Ahead of the Curve

Grade Signals, Gender, and College Major Choice

Stephanie Owen

April 8, 2021

Abstract

Grades students receive are likely a key input into major choice, and prior observational work suggests the relationship varies by gender. I study a natural experiment within the economics department of a large university, which changed its grading policy to give out higher grades in its introductory economics courses. I leverage this variation to compare students with the same underlying performance but who received different letter grades. I find that receiving a higher grade in introductory economics increases the likelihood that a student will take the next course in the sequence by between 2 and 3 percentage points, with much smaller effects on economics major choice. Higher economics grades lead more students to declare a major in business—the highest-earning major at the studied institution. I find little evidence that women are more responsive to grades than men. My findings suggest that grade inflation as a policy may work to retain more students within a field, but is unlikely to close gender gaps. Given the complexity of the college major choice and the interdependence across subjects and courses, the optimal policy may not be straightforward.

1 Introduction

The choice of college major is a consequential at both the individual and aggregate level. There is wide dispersion in earnings by major, with the highest-earning making substantially more than the lowest-earning over their lifetime (Webber 2019). Since different types of students tend towards different types of majors, earnings differences by major have implications for income gaps by gender, race, and other characteristics. Furthermore, the allocation of students across majors and workers across jobs affects overall economic efficiency. If there are barriers or frictions keeping some individuals out of certain fields, then removing these barriers would lead to more efficient sorting and increase overall productivity (Hsieh et al. 2019). Finally, there is considerable policy interest in encouraging students to study certain fields, both to address shortages as well as to close gender and race gaps and address equity concerns.

Prior work has suggested that grades student receive are a key input into major choice, and that the relationship varies by gender (e.g., Avilova and Goldin 2020). However, much of this work is observational, and it is not clear how much the relationship between grades and behavior is about the effort, ability, and motivation behind the grade versus the signal the grade itself provides for students or the hurdles it may remove (psychological or administrative). If the grade itself matters and matters differently for women versus men, educators could use grades as a policy lever to induce students into certain fields and potentially close gender gaps. Doing this effectively requires knowing if and how grades affect choices conditional on endogenous inputs, as well as if and how the effect varies by gender.

To answer these questions, I study a natural experiment within the economics department of a large university, which changed the grading policy in its introductory economics courses to award higher grades. Specifically, the instructors went from a curve guaranteeing that 25 percent of students got a grade in the A range and 33 percent in the B range to guaranteeing at least 35 percent got some type of A and at least 75 percent got an A or B grade. This policy introduced variation in the letter grade students received, even conditional on their underlying effort and performance.

Rich administrative data provide information on students' final grades as well as their raw scores; I track their subsequent course-taking and academic decisions with full transcript data. I leverage the variation introduced by the change in the grading policy as well as naturally-occurring variation in letter grade cutoffs across semesters to compare students with the same underlying performance but who received different letter grades. This allows me to isolate the effect of the grade itself from confounding variables such as academic preparation and ability, and endogenous inputs such as interest and effort.

The results suggest that receiving one higher letter grade in introductory economics—for example, receiving a C over a C-, or an A- over a B+—increases the likelihood that a student will take the next course in the sequence by approximately 2.5 percentage points. These short term effects translate to small effects on later economics coursetaking and major declaration. In my preferred specification, the effect on taking the third course in the economics sequence is a statistically insignificant 0.6 percentage points, and the effect on declaring an economics major is 0.8 percentage points. I study not just economics course-taking and major choice but also behavior across subjects. In the setting I study, introductory economics is required for admission to the selective business school, and my findings imply that relatively higher economics grades may enable students to switch to related, competitive majors like business. A student who receives a higher grade in introductory economics is 2.3 percentage points more likely to declare a business major. Students who were induced by higher grades to major in economics or business appear to have switched from other social science majors. Business is the major with the highest post-graduate earnings at the studied institution, so these changes are likely welfare-improving for the affected students. I find little evidence that women are more (or less) responsive to grades than men. The estimated effects of grades on academic outcomes are very similar and statistically indistinguishable for men versus women.

My findings suggest that grade inflation as a policy may work to retain more students— including women—within a field, but will not necessarily close gender gaps. From the perspective of an academic department or institution, the optimal policy will depend on the objective (increasing the number of economics majors versus shifting students into majors with the highest earnings) and the structure of major requirements across departments (e.g., requiring economics courses for business majors).

The paper proceeds as follows. I summarize prior related work and my contribution in Section 2, introduce the setting and data in Section 3, and describe my empirical approach in Section 4. I present main results in Section 5 and robustness checks in Section 6. Section 7 concludes.

2 Conceptual Framework and Prior Work on the Effect of Grades

Theoretically, there are several reasons why we would expect letter grades to causally affect students' behavior, and in particular why the effects might differ between women and men. First, many academic programs, including college majors and scholarships, set thresholds for entry or continued eligibility. Receiving, say, a C over a C- in introductory economics could mechanically allow a student to major in economics or business—which set C as the minimum grade required for entry—over a field that doesn't require economics courses. As a more extreme example, a higher grade could allow a student to retain their financial aid and make the difference between persisting and dropping out of college altogether. Students also care about their GPAs for future plans such as graduate or professional school and entering the workforce; both employers and graduate institutions use college performance to evaluate candidates. For students trying to maximize their GPAs for any of the above reasons, the grade in an introductory course helps form expectations for their GPA if they major in that subject. All else equal, a lower grade means a lower expected GPA and could nudge them towards an easier-grading major.

These explanations could play out differently by gender for several reasons. If women and men are at different points in the grade distribution and students respond to certain grades, this could result in different observed behavioral responses, even if men and women respond similarly to a given grade. For example, if more women are on the margin between a C- and C grade and receiving a C is particularly consequential, we could see a stronger response for women. In a model of comparative advantage, a student's next best major option and their grades in that subject relative to economics will determine whether a higher economics grade pushes them over the margin into economics or a related major. If women's grades (or interest) in other subjects are much higher than in economics, a higher economics grade may not be enough to shift them into economics; this would manifest as women appearing less responsive to grades. Similarly, if men's non-economics GPAs are low enough, they could be inframarginal economics majors, resulting in their behavior changing less in reponse to higher economics grades.

Grades could also affect behavior through their signaling value. A higher letter grade may confer some positive utility or signal of ability to students above and beyond the other signals of their ability that they've observed. Much of the theoretical economics literature conceptualizes grades in this way, as signals by which students learn about their field-specific ability over time (Arcidiacono 2004; Altonji et al. 2016). The existence of stereotypes about groups, such as men being relatively better at quantitative subjects, could affect how much weight is given to signals and lead to different behavioral responses by gender (Bordalo et al. 2016). There are a number of empirical studies suggesting women and men may interpret performance feedback differently, though they do not always agree on whether men versus women respond more (see, e.g., Mobius et al. 2014 compared to Goulas and Megalokonomou 2015).

The empirical work on this topic consistently finds that grades are related to behavior, but there is a lack of strong causal evidence. A number of studies using selection-on-observables designs—i.e., comparing observationally similar students with different grades—find, unsurprisingly, that grades are associated with subsequent course-taking and major choice (Chizmar 2000; Jensen and Owen 2001; Rask and Bailey 2002; Rask and Tiefenthaler 2008; Ost 2010; Emerson et al. 2012; Astorne-Figari and Speer 2019; Kaganovich et al. 2020; Kugler et al. 2021). These studies come to different conclusions about whether women are more sensitive to grades than men, and whether easing grading standards in traditionally male-dominated fields could be an effective policy to close gender gaps. Rask and Bailey (2002), Rask and Tiefenthaler (2008), and Ost (2010) find support for the hypothesis that women respond more than men to grade signals in early courses. Two recent papers by Kaganovich et al. (2020) and Kugler et al. (2021) find that women are more likely to leave male-dominated STEM fields (including economics) in response to lower grades, but don't find different gender responses in other fields. Kugler et al. (2021) interpret this as women needing multiple negative signals (low grades, the presence of few other women, and stereotypes about whether a field is male or female) to leave a major at a higher rate than men. Kaganovich et al. (2020) argue that what appears to be greater grade sensitivity actually reflects a weaker underlying preference for those fields. However, Astorne-Figari and Speer (2019) and Chizmar (2000) use a similar approach and data but find no evidence that women are more likely to switch majors in response to low grades.

Studies that take a more structural approach similarly find that, consistent with a framework of learning about field-specific ability through grades, students who perform worse are more likely to switch majors (Arcidiacono 2004; Zafar 2011; Stinebrickner and Stinebrickner 2014). Several of these examine differences by gender, and while Calkins (2020) suggests that women respond more to grades and improving women's grades could close gender gaps in STEM, Zafar (2013) finds that differences in preferences rather than in beliefs about ability are responsible for most of the gender gap in major choice.

However, none of these studies have any exogenous variation in grades, and do not have finer measuers of underlying performance than letter grade or overall GPA. In this study, I exploit plausibly exogenous variation in grades, and furthermore am able to much more precisely control for students' underlying effort, motivation, and preparation, to the extent they are reflected in raw course performance.

More closely related are a handful of studies exploiting plausibly exogenous variation in letter grades. Owen (2010) and Main and Ost (2014) both use regression discontinuity designs, controlling for raw score and comparing students above and below the cutoffs for letter grades in introductory economics courses. Though they use similar approaches and study similar settings, these two studies come to different conclusions. Main and Ost (2014) find no effect of receiving a higher letter grade on subsequent coursetaking or major choice for any students, and no evidence of different responses by gender. Owen (2010), on the other hand, finds that receiving a higher grade in introductory economics increases the probability of majoring in economics for women but not for men. Though these RD designs are conceptually similar to my approach, they rely on a different identifying assumption and exploit a different source of variation. The validity of the RD requires that within a course, students who end up with slightly different raw scores and therefore grades are not different in other ways that could affect their outcomes. If a student has a slightly higher score as a result of a targeted effort to achieve their desired grade, or successful advocacy for a re-grade, both of which may reflect a higher interest in the subject, this assumption would be violated. My approach, on the other hand, compares students across courses and semesters rather than within, and uses an external change to the way grades are assigned. Crucially, any manipulation in underlying scores is already accounted for in the measure of raw score that I use, and does not threaten my idenfitication strategy.

Finally, Butcher et al. (2014) use a type of policy variation similar to the current study to examine the effect of grade inflation on major choice. They exploit an anti-grade inflation policy that affected different academic departments differently. They compare previously lenient-grading departments to harder-grading departments unaffected by the new rule (including economics) and find that the policy decreased the rate of students enrolling and majoring in departments that saw grade deflation relative to those that didn't. However, the setting of that paper—Wellesley College, an all-women's institution—does not allow the authors to say anything about differential response by gender. The current paper provides the most convincing causal evidence to date on the effect of letter grades on students' academic choices, and how those effects vary by gender.

3 Setting, Policy Background, and Data

I study a large, selective, public flagship university in the Midwest, which I will refer to as Midwestern University or MU. In the fall of 2016, the economics department at MU changed the grading curve in its introductory courses to give out more grades in the A and B ranges. Prior to the change, instructors in the Principles of Economics courses guaranteed that 25 percent of students received a grade in the A range (A-, A, or A+) and 33 percent received a grade in the B range. After the change, at least 35 percent of students were guaranteed some type of A and at least 75 percent an A or B. This change reflected concern among economics faculty that the department was not keeping up with grade inflation across the university, and that the harsher grading was deterring students from enrolling and persisting in its courses.[1] This change was not an official one voted on by the full department; rather, the instructors teaching Principles collectively agreed to give out higher grades. There was no enforcement from above and there were no official sanctions for not complying. At MU, economics instructors have considerable independence in teaching their sections. They write their own assignments and exams, weight assignments how they like, and are not required to use the same textbook. Although the Principles instructors agreed to this new common grading curve, they had discretion over their own students' grades and implementation of the policy. There was no official announcement about the policy change, and, according to instructors, students would not have known about it when registering for fall 2016 classes. Some of the instructors announced it to their students at the beginning of the fall 2016 term, while others did not bring it to students' attention.

[1]Source: internal department memo comparing undergraduate grading policies in economics to other fields.

Though I focus on Principles I - Microeconomics, the first course in the economics sequence, the policy also extended to Principles II - Macroeconomics, the second course in the sequence.[2] Instructors in Principles II adjusted their curve one semester later than Principles I (spring 2017 vs. fall 2016) so that the change occurred starting with a single cohort. In this sense, the policy can be thought of as inflating the grades for both of the first two core courses in the department and the major.

I combine two sources of administrative data to leverage this policy change and study the effect of grades on student behavior. Learning management system (LMS) data allow me to measure student performance in Principles I. Crucially, LMS data contain students' continuous raw scores on assignments and in the course overall before letter grades are assigned. I merge these data with university student record data, which include individual-level academic and demographic characteristics (standardized test scores, high school GPA, gender, race, parental education, family income, etc.), as well as full longitudinal academic transcripts (official letter grades, courses taken, and declared major).

Since each instructor manages their own LMS page, the structure of LMS data varies across instructors and sections. Between 2014 and 2016, MU transitioned from one LMS to another, so the structure also varies over time. My empirical approach requires a measure of students' final total score in Principles I. In some cases, instructors entered a final score into the LMS, and little additional cleaning was required. In other cases, I constructed final scores based on individual assignment scores and the weighting of assignments detailed in course syllabuses. There is likely a non-trivial amount of measurement error arising from this process. I discuss this issue more in Section 6.2, and show that it is not substantively altering my conclusions.

Because I am studying the effect of letter grades, I limit the sample to students who complete the class and receive a letter grade, which excludes students who elect to take the course Pass/Fail. I also restrict to students with observations in the LMS

[2]Principles I is an advisory prerequisite for Principles II, meaning it is highly recommended but not strictly required. In practice, only around one percent of students take Principles II without having first taken Principles I.

data. If a student repeated the course, I use their first observation. The final sample includes 11,836 students, covering students who took Principles of Economics I for the first time between Spring 2013 and Spring 2018 (inclusive).[3] The dataset includes eight unique instructors, 11 academic terms, and 49 lecture sections.[4] Of the eight instructors, five taught both before and after the grading curve changed.

4 Empirical Strategy

My empirical strategy compares students with the same underlying economics course performance (measured as their final percentile rank within their instructor's section of Principles I) and observable characteristics, but who receive different letter final grades. By holding underlying performance constant, this strategy controls for all of the inputs that determine grades—such as effort, motivation, and prior academic preparation—and could also affect subsequent academic outcomes.

This approach uses two types of plausibly exogenous variation in letter grades. The first is variation introduced by the policy change, which increased a student's expected letter grade conditional on their raw score or percentile. Consider a simple example where an instructor strictly implemented the curve. Under the old grading regime 25 percent of students were guaranteed an A grade, while under the new regime at least 35 percent were. Comparing two students who performed at the 74th percentile (right below the old cutoff), the student who took the course before the change would receive a B+, while the student who took the course after would receive (at least) an A-.

The second source of variation is naturally occurring variation in letter grades across semesters under the same official grading curve policy. Because the instructors have ultimate discretion in assigning letter grades and cannot perfectly control the composition of students in their courses or the difficulty of their exams and assignments,

[3] I include fall and spring courses only. MU does offer Principles I during its summer term, but the courses have much smaller enrollments and are structured differently.

[4] This is not the universe of Principles I course offerings during this time, which included 61 lecture sections. A handful of instructors did not have archived LMS data available.

the same instructor may give out more A's in one semester than another. The new grading policy, which states that *at least* 35 percent of students receive A's, explicitly allows for different grade distributions. Under both policies, there is considerable variation in how many A's and B's instructors awarded. (I show evidence of this in Section 5.2.)

Formally, I estimate the following equation:

$$\begin{aligned} Y_{ijt} = \beta_0 + \beta_1 Grade_{ijt} + \beta_2 Percentile_{ijt} + \beta_3 Percentile_{ijt}^2 + \beta_4 Percentile_{ijt}^3 \\ + \sum_j \alpha_j + \delta Fall_t + \lambda Year_t + \boldsymbol{\gamma} \boldsymbol{X_i'} + \epsilon_{ijt} \end{aligned} \tag{1}$$

where i indexes students, j instructors, and t time periods (academic terms). The term β_1 is the estimand of interest and represents the effect of receiving a higher letter grade in Principles of Economics I. In my main specification, I estimate a constant linear effect of a higher grade; a one unit increase in letter grade is equivalent to receiving an A- over a B+, or a C over a C-. I also estimate a version of Equation 1 where I replace the $Grade$ term with indicators for each letter grade to separately identify the effect of an A, A-, etc. Since I largely lack the power to compare effects by each grade, I prefer the linear specification.

In controlling for underlying performance, I calculate percentiles within academic term and lecture section, since this is the level at which grades are curved. (Note I calculate percentiles before excluding any students from the sample, such as those who took the course pass/fail or were taking it for a second time.) I control for a cubic in percentile to allow for a flexible relationship between percentile and the outcome.

The vector $\boldsymbol{\alpha}$ represents instructor fixed effects, and δ captures seasonality effects (the absolute performance thresholds tend to be higher in spring terms, when more engineering majors take Principles I). I include a linear time trend λ (where $Year$ denotes academic year) to allow for upward (or downward) trends in the outcomes. The vector $\boldsymbol{X_i}$ includes student gender, race/ethnicity (indicators for Black; Hispanic;

Asian; Native American, Native Hawaiian, or other Pacific Islander; and multiple races), class standing (indicators for second, third, and fourth and higher year), family income (indicators for $25,000-$49,999, $50,000-$74,999, $75,000-$99,999, and $100,000 and above), parent education (indicators for high school, some college, bachelor's degree, and graduate degree), high school GPA, an indicator for taking calculus in high school, SAT and ACT subscore percentiles, and score on the university's math placement test. Some of these characteristics are self-reported or not collected for all students, so I also includes missingness indicators for background characteristics with any missing values.

I estimate the effect of receiving a higher letter grade in Principles of Economics I on three measures of persistence within economics: indicators for taking the second course in the sequence (Principles of Economics II - Macroeconomics), taking the third course (Intermediate Microeconomic Theory), and declaring an economics major; I measure all of these within two years of completing Principles I. The outcomes are all indicators for *ever* doing the outcome. For coursetaking, this simply means the student took the course at some point in the two years following Principles I. For major choice, a student gets counted as an economics major as long as they appear as an economics major in any of the subsequent semesters (even if they double major or later switch to a different major). I also study effects on major choice beyond economics, by measuring declaration of a business major, a STEM major, or a non-economics social science major. I classify subjects using two-digit Classification of Instructional Program (CIP) codes, developed and maintained by the U.S. Department of Education's National Center for Education Statistics.[5] All effects are estimated with a linear probability model. I report robust standard errors calculated with the sandwich estimator of variance.

[5] STEM includes natural resources and conservation (CIP code 03), computer and information sciences (11) engineering (14), biological and biomedical sciences (26), mathematics and statistics (27), and physical sciences (40). Social sciences (CIP code 45, excluding 45.06 - Economics) includes anthropology, political science, and sociology. Business majors (CIP code 52) include business administration and organizational studies.

4.1 Possible threats to identification

The identifying assumption required to interpret β_1 causally is that conditional on instructor, observable characteristics, and percentile rank in the course, the final letter grade is orthogonal to the error term. The thought experiment takes two observably similar students who take Principles I with the same instructor at the same time of year (fall or spring) and perform equally well in the class, and assigns one a higher letter grade than the other (e.g, an A- rather than a B+). The primary research question is whether receiving the higher grade makes students more likely to persist in economics or changes their academic trajectory in some way.

This assumption would be violated if letter grades are not exogenous conditional on performance. For example, if certain students are able to advocate for higher grades, this could result in students with the same performance being assigned different grades. While re-grades do happen, this is only an issue for my identification strategy if the instructors change the final grade but not the underlying score. From conversations with instructors, this is exceedingly rare. Any changes tend to happen at the individual assignment level, and are entered in gradebooks. Furthermore, instructors require students to go through an appeals process and rarely if ever grant end-of-semester requests to bump up a grade close to the margin. To test for this type of selection into letter grades more formally, I examine whether, conditional on underlying performance, instructor, academic year, and time of year, letter grade predicts observable characteristics such as gender, race, family background, and academic preparation. While I control for all of these observable characteristics in my analyses, significant effects on these falsification tests could indicate differences in unobservable characteristics that could be determining both grades and outcomes. While I do find some differences (see Section 6.4), they are substantively small and unlikely to be responsible for effects of the magnitude that I find.

Another possible threat to identification would be another policy change contemporaneous to the grade curve change which could also affect persistence in economics and major choice. There was a substantial change to the admissions policy of the institution's business school around the same time, which could be confounding the effect of the grade policy change. I discuss this in more detail in Section 6.3 and argue that it is not substantively biasing my results.

5 Results

5.1 Descriptive sample statistics

Table 1 presents mean characteristics for the sample, both overall and by gender. 39 percent of students who took Principles of Economics I during the sample period are women. Nearly two thirds of students are white, and another quarter are Asian. Very few underrepresented minority students are in the sample: 3 percent are Black, 5 percent are Hispanic, less than 1 percent are Native American, Native Hawaiian, or other Pacific Islander, and 3 percent are more than one race. The majority (77 percent) of students who take Principles I do so for the first time in their first year of college, while another 18 percent took the course in their second year; fewer than 6 percent took the course in their third year or later.

Students at MU come from very socioeconomically advantaged backgrounds. The majority of students (53 percent) have family income in the highest category of $100,000 and above; conditional on having a reported family income, over two thirds are in this category. Similarly, 58 percent of students have a parent with a graduate or professional degree, and only 8 percent are first-generation college students (meaning neither parent has a bachelor's degree or higher). MU is considered a highly competitive institution, and this is reflected in the academic background of students. The average high school GPA is 3.82 on a 4.0 scale, 72 percent took calculus in high school, and they performed at the 70th percentile on the quantitative section of the SAT or ACT, on average.

In general, the female and male students in the sample are similar in their mean characteristics. The women in the sample are less likely than men to be White (63 vs.

66 percent) and more likely to be Black (3 vs. 2 percent), though the differences are small. Men are slightly more likely to take the course in their first year (77 vs. 76 percent) and third year (4 vs. 3 percent), while women are more likely to take it in their second year (20 vs. 17 percent). Female economics students seem to have lower family incomes (more likely to be in the lowest category and less likely to be in the highest category), but they are also more likely to not have reported income. Notably, women have higher high school GPAs (3.84 vs. 3.81), but men are slightly more likely to have taken calculus (73 vs. 71 percent) and have higher standardized quantitative test scores (74th vs. 65th percentile).

5.2 Evidence of policy change

I first present descriptive evidence that the stated changes to the grading curve in Principles of Economics I courses did in fact change the distribution of grades instructors awarded. Panel (a) of Figure 1 shows the distribution of grades students received by whether they took the course before the curve changed or after, with Fall 2016 the first semester under the post regime. From Spring 2013 to Spring 2016, 31 percent of grades awarded were in the A range (A-, A, or A+); from Fall 2016 to Spring 2018, 42 percent of grades were some type of A. After the curve changed, instructors gave fewer E, D, or C grades (31 percent pre vs. 19 percent post) and more B and A grades (69 vs. 81 percent).

Panel (b) of Figure 1 shows variation in grades awarded at the lecture section level, where a section is a unique instructor, term, and course catalog number. (Most instructors teach one section a semester, and some teach two.) This figure plots the distribution of the proportion of students in a section receiving A grades. Even under the same official curve, there is variation in how many A's instructors award, but the distribution clearly shifts right under the new policy. In the pre-period, the average proportion of A grades was 30 percent, but the proportion ranged from 25 to 52. In the post period, the average proportion of A grades was 41 percent, with a range of 32 to 57. Note that both panels of Figure 1 suggest that under both policies, instructors were

Table 1: Sample Descriptive Statistics

	All	Women (W)	Men (M)	p-value, W vs. M
Female	0.39	1.00	0.00	
White	0.65	0.63	0.66	0.00
Asian	0.19	0.20	0.19	0.15
Black	0.03	0.03	0.02	0.00
Hispanic	0.05	0.05	0.05	0.87
Native American, Hawaiian, or Pacific Islander	0.00	0.00	0.00	0.72
Multiple races	0.03	0.03	0.03	0.59
Race/ethnicity missing	0.05	0.05	0.05	0.97
First year at MU	0.77	0.76	0.77	0.05
Second year	0.18	0.20	0.17	0.00
Third year	0.04	0.03	0.04	0.00
Fourth+ year	0.02	0.02	0.02	0.90
Family income less than $25,000	0.03	0.03	0.03	0.02
$25,000-$49,999	0.05	0.05	0.05	0.13
$50,000-$74,999	0.05	0.05	0.05	0.51
$75,000-$99,999	0.07	0.07	0.07	0.92
$100,000 and above	0.53	0.51	0.54	0.00
Family income missing	0.23	0.25	0.22	0.00
Max parent ed less than high school	0.01	0.01	0.01	0.12
High school	0.03	0.04	0.03	0.21
Some college	0.04	0.03	0.04	0.46
Bachelor's degree	0.25	0.25	0.25	0.69
Graduate or professional degree	0.58	0.59	0.58	0.47
Parent education missing	0.09	0.08	0.09	0.13
High school GPA	3.82	3.84	3.81	0.00
HS GPA missing	0.11	0.10	0.12	0.00
Took calculus in high school	0.72	0.71	0.73	0.03
SAT or ACT math percentile	70.41	65.22	73.77	0.00
Missing test score	0.13	0.12	0.14	0.00
N	11,836	4,592	7,244	

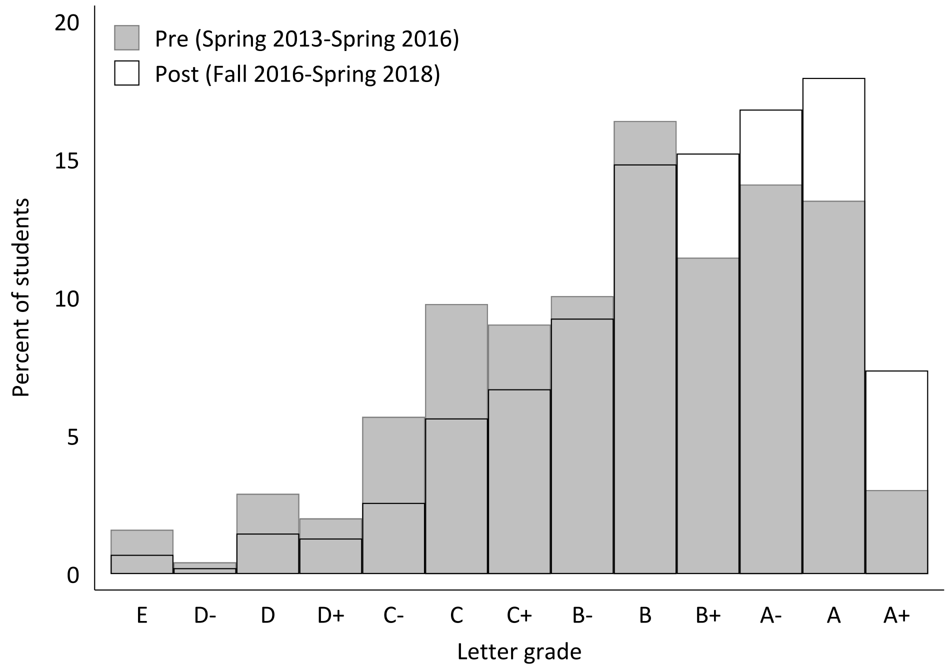
somewhat more generous than what was written in their syllabuses, which guaranteed 25 percent of students some type of A in the pre-period and at least 35 percent in the post period.

5.3 Causal effect of higher letter grades

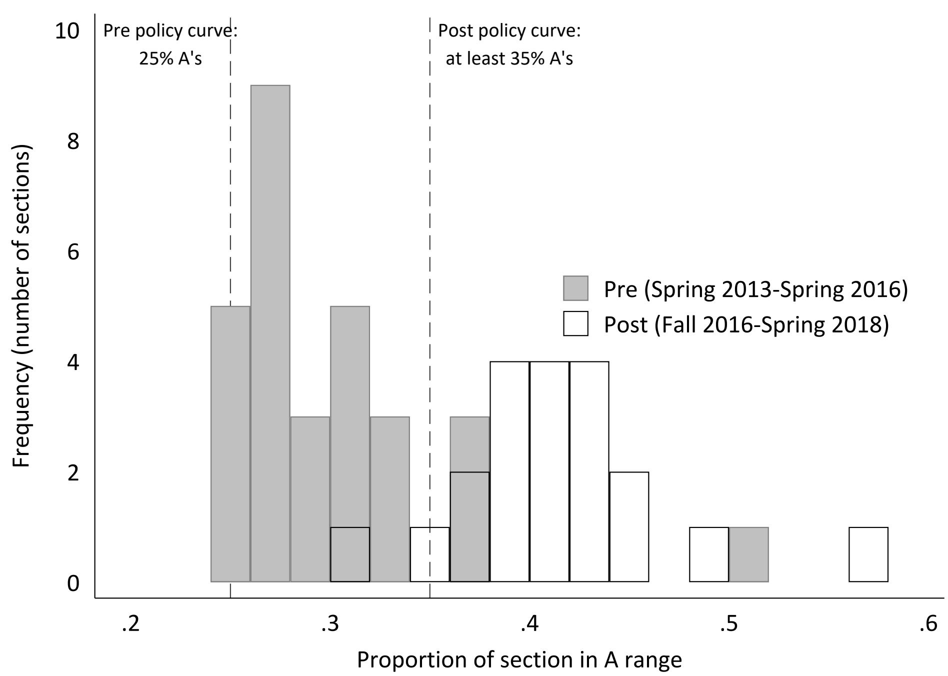
Table 2 presents the main findings, the estimated effect of receiving a higher letter grade in Principles I on subsequent economics coursetaking and major declaration. The six outcome variables are indicators for whether a student took Principles of Economics II (the second course in the economics sequence), took Intermediate Microeconomic Theory (the third course in the sequence), or declared an economics, business, STEM, or non-economics social science major, all measured as ever doing so in the two years after they took Principles I. The "treatment" of a higher letter grade is for one higher grade on a scale with pluses and minuses. For example, an A- is one grade higher than a B+.

The first column of Table 2 shows effects estimated on the full sample of students who took Principles of Economics I. I find that receiving a higher letter grade in Principles I makes students 2.5 percentage points more likely ($p<0.05$) to take Principles II, the next course in the sequence. The effect on taking Intermediate Micro Theory is a statistically insignificant 0.6 percentage points; the effect on declaring an economics major is similar in magnitude at 0.8 percentage points and marginally statistically significant ($p<0.1$). In terms of non-economics major outcomes, it appears that higher grades in introductory economics make students more likely to major in business, by 2.3 percentage points ($p<0.01$). At MU, taking Principles I and II in the economics department is required for business majors. Paired with the fact that the business major is considered more selective and is associated with higher earnings, it is not surprising that awarding higher economics grades increases the rate of students majoring in business more than the rate majoring in economics. I detect no change in the rate of STEM majoring (a statistically insignificant -0.3 percentage points). The increase in economics and business majors may correspond to a decrease in social

Figure 1: Grade Distributions Pre- and Post-Grading Curve Policy Change



(a) Distribution of letter grades (student level)



(b) Distribution of proportion of A range grades awarded (section level) science majors by 0.7 percentage points ($p<0.1$). Since the way I identify majors is not mutually exclusive, the remainder of the increase in business and economics could correspond to a decrease in all other majors such as humanities, arts, and communications, or to an increase in the rate of double-majoring. In a separate analysis (not shown) I find no increase in the likelihood of having two or more declared majors, implying students are not adding economics or business as second majors but rather shifting from other fields.

The final three columns of Table 2 show effects estimated separately for women and men and a test for equality between the groups. I find no evidence of heterogeneity by gender in the response to higher grades. The point estimates for women and men are consistently very similar, and I cannot reject the hypothesis that they are equal for any of the outcomes. For example, a higher grade makes women 2.7 percentage points more likely to take Principles II, and men 2.4 percentage points more likely (p-value for difference: 0.806). The effect on declaring a business major is 2 percentage points for women and 2.6 percentage points for men (p-value for difference: 0.517).

5.4 Effect of higher letter grade, by grade

For statistical power reasons, my preferred specification estimates a constant linear effect of receiving a higher letter grade. However, the marginal effect of receiving, say, an A- over a B+ could be different than receiving a C over a C-. This could be true for multiple reasons. Perhaps students psychologically value grades in the A-range, or employers only care about grades above a certain threshold. Both the economics department and the business school require a C grade or higher in Principles I, so the marginal effect of a C might be particularly salient. To investigate this, I estimate a specification similar to Equation 1, but with indicators for each grade (A+, A, A-, B+, B, B-, C+, and C, with C- or below the omitted category) rather than a single grade variable. I present estimated effects of each grade relative to the grade just below it, since the "treatment" can be thought of as increasing a student's grade on the margin (and this is the analog of the linear effect). For example, the effect of receiving an A-

Table 2: Estimated Effect of Higher Letter Grade in Introductory Economics

Effect of higher grade on:	All	Women (W)	Men (M)	p-value, W vs. M
Took Principles II	.025***	.027***	.024***	0.806
	(0.006)	(0.010)	(0.008)	
Took Interm. Micro	.006	.005	.006	0.839
	(0.005)	(0.007)	(0.007)	
Declared Econ Major	.008*	.009	.007	0.866
	(0.004)	(0.006)	(0.006)	
Declared Business Major	.023***	.02**	.026***	0.517
	(0.005)	(0.008)	(0.006)	
Declared STEM Major	-.003	.002	-.005	0.504
	(0.006)	(0.008)	(0.007)	
Declared Social Science Major	-.007*	-.007	-.007	1.000
	(0.004)	(0.007)	(0.005)	
N	11,836	4,592	7,244	

Notes: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Robust standard errors reported. The results in each column come from a regression of the outcome on a linear term for letter grade. The effects are of one higher letter grade, e.g. going from a B+ to an A- or a C to a C+. Regressions also control for a cubic in percentile rank in the course, gender, race, class standing, family income, parental education, SAT and ACT scores, high school GPA, taking calculus in high school, score on the university's math placement test, instructor, academic year, and season of course (fall vs. spring term). Effects for women and men are estimated in separate regressions. All outcomes are measured in the two years following the term the student took Principles of Economics I. Major categories based on 2-digit CIP codes.

is relative to a B+, and is calculated as the coefficient on A- minus the coefficient on B+.

Table 3 shows the effect of each letter grade (relative to the grade below) on each of the six outcomes, for all students and separately by gender. For economics coursetaking, the largest effects of grades on taking Principles II are of receiving an A- (4.6 percentage points, $p<0.05$), a B (4.4 percentage points, $p<0.05$), and a C (5 percentage points, $p<0.05$). I find a marginally significant effect of receiving a C on taking Intermediate Micro (2.8 percentage points, $p<0.1$) and a significant effect of a C on declaring an economics major (2.7 percentage points, $p<0.05$). The effect of a B on majoring in economics is a marginally signficant 2.5 percentage points. Though the comparison of magnitudes is consistent with students particularly valuing A range grades and needing a C to meet major requirements, the confidence intervals around the estimates are wide enough that I can't make precise comparisons.

Turning to the effects on declaring a business major, the largest and statistically significant effects appear for relatively lower letter grades (C, B-, and B, all with effects of around 4 percentage points). I find no effect of any grade on majoring in STEM, and the only significant grade effect on majoring in social science is of a B- (-2.9 percentage points, $p<0.05$).

In terms of gender differences, I find few significantly different effects by gender and no clear pattern. Furthermore, I am conducting many hypothesis tests. The tests for heterogeneity by gender imply that women react more positively to an A+ grade than men in terms of taking Intermediate Micro (8.3 vs. 0.2 percentage points, p-value for difference: 0.095), while men increase their probabilities of taking intermediate micro and declaring an economics major more in response to a B- (5.9 vs. -2.9 percentage points, p-value for difference: 0.006). The effects of an A (4.6 percentage points for women, -2.1 for men, p-value for difference: 0.041) and a B (5.3 for women, 0.5 for men, p-value for difference: 0.095) on declaration of an economics major are more positive for women, but the opposite is true for a B- grade (-1.1 for women, 4.4 for men, p-value for difference: 0.066). Panel B of Table 3 suggests that grades of A and A- may induce more men into business relative to women. Consistent with the results in Table 2, I interpret all of this as providing no strong evidence that women are particularly responsive to grades.

6 Alternative Specifications and Robustness

6.1 Measuring performance as percentage points vs. percentile

Thus far, in controlling for students' raw performance in Principles I, I have used a measure of their percentile rank within their course. The advantage of operationalizing performance as percentile rank is that it has the same meaning and distribution across classes and explicitly maps to grade curve policies, which specify a certain percent of students to receive certain grades. Because instructors vary in the raw scores they give out, students with the same raw score in different courses might fall in very different parts of their course distribution—and therefore the grade distribution. However, one argument against using percentile rank is that students generally do not observe this measure; rather, they observe their total points or percent score. Conceptually, controlling for percent better captures the signal value of the letter grade over and above the information contained in their raw score. In practice, percentile rank and percentage score are monotonic transformations between each other within a course. (This is not necessarily true across courses, but by including instructor fixed effects I account for different grading norms across instructors.)

Table 3: Estimated Effects of Letter Grades in Introductory Economics, with Separate Effects by Grade

Panel A: Economics Coursetaking and Major Outcomes

	Took Principles II				Took Interm. Micro				Declared Econ Major			
	All	Women (W)	Men (M)	p-value, W vs. M	All	Women (W)	Men (M)	p-value, W vs. M	All	Women (W)	Men (M)	p-value, W vs. M
(A+)	.017	.035	.012	0.683	.022	.083**	.002	0.095	-.003	.034	-.015	0.251
	(0.025)	(0.048)	(0.030)		(0.021)	(0.042)	(0.025)		(0.018)	(0.037)	(0.021)	
(A)	-.02	-.049	-.007	0.334	0	.036	-.015	0.171	0	.046*	-.021	0.041
	(0.020)	(0.037)	(0.024)		(0.017)	(0.031)	(0.020)		(0.015)	(0.027)	(0.018)	
(A-)	.046**	.04	.047**	0.851	.02	.052**	0	0.100	.013	.037*	-.002	0.179
	(0.018)	(0.031)	(0.023)		(0.016)	(0.024)	(0.020)		(0.014)	(0.022)	(0.018)	
(B+)	.011	.028	-.002	0.432	-.024	-.034	-.02	0.654	-.015	-.034	-.003	0.282
	(0.018)	(0.030)	(0.023)		(0.016)	(0.024)	(0.021)		(0.014)	(0.022)	(0.019)	
(B)	.044**	.048	.043*	0.898	.014	.042**	-.007	0.114	.025*	.053***	.005	0.095
	(0.020)	(0.031)	(0.026)		(0.016)	(0.021)	(0.023)		(0.015)	(0.020)	(0.021)	
(B-)	.032	.047	.018	0.507	.022	-.029	.059**	0.006	.02	-.011	.044*	0.066
	(0.022)	(0.033)	(0.029)		(0.017)	(0.021)	(0.024)		(0.015)	(0.019)	(0.023)	
(C+)	-.002	-.027	.02	0.314	-.003	-.002	0	0.942	.015	.011	.023	0.670
	(0.023)	(0.034)	(0.032)		(0.017)	(0.022)	(0.025)		(0.015)	(0.019)	(0.022)	
(C)	.05**	.06*	.041	0.690	.028*	.008	.043*	0.268	.027**	.017	.036*	0.468
	(0.024)	(0.034)	(0.034)		(0.016)	(0.020)	(0.025)		(0.013)	(0.016)	(0.021)	

Notes: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors reported. The results in each column come from a regression of the outcome on indicators for each letter grade, with C- or below as the omitted category. The effect of each grade is relative to the grade below it. For example, the effect of an A+ is relative to an A, calculated by subtracting the coefficient on A from the coefficient on A+. Regressions also control for a cubic in percentile rank in the course, gender, race, class standing, family income, parental education, SAT and ACT scores, high school GPA, taking calculus in high school, score on the university's math placement test, instructor, academic year, and season of course (fall vs. spring term). Effects for women (N=4,592) and men (N=7,244) are estimated in separate regressions. Total N=11,836. All outcomes are measured in the two years following the term the student took Principles of Economics I.

Table 3: Estimated Effects of Higher Letter Grades in Introductory Economics, with Separate Effects by Grade
Panel B: Non-Economics Major Choice Outcomes

	Declared Business Major				Declared STEM Major				Declared Social Science Major			
	All	Women (W)	Men (M)	p-value, W vs. M	All	Women (W)	Men (M)	p-value, W vs. M	All	Women (W)	Men (M)	p-value, W vs. M
(A+)	-.019	-.073	-.001	0.192	-.026	.022	-.049*	0.191	-.011	-.009	-.012	0.855
	(0.024)	(0.048)	(0.027)		(0.025)	(0.046)	(0.029)		(0.008)	(0.016)	(0.009)	
(A)	-.001	-.053	.018	0.089	-.014	.003	-.021	0.552	0	-.003	.002	0.775
	(0.019)	(0.036)	(0.022)		(0.019)	(0.033)	(0.022)		(0.007)	(0.015)	(0.008)	
(A-)	.026	-.041	.06***	0.004	-.021	.006	-.031	0.293	.005	.017	-.001	0.316
	(0.017)	(0.030)	(0.020)		(0.017)	(0.028)	(0.021)		(0.008)	(0.015)	(0.009)	
(B+)	.029*	.029	.03	0.986	-.001	.014	-.006	0.548	-.007	.005	-.016	0.276
	(0.015)	(0.027)	(0.018)		(0.016)	(0.026)	(0.021)		(0.009)	(0.017)	(0.010)	
(B)	.041***	.038	.044**	0.846	.006	-.021	.024	0.182	-.008	.005	-.017	0.352
	(0.015)	(0.026)	(0.018)		(0.017)	(0.025)	(0.023)		(0.011)	(0.020)	(0.014)	
(B-)	.041***	.059**	.026	0.241	-.015	-.019	-.009	0.810	-.029**	-.022	-.036*	0.638
	(0.014)	(0.023)	(0.017)		(0.019)	(0.027)	(0.026)		(0.015)	(0.023)	(0.019)	
(C+)	.008	.002	.013	0.696	.005	.008	.005	0.940	.001	-.018	.016	0.319
	(0.014)	(0.022)	(0.018)		(0.020)	(0.027)	(0.029)		(0.017)	(0.025)	(0.022)	
(C)	.045***	.067***	.035**	0.223	-.008	-.009	-.008	0.989	-.009	-.019	-.001	0.628
	(0.013)	(0.021)	(0.017)		(0.021)	(0.029)	(0.030)		(0.019)	(0.029)	(0.024)	

Notes: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Robust standard errors reported. The results in each column come from a regression of the outcome on indicators for each letter grade, with C- or below as the omitted category. The effect of each grade is relative to the grade below it. For example, the effect of an A+ is relative to an A, calculated by subtracting the coefficient on A from the coefficient on A+. Regressions also control for a cubic in percentile rank in the course, gender, race, class standing, family income, parental education, SAT and ACT scores, high school GPA, taking calculus in high school, score on the university's math placement test, instructor, academic year, and season of course (fall vs. spring term). Effects for women (N=4,592) and men (N=7,244) are estimated in separate regressions. Total N=11,836. All outcomes are measured in the two years following the term the student took Principles of Economics I. Major categories based on 2-digit CIP codes.

To test the sensitivity of my results to using percentile rank versus percent score, I present results controlling for the latter in Table 4. These results are equivalent to Table 2 and estimate Equation 1, except with percent score (out of 100) rather than percentile. The estimated effects of higher grades are qualitatively similar across the two specifications, but conditioning on percent score rather than percentile results in somewhat larger point estimates and smaller standard errors. The first column of results finds that receiving a higher grade in Principles I makes students 4 percentage points more likely to take Principles II, 2.3 percentage points more likely to take Intermediate Micro, and 2.1 percentage points more likely to major in economics, all significant at the $\alpha = 0.01$ level. Higher grades in introductory economics make students 3.2 percentage points more likely to major in business and 1 point less likely to declare a non-economics social science major.

In this specification, I find some evidence of heterogeneity by gender, but in a way that suggests men are more likely to continue in economics as a result of a higher grade. The effect of a higher grade on taking Intermediate Micro is 2.9 percentage points for men and an insignificant 1 percentage point for women (p-value for difference: 0.013). Similarly, the effect on declaring an economics major is 2.5 points for men and an insignficant 0.9 points for women (p-value for difference: 0.018). Tables 2 and 4 together suggest that, at best, higher grades do not close gender gaps in economics persistence, and at worst they may exacerbate gaps.

6.2 Dropping sections with grade rank inconsistencies

My identification strategy hinges on controlling for students' underlying course performance, which I argue captures the characteristics and inputs that could affect both letter grade and academic choices. This requires a measure of students' final total score in Principles I, which comes from what instructors have entered in learning management system gradebook data. To calculate final score and final percentile, in some cases I could use a "final score" grade entered by instructors with no additional cleaning. In other cases, there was no final score, so I constructed final scores based on individual assignment scores and the weighting of assignments detailed in course syllabuses.

Table 4: Estimated Effect of Higher Letter Grade in Introductory Economics, Controlling for Percent Score Rather than Percentile

Effect of higher grade on:	All	Women (W)	Men (M)	p-value, W vs. M
Took Principles II	.04***	.038***	.039***	0.882
	(0.005)	(0.008)	(0.006)	
Took Interm. Micro	.023***	.01	.029***	0.013
	(0.004)	(0.006)	(0.005)	
Declared Econ Major	.021***	.009	.025***	0.018
	(0.003)	(0.005)	(0.005)	
Declared Business Major	.032***	.041***	.026***	0.059
	(0.004)	(0.006)	(0.004)	
Declared STEM Major	.003	.007	.003	0.657
	(0.004)	(0.007)	(0.006)	
Declared Social Science Major	-.01***	-.008	-.011***	0.686
	(0.003)	(0.005)	(0.003)	
N	11,836	4,592	7,244	

Notes: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Robust standard errors reported. The results in each column come from a regression of the outcome on a linear term for letter grade. The effects are of one higher letter grade, e.g. going from a B+ to an A- or a C to a C+. Regressions also control for a cubic in percent score in the course, gender, race, class standing, family income, parental education, SAT and ACT scores, high school GPA, taking calculus in high school, score on the university's math placement test, instructor, academic year, and season of course (fall vs. spring term). Effects for women and men are estimated in separate regressions. All outcomes are measured in the two years following the term the student took Principles of Economics I. Major categories based on 2-digit CIP codes.

I checked for agreement between the gradebook data and official transcript data by comparing the ranking of raw scores to the ranking of official letter grades. If scores were entered correctly and completely by the instructors and (as needed) calculated correctly by the researcher, then the ranking of raw scores within a class should align with the ranking of letter grades. For example, a student whose final score was an 80 should have a (weakly) lower letter grade than a student with an 85, conditional on instructor, term, and class section. I do find some violations of this rank restriction. Out of 49 sections, 22 have rank inconsistencies between final score and final letter grade.

These inconsistencies are likely due to measurement error; this could occur if for example if an instructor did not input all assignment scores, or calculated final scores outside of the LMS and changed their weighting scheme from what appeared on the initial syllabus. However, it could also indicate some instructors changing final grades, possibly because of some advocacy on the part of students. The latter case is more problematic for identification because it suggests that students with the same performance could have different grades because of manipulation by the student rather than randomness in the grading curve. The case of measurement error could also bias results because it would mean I am less precisely controlling for underlying performance and effort, which are positively correlated with raw score and grade. I test for robustness to excluding these potentially problematic sections below.

Table 5 shows grade effects estimated only on the sample of course sections with rank consistency between raw scores and letter grades. Although I lose power with a smaller sample, the results are generally robust to this sample restriction. I still find a significant effect of higher grade on taking Principles II (3.8 percentage points, $p<0.01$) and on declaring a business major (1.9 percentage points, $p<0.01$). The estimated effect on declaring an economics major is no longer significant, but the point estimate of 0.4 percentage points is similar to the estimate of 0.8 points with the full sample. The small, negative, marginally significant effect of higher grades on social science major declaration also disappears with this sample: the point estimate is a precise 0.00 percentage points and not significant. Like in Table 2, I detect no significant differences by gender.

Table 5: Estimated Effect of Higher Letter Grade in Introductory Economics, Dropping Course Sections with Grade Rank Inconsistencies

Effect of higher grade on:	All	Women (W)	Men (M)	p-value, W vs. M
Took Principles II	.038***	.043***	.032***	0.563
	(0.009)	(0.014)	(0.012)	
Took Interm. Micro	.004	-.005	.01	0.295
	(0.007)	(0.010)	(0.011)	
Declared Econ Major	.004	.006	.002	0.773
	(0.007)	(0.009)	(0.010)	
Declared Business Major	.019***	.018	.023**	0.726
	(0.007)	(0.011)	(0.009)	
Declared STEM Major	-.004	.001	-.007	0.598
	(0.008)	(0.012)	(0.012)	
Declared Social Science Major	0	0	0	0.999
	(0.006)	(0.011)	(0.008)	
N	6,902	2,829	4,073	

Notes: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Robust standard errors reported. Sample excludes course sections for which the ranking of raw score (from learning management system data) is inconsistent with the ranking of final letter grade (from university transcript data). The results in each column come from a regression of the outcome on a linear term for letter grade. The effects are of one higher letter grade, e.g. going from a B+ to an A- or a C to a C+. Regressions also control for a cubic in percentile rank in the course, gender, race, class standing, family income, parental education, SAT and ACT scores, high school GPA, taking calculus in high school, score on the university's math placement test, instructor, academic year, and season of course (fall vs. spring term). Effects for women and men are estimated in separate regressions. All outcomes are measured in the two years following the term the student took Principles of Economics I. Major categories based on 2-digit CIP codes.

6.3 Change to business school admissions

One potentially confounding policy change occurred during the same period. At MU, the economics department and the business school are closely related, sharing faculty and students. Undergraduate business majors are required to take introductory micro- and macroeconomics in the economics department, so many students who take Principles are aspiring business majors. Traditionally, students applied for admission to the business major after they had already enrolled at MU. Starting with the entering class of fall 2017, the business school started admitting the majority of its students as pre-admits, meaning they applied as seniors in high school and arrived on campus as already declared business majors. This changed the default major from undeclared to business for a number of students in the sample. In addition, the business school expanded its undergraduate class size. These changes could affect the analysis in several ways. The class size expansion could bias effects on business majoring upwards. On the other hand, the fact that students are already in the business school could potentially make them less responsive to grades, since they are no longer competing to get in (though they must still achieve minimum grades in Principles of Economics).

To see how much this business admissions policy change is affecting the results, I estimate Equation 1 using only observations prior to the 2017-2018 school year. Since the economics grading policy changed in 2016, this still leaves one year of data after the curve change. The results are in Table 6. The estimated effects are similar to the effects using the full time period. Notably, when I exclude years affected by the new business school admissions policy, the effect of economics grades on economics majoring is slightly higher (1.1 percentage points compared to 0.8 in Table 2) and the effect on business majoring is somewhat lower (1.5 compared to 2.3 percentage points), though the confidence intervals from the two sets of estimates overlap. I conclude that the change to the business school is not responsible for the substantive findings.

6.4 Checking for selection into grades and courses

My identification strategy assumes that after controlling for underlying performance, letter grades in Principles I are orthogonal to subsequent academic decisions. If there are factors affecting students' grades which are not captured by the

Table 6: Estimated Effect of Higher Letter Grade in Introductory Economics, Excluding 2017-18 Observations

Effect of higher grade on:	All	Women (W)	Men (M)	p-value, W vs. M
Took Principles II	.018***	.021*	.015*	0.666
	(0.007)	(0.011)	(0.009)	
Took Interm. Micro	.008	.004	.011	0.558
	(0.005)	(0.007)	(0.007)	
Declared Econ Major	.011**	.008	.013**	0.547
	(0.005)	(0.007)	(0.007)	
Declared Business Major	.015***	.014	.017**	0.776
	(0.005)	(0.009)	(0.007)	
Declared STEM Major	-.002	.004	-.004	0.527
	(0.006)	(0.009)	(0.008)	
Declared Social Science Major	-.005	-.006	-.004	0.889
	(0.004)	(0.008)	(0.005)	
N	9,102	3,486	5,616	

Notes: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Robust standard errors reported. The results in each column come from a regression of the outcome on a linear term for letter grade. The effects are of one higher letter grade, e.g. going from a B+ to an A- or a C to a C+. Regressions also control for a cubic in percentile rank in the course, gender, race, class standing, family income, parental education, SAT and ACT scores, high school GPA, taking calculus in high school, score on the university's math placement test, instructor, academic year, and season of course (fall vs. spring term). Effects for women and men are estimated in separate regressions. All outcomes are measured in the two years following the term the student took Principles of Economics I. Major categories based on 2-digit CIP codes. Sample includes only students who took Principles I prior to the 2017-18 school year.

raw measure of performance and which also affect outcomes, this exogeneity assumption would be violated. To test for this type of selection into letter grades, I examine whether, conditional on underlying performance, instructor, academic year, and time of year, letter grade predicts observable characteristics such as gender, race, family background, and academic preparation. This is analogous to testing for discontinuities at the cutoff in observable characteristics in a regression discontinuity setting. Although I control for a rich set of observable characteristics in all of the above analyses, significant relationships between grades and observable characteristics could suggest a relationship between grades and unobservable characteristics, which may be biasing my effects upwards.

I estimate:

$$
\begin{aligned}
X_{ijt} = {} & \beta_0 + \beta_1 Grade_{ijt} + \beta_2 Percentile_{ijt} + \beta_3 Percentile_{ijt}^2 + \beta_4 Percentile_{ijt}^3 \\
& + \alpha Instructor_j + \delta Fall_t + \lambda Year_t + \epsilon_{ijt}
\end{aligned}
\tag{2}
$$

For X variables including gender (female indicator), race (indicators for White, Black, Hispanic, and Asian), high school GPA, whether the student took calculus in high school, score on the university's math placement test, and parent education (whether they have a parent with a graduate or professional degree). All other terms are defined as before.

Table 7 shows the results of these falsification tests. Conditional on performance, instructor, and term, students with higher grades are no more likely to be women. A higher grade is associated with a 0.9 percentage point lower chance of a student being Black ($p<0.01$). Higher grades do not predict whether a student took calculus in high school, but they do predict higher high school GPA; one higher letter grade is associated with 0.008 high school grade points on a 4.0 scale. Higher grades also predict a student's performance on MU's math placement test, by 0.2 points on a 25-point scale. Finally, students with higher grades are somewhat more likely to have a parent with a graduate degree (1.3 percentage points, $p<0.1$).

These results suggest that even conditional on performance, students who receive higher letter grades are different than those with lower grades. The conditional exoegeneity assumption may not be fully satisfied and the effects of letter grades I estimate may be upwardly biased. However, the sizes of the associations in Table 7 are

Table 7: Falsification Test: Does Letter Grade Predict Student Characteristics, Conditional on Performance, Instructor, Year, and Season

	Coefficient on grade in Principles I	N in regression
Female	-0.000 (0.007)	11,836
White	0.008 (0.007)	11,203
Asian	-0.001 (0.006)	11,203
Black	-0.009*** (0.003)	11,203
Hispanic	0.000 (0.003)	11,203
High School GPA	0.008*** (0.003)	10,529
Took calculus in high school	0.007 (0.006)	11,836
Math placement score (out of 25)	0.213*** (0.064)	11,357
Parent has graduate degree	0.013* (0.007)	10,814

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors reported. Each coefficient is from a regression of the student characteristic on a linear term for letter grade, controlling for a cubic in percentile rank in the course, instructor, academic year, and season of course (fall vs. spring term). Observations missing a value for the characteristc are not included.

substantively small, and, once multiplied by correlations between the characteristics and outcomes, unlikely to account for treatment effects of the magnitudes I find.

7 Conclusion

Many economics and STEM departments, as they consider ways to attract and retain students and improve representation by gender and other dimensions, are thinking about using grading and evaluation systems as a policy tool that could be used to achieve these goals. For example, the economics department at Duke University instituted a pass/fail grading system for its introductory courses in 2019, motivated by a desire to make the major "more welcoming to students" (Li 2019). Furthermore, the policy change examined in the current study was partially motivated by prior work suggesting women may be particularly deterred by poor grades.

Using grades as an effective policy lever requires understanding how students' academic decisions change in response to grades and how those responses vary across groups. However, grades are not randomly assigned, so estimating causal effects can be challenging using most observational data. To overcome this challenge, I implement an identification strategy that controls for students' raw, continuous performance in introductory economics courses, which captures much of the often unobservable inputs that determine grades, including effort, motivation, and academic preparation. I exploit variation in letter grades conditional on underlying performance, which comes from both naturally occurring variation in grade cutoffs across semesters as well as a discrete change to the grading curve in introductory economics.

I find that receiving a higher grade in Principles of Economics I - Microeconomics makes students 2.5 percentage points more likely to take the next course in the sequence, Principles of Economics II - Macroeconomics. The effect of eventually declaring an economics major is small at 0.8 percent, and marginally statistically significant. However, I find a substantial increase—2.3 percentage points—in the probability of declaring a business major, suggesting that higher economics grades allowed students to gain admission to the prestigious business school.

According to Census Bureau data on postsecondary labor market outcomes, business majors at the studied institution have the highest earnings of any major ten years after graduating—more than economics and even more than engineering and computer science majors. While the median MU economics graduate earns $112,000 ten years after graduating, the median business graduate makes over $160,000.[6] This implies that while the economics department may not have succeeded in its goal of attracting more economics majors, giving out higher economics grades likely improved the labor market outcomes of students by giving them access to the highly selective business school.

I find very little support for the hypothesis that women respond differently to letter grades than men. In my primary specification, the effect of grades on coursetaking and major outcomes are very similar and statistically indistinguishable by gender. In an alternate specification where I control for a student's percent score rather than their percentile ranking, I find differential effects on economics coursetaking and major choice, but in the direction of *men* changing their behavior more. The combined results suggest that at best, giving out higher grades in introductory economics did not affect the gender gap in economics; at worst, it could have widened the gap. Economics departments interested in closing gender gaps would be better off pursuing policies that have proved more effective, such as role model interventions (Porter and Serra 2019) and providing more information about the field of economics (Li 2018; Bayer et al. 2019).

[6]Based on publicly available Post-Secondary Employment Outcomes (PSEO) data from the U.S. Census Bureau, using the 2004-2006 graduating cohorts.

References

Altonji, J. G., P. Arcidiacono, and A. Maurel (2016). The analysis of field choice in college and graduate school: Determinants and wage effects. In *Handbook of the Economics of Education*, Volume 5, pp. 305–396. Elsevier.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics 121*(1-2), 343–375.

Astorne-Figari, C. and J. D. Speer (2019). Are changes of major major changes? The roles of grades, gender, and preferences in college major switching. *Economics of Education Review 70*, 75–93.

Avilova, T. and C. Goldin (2020). What can UWE do for economics?. In S. Lundberg (Ed.), *Women in Economics*. A CEPR Press VoxEU.org book.

Bayer, A., S. P. Bhanot, and F. Lozano (2019). Does simple information provision lead to more diverse classrooms? Evidence from a field experiment on undergraduate economics. *AEA Papers and Proceedings 109*, 110–14.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2016). Stereotypes. *The Quarterly Journal of Economics 131*(4), 1753–1794.

Butcher, K. F., P. J. McEwan, and A. Weerapana (2014). The effects of an anti-grade-inflation policy at Wellesley College. *Journal of Economic Perspectives 28*(3), 189–204.

Calkins, A. (2020). Gender, grades, and college major during the dot-com crash. Working paper.

Chizmar, J. F. (2000). A discrete-time hazard analysis of the role of gender in persistence in the economics major. *The Journal of Economic Education 31*(2), 107–118.

Emerson, T. L., K. McGoldrick, and K. J. Mumford (2012). Women and the choice to study economics. *The Journal of Economic Education 43*(4), 349–362.

Goulas, S. and R. Megalokonomou (2015). Knowing who you are: The effect of feedback information on short and long term outcomes. Working paper.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2019). The allocation of talent and US economic growth. *Econometrica 87*(5), 1439–1474.

Jensen, E. J. and A. L. Owen (2001). Pedagogy, gender, and interest in economics. *The Journal of Economic Education 32*(4), 323–343.

Kaganovich, M., M. Taylor, and R. Xiao (2020). Gender differences in persistence in a field of study. Working paper.

Kugler, A. D., C. H. Tinsley, and O. Ukhaneva (2021). Choice of majors: Are women really different from men? *Economics of Education Review 81*, 1–19.

Li, H.-H. (2018). Do mentoring, information, and nudge reduce the gender gap in economics majors? *Economics of Education Review 64*, 165–183.

Li, X. (2019). Duke's introductory economics course set to institute pass/fail grading system. *The Chronicle*. https://www.dukechronicle.com/article/2019/04/duke-econ- economics-101-pass-fail-satisfactory-unsatisfactory-fullenkamp, April 3, 2019.

Main, J. B. and B. Ost (2014). The impact of letter grades on student effort, course selection, and major choice: A regression-discontinuity analysis. *The Journal of Economic Education 45*(1), 1–10.

Mobius, M. M., M. Niederle, P. Niehaus, and T. S. Rosenblat (2014). Managing self-confidence. Working paper.

Ost, B. (2010). The role of peers and grades in determining major persistence in the sciences. *Economics of Education Review 29*(6), 923–934.

Owen, A. L. (2010). Grades, gender, and encouragement: A regression discontinuity analysis. *The Journal of Economic Education 41*(3), 217–234.

Porter, C. and D. Serra (2019). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics 12*(3), 226–254.

Rask, K. and J. Tiefenthaler (2008). The role of grade sensitivity in explaining the gender imbalance in undergraduate economics. *Economics of Education Review 27*(6), 676–687.

Rask, K. N. and E. M. Bailey (2002). Are faculty role models? Evidence from major choice in an undergraduate institution. *The Journal of Economic Education 33*(2), 99–124.

Stinebrickner, R. and T. R. Stinebrickner (2014). A major in science? Initial beliefs and final outcomes for college major and dropout. *Review of Economic Studies 81*(1), 426–472.

Webber, D. A. (2019). Projected lifetime earnings by major. Technical report.

Zafar, B. (2011). How do college students form expectations? *Journal of Labor Economics 29*(2), 301–348.

Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources 48*(3), 545–595.