(0.058)	(0.121)	(0.121)
No Econ preference	0.308***	-0.042	0.162***	0.144***
	(0.009)	(0.040)	(0.051)	(0.054)
p-value	.51	.56	.41	.63
Any financial aid	0.312***	-0.094**	0.197***	0.226***
	(0.011)	(0.043)	(0.061)	(0.062)
No financial aid	0.311***	0.017	0.066	-0.010
	(0.012)	(0.062)	(0.077)	(0.081)
p-value	.99	.14	.19	.02
Female instructor	0.317***	-0.065	0.200***	0.213***
	(0.008)	(0.056)	(0.069)	(0.072)
Male instructor	0.307***	-0.035	0.088	0.057
	(0.013)	(0.046)	(0.065)	(0.068)
p-value	.52	.69	.24	.11
Panel B: RD sample; Economics 102				
Lower QR score	0.311***	0.015	0.238***	0.210***
	(0.009)	(0.056)	(0.072)	(0.074)
Higher QR score	0.299***	-0.001	0.199***	0.247***
	(0.010)	(0.065)	(0.069)	(0.070)
p-value	.31	.84	.66	.69
Econ preference	0.308***	-0.042	0.019	0.091
	(0.020)	(0.086)	(0.108)	(0.115)
No Econ preference	0.304***	0.027	0.274***	0.266***
	(0.009)	(0.058)	(0.065)	(0.066)
p-value	.85	.51	.05	.20
Any financial aid	0.301***	0.035	0.244***	0.228***
	(0.011)	(0.061)	(0.070)	(0.071)
No financial aid	0.311***	-0.042	0.170*	0.226**
	(0.011)	(0.076)	(0.091)	(0.093)
p-value	.54	.43	.53	.99
Female instructor	0.298***	-0.059	0.329***	0.376***
	(0.019)	(0.098)	(0.108)	(0.107)
Male instructor	0.306***	0.033	0.165**	0.156**
	(0.009)	(0.054)	(0.064)	(0.065)
p-value	.72	.41	.19	.08

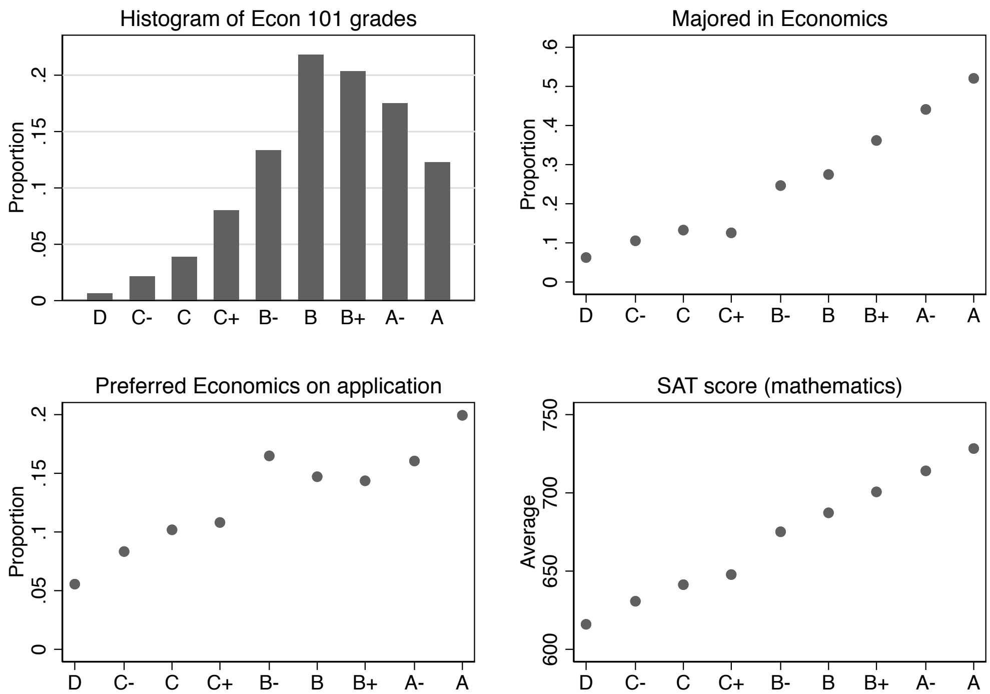
Notes: Each cell reports an estimate of β from equation (2), applying triangular weights and a bandwidth of 4 percentage points. Robust standard errors are in parentheses. *** indicates statistical significance at 1%; ** at 5%; * at 10%. All regressions include the controls listed in panel A of Table 1, as well as dummy variables indicating missing values. The p-values correspond to a test of the null that the coefficients are jointly equal.

Figure 1: Bachelor's degrees by major and gender, 2015



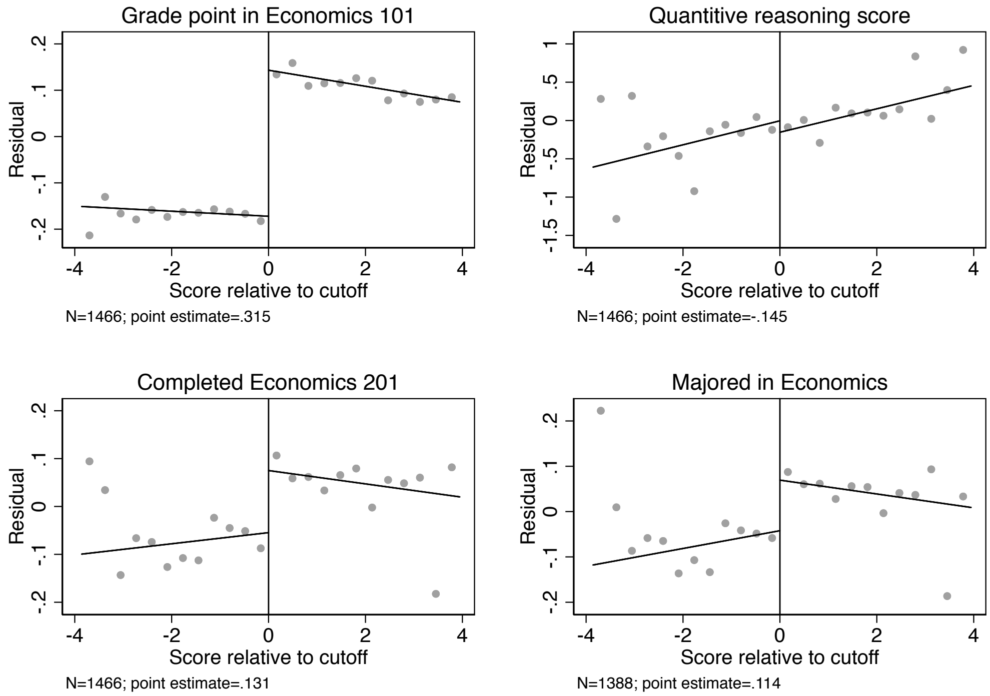
Source: Integrated Postsecondary Education Data System, 2015, and authors' calculations.
Notes: The "top 50 math SAT" category is ranked by the 25th percentile of math SAT scores.

Figure 2: Introductory grades and the Economics major



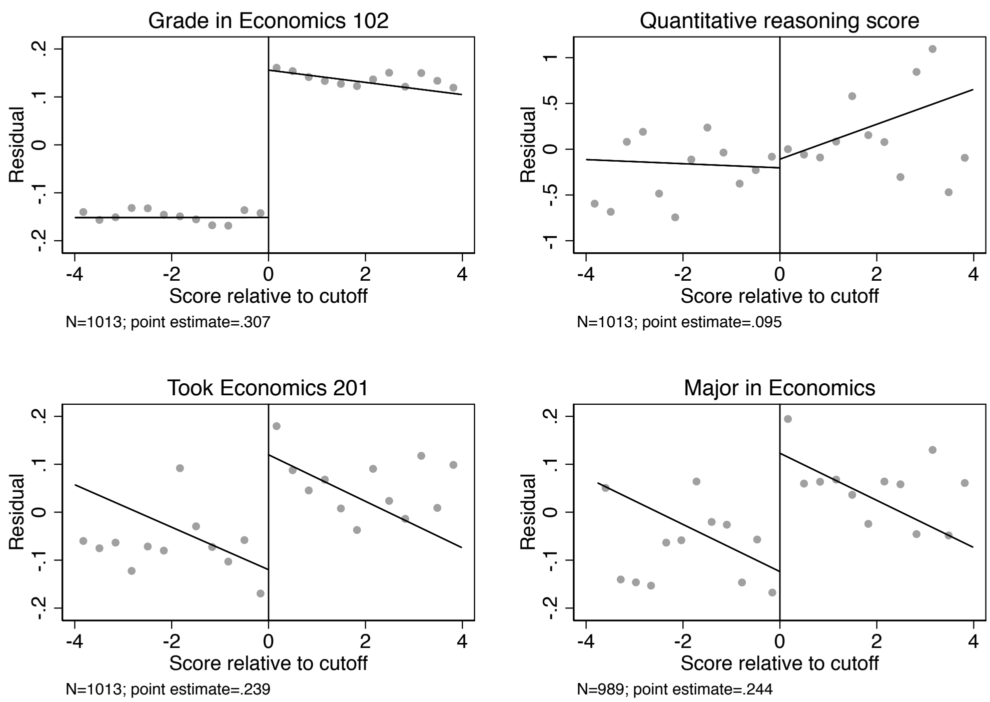
Notes: The panels include 2774 students that took Economics 101 between Fall 2004 and Spring 2014 (except the top-right, which includes 2657 students that graduated by Spring 2017).

Figure 3: Student outcomes at grade cutoffs (Economics 101)



Note: The sample includes 1466 students who took Economics 101 between Fall 2004 and Spring 2014 (see text for details). The circles indicate unsmoothed means of the residuals—taken within bins of 1/3 percentage point—from a regression of the dependent variable on semester-by-instructor-by-grade-cutoff fixed effects. The solid lines are fitted values from linear regressions estimated on either side of the cutoff, applying triangular weights.

Figure 4: Student outcomes in grade cutoffs (Economics 102)



Note: The sample includes 1013 students who took Economics 102 between Fall 2004 and Spring 2014 (see text for details). The circles indicate unsmoothed means of the residuals—taken within bins of 1/3 percentage point—from a regression of the dependent variable on semester-by-instructor-by-grade-cutoff fixed effects. The solid lines are fitted values from linear regressions estimated on either side of the cutoff, applying triangular weights

Table A1: Pre-college variables at grade cutoffs (Economics 101 and 102)

	RD sample; Economics 101			RD sample; Economics 102		
	Bandwidth			Bandwidth		
	4 p.p.	MSE (1)	MSE (2)	4 p.p.	MSE (1)	MSE (2)
Preference for Econ major	-0.022	-0.026	-0.028	0.024	0.024	0.024
	(0.038)	(0.039)	(0.054)	(0.053)	(0.049)	(0.079)
N	1,466	1,446	1,141	1,013	1,037	722
Quantitative reasoning score	-0.213	-0.144	-0.117	0.079	0.098	0.140
	(0.219)	(0.276)	(0.350)	(0.259)	(0.283)	(0.457)
N	1,466	1,274	978	1,013	959	634
ACT	-0.042	0.313	0.835	0.014	0.294	0.423
	(0.451)	(0.535)	(0.749)	(0.629)	(0.786)	(1.040)
N	522	470	340	326	272	212
SAT (mathematics)	-8.390	-9.698	-1.018	2.429	5.545	3.706
	(6.236)	(7.090)	(10.052)	(7.304)	(11.533)	(14.636)
N	1,248	1,167	849	893	669	516
SAT (verbal)	4.623	16.035*	21.622*	-17.031*	-17.838**	-10.956
	(7.284)	(9.457)	(11.514)	(8.678)	(8.320)	(13.712)
N	1,248	1,041	784	893	910	615
Receives no financial aid	-0.033	-0.045	-0.072	0.004	0.012	0.028
	(0.036)	(0.038)	(0.051)	(0.040)	(0.043)	(0.056)
N	1,465	1,404	1,079	1,012	953	666
Admissions score	0.009	0.002	-0.003	-0.020	-0.012	-0.077
	(0.051)	(0.051)	(0.075)	(0.057)	(0.057)	(0.084)
N	1,466	1,504	1,116	1,013	1,023	685
Did not graduate	-0.014	-0.005	-0.003	0.028*	0.019	0.015
	(0.026)	(0.028)	(0.040)	(0.016)	(0.019)	(0.024)
N	1,466	1,421	1,142	1,013	896	654

Notes: Each cell reports an estimate of β from equation (2), applying triangular weights within the bandwidth specified in the column header (see text for an explanation of MSE-optimal bandwidths). Robust standard errors are in parentheses. *** indicates statistical significance at 1%; ** at 5%; * at 10%.

Table A2: Descriptive statistics (Computer Science 111)

	Ever took CS 111	
	All	RD sample
	Mean	Mean
Panel A: Pre-college variables		
Preference for Computer Science major (1/0)	0.07	0.07
Quantitative reasoning score	14.1	14.4
ACT	31.0	31.4
SAT (Mathematics)	718.8	727.8
SAT (Verbal)	705.9	704.7
Receives no financial aid	0.37	0.36
Admissions score	4.1	4.1
Panel B: College variables		
Grade point in 111	3.11	3.30
Grade point average in non-CS courses (in semester of 111)	3.31	3.37
Majored in Computer Science (1/0)	0.23	0.28
Cumulative grade point average	3.36	3.42
Did not graduate (1/0)	0.05	0.06
Panel C: Grade cutoff samples		
A/A- (1/0)		0.41
A-/B+ (1/0)		0.16
B+/B (1/0)		0.11
B/B- (1/0)		0.15
B-/C+ (1/0)		0.04
C+/C (1/0)		0.07
C/C- (1/0)		0.03
C-/D (1/0)		0.02
D/F (1/0)		0.01
Maximum number of observations	718	395

Notes: “Ever took” samples include the subset of students who took and received a grade in Computer Science 111. “RD samples” are further limited to students with complete data from instructor grading spreadsheets in 111.

Table A3: Student outcomes at letter-grade cutoffs in Computer Science 111

	Bandwidth for local linear regressions			
			MSE-optimal	
	4 percentage points		(1)	(2)
Grade point in CS 111 (2.88)	0.309***	0.308***	0.318***	0.310***
	(0.010)	(0.011)	(0.012)	(0.021)
N	342	342	287	209
G.P.A. in non-CS courses (semester of 111) (3.30)	0.041	0.020	0.116	0.122
	(0.075)	(0.076)	(0.087)	(0.115)
N	341	341	255	188
Majored in CS (0.22)	0.115	0.136	0.137	0.188
	(0.093)	(0.096)	(0.097)	(0.128)
N	323	323	307	260
Controls?	N	Y	Y	Y

Notes: Each cell reports an estimate of β from equation (2), applying triangular weights within the bandwidth specified in the column header (see text for an explanation of MSE-optimal bandwidths). Robust standard errors are in parentheses. *** indicates statistical significance at 1%; ** at 5%; * at 10%. The final row indicates regressions that contain the controls listed in panel A of Table A2, as well as dummy variables indicating missing values. In the first column, the number in parentheses is the mean of the dependent variable to the left of grade cutoffs (within 4 percentage points of the cutoff).