

NBER WORKING PAPER SERIES

EQUILIBRIUM GRADE INFLATION WITH IMPLICATIONS FOR FEMALE INTEREST IN
STEM MAJORS

Thomas Ahn
Peter Arcidiacono
Amy Hopson
James R. Thomas

Working Paper 26556
<http://www.nber.org/papers/w26556>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2019

The views expressed in this article are those of the authors and do not necessarily reflect those of the Federal Trade Commission. The contributions of Amy Hopson to this article were prepared in her former capacity as a researcher at Duke University. The opinions expressed in the article are the authors' own and do not reflect the views of the Bureau of Labor Statistics or the Department of Labor or the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Thomas Ahn, Peter Arcidiacono, Amy Hopson, and James R. Thomas. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Equilibrium Grade Inflation with Implications for Female Interest in STEM Majors
Thomas Ahn, Peter Arcidiacono, Amy Hopson, and James R. Thomas
NBER Working Paper No. 26556
December 2019
JEL No. I23,J16

ABSTRACT

Substantial earnings differences exist across majors with the majors that pay well also having lower grades and higher workloads. We show that the harsher grading policies in STEM courses disproportionately affect women. To show this, we estimate a model of student demand courses and optimal effort choices of students conditional on the chosen courses. Instructor grading policies are treated as equilibrium objects that in part depend on student demand for courses. Restrictions on grading policies that equalize average grades across classes helps to close the STEM gender gap as well as increasing overall enrollment in STEM classes.

Thomas Ahn
Naval Postgraduate School
247 Ingersoll Hall
Monterey, CA 93943
sahn1@nps.edu

Peter Arcidiacono
Department of Economics
201A Social Sciences Building
Duke University
Durham, NC 27708
and NBER
psarcidi@econ.duke.edu

Amy Hopson
Bureau of Labor Statistics
akhopson@gmail.com

James R. Thomas
Federal Trade Commission
600 Pennsylvania Ave NW
Washington, DC 20580
james.robinson.thomas@gmail.com

Equilibrium Grade Inflation with Implications for Female Interest in STEM Majors*

Thomas Ahn Peter Arcidiacono Amy Hopson James Thomas
Naval Postgraduate Duke University Bureau of Federal Trade
School NBER & IZA Labor Statistics Commission

December 7, 2019

Abstract

Substantial earnings differences exist across majors with the majors that pay well also having lower grades and higher workloads. We show that the harsher grading policies in STEM courses disproportionately affect women. To show this, we estimate a model of student demand courses and optimal effort choices of students conditional on the chosen courses. Instructor grading policies are treated as equilibrium objects that in part depend on student demand for courses. Restrictions on grading policies that equalize average grades across classes helps to close the STEM gender gap as well as increasing overall enrollment in STEM classes.

1 Introduction

Even after accounting for selection, substantial earnings differences exist across majors. Majors in engineering and the sciences, as well as economics and business, pay substantially more than other fields.¹ Further, earnings disparities across majors have increased substantially over time (Altonji et al. (2014) and Gemici & Wiswall (2014)). Despite their value in the marketplace, STEM (Science, Technology, Engineering, and Mathematics) fields are perceived to be under-subscribed. A report

*The views expressed in this article are those of the authors and do not necessarily reflect those of the Federal Trade Commission. The contributions of Amy Hopson to this article were prepared in her former capacity as a researcher at Duke University. The opinions expressed in the article are the authors' own and do not reflect the views of the Bureau of Labor Statistics or the Department of Labor.

¹See Altonji et al. (2012), Altonji, Kahn & Speer (2016), Xue & Larson (2015) and Altonji, Arcidiacono & Maurel (2016) for reviews.

by the President’s Council of Advisors on Science and Technology (2012) highlighted difficulties in the U.S. retaining “its historical preeminence in science and technology” due to a potential shortage of approximately 1 million STEM professionals over a decade. More recently, Marcia K. McNutt, President of the National Academy of Science, in a statement before the Congressional Committee on Science, Space, and Technology (2019), testified “there are troubling signs that … U.S.-born students are not entering STEM fields in sufficient numbers, and foreign STEM students are no longer coming to the U.S. … as they did before.” In response, there has been a flurry of recent legislative activity to encourage majoring in STEM and to retain STEM degree holders.²

But do universities want more STEM majors? Universities appear to at least tacitly discourage majoring in STEM by being *laissez-faire* with regard to differences in grading and study time across fields. The same majors that pay well also give (on average) significantly lower grades (Sabot & Wakeman-Linn (1991), Johnson (2003)) and are associated with more study time (Brint et al. (2012), Stinebrickner & Stinebrickner (2014)).³ Lower grades and higher study times deter enrollment. Sabot & Wakeman-Linn (1991) show that the absolute level of grades was a far more important indicator of taking further courses in the subject than their ranking within the class. When colleges have intervened in grading practices, enrollment in STEM classes has increased. Butcher et al. (2014) showed that a policy of capping the fraction of A’s given at Wellesley College resulted in shifts towards science classes and science majors.⁴

These differences in grading and workloads affect major decisions in predictable ways. Those who have relatively weak academic preparation (e.g. SAT scores or HS grades) are substantially more likely to leave STEM fields.⁵ At Duke University, African-American men were more likely to

²A number of state legislatures have introduced or are planning to introduce (as of 2019) bills to waive or reduce tuition and/or forgive student loans for STEM students who agree to teach in the state for a number of years (Hinz (2019), Latek (2019), Chapman (2014)). The GI Bill was recently expanded to support veterans seeking a STEM degree (Veteran STEM Scholarship Improvement Act) (Gross 2019). Legislation has also been introduced to keep more international students who graduate in STEM fields in the U.S. by easing restrictions on green cards issuance (Keep STEM Talent Act of 2019, STEM Jobs Act of 2012 (H.R. 6429)) (Graeml 2019).

³There is also evidence that students enter unaware of the extent of cross-department differences in grading standards. Stinebrickner & Stinebrickner (2014) show that the over-optimism regarding performance at Berea College is primarily driven by students over-predicting their performance in the sciences. As students take more classes, they generally revise their expected performance in the sciences downward. This holds true even for students who persist in the sciences who ought to have received relatively positive grade realizations.

⁴Another example is Minaya (2018) that shows that a policy change which reduced the differential in average grades between STEM and non-STEM fields led to an increase in STEM participation.

⁵See Arcidiacono (2004), Arcidiacono et al. (2012), and Arcidiacono et al. (2016).

express an initial interest in STEM majors and economics, yet left these majors at a rate of over 54% compared to just 8% for white men (Arcidiacono et al. 2012). Yet the same study shows that racial differences in STEM persistence disappear when one accounts for differences in pre-college preparation.

Women are also more likely to switch out of STEM, though for different reasons; the existing literature generally finds that this differential attrition cannot be explained by differences in preparation.⁶ This is perhaps surprising because women study substantially more than men (Stinebrickner & Stinebrickner 2004, Arcidiacono et al. 2012) suggesting the higher study times required for STEM classes would not be a deterrent. But if the reason women study more is the higher perceived benefits of the grades themselves—as opposed to a lower cost of studying—then the low grades given in STEM may contribute to the lack of persistence in STEM majors.⁷ In this case, reducing the grading differences across STEM and non-STEM classes would work to mitigate the gender gap in STEM.

In this paper we examine how students make course enrollment decisions and in particular the sensitivity of these enrollment decisions to grading policies. We treat the difference in grading and study time across classes as equilibrium objects. Professors choose these grading policies recognizing how their policies affect the number and type of students who enroll in their courses.

To estimate both the demand and supply side of the model, we use registrar and course evaluation data from the University of Kentucky. The registrar data contains a number of characteristics of the students as well as the course enrollments and grades. The course evaluation data provides information on how many hours a week students are studying in each course.

The raw data show that STEM classes are almost twice as large as their non-STEM counterparts. They also are associated with grades over 0.3 points lower and average study time that is almost 40% higher. Women have higher grades in both STEM and non-STEM classes but are also significantly under-represented in STEM classes.

Our demand-side model helps explain this lack of representation. The demand-side of the model has students choosing courses and exerting effort based on their preferences for classes and

⁶See Arcidiacono et al. (2012), Astorne-Figari & Speer (2019, 2018), Kugler et al. (2017), Ost (2010), and Price (2010). Griffith (2010) is a notable exception which finds that differences in preparation can explain a large portion of the gender difference in STEM attrition.

⁷Using elicited expectations data, Zafar (2013) shows that women have a higher preferences for good grades. Saltiel (2019) shows that, for the same math ability, self-efficacy in math is lower for women. This is perhaps in part due to higher responsiveness to grading differentials across fields.

departments, costs of studying, and expected grade grades. Expected grades in turn depend on the optimal choice of study time. Identification of preferences for grades—and how those preferences differ between men and women—comes from how students sort into classes. The more students sort into classes based on their comparative advantage in grades, the higher the estimated preference for grades.

The results show that comparative advantage in both observed and unobserved ability, differences in preferences for departments, and differences in female representation among STEM faculty all contribute to the gender gap in STEM.⁸ But so too do the differences in grading policies across STEM and non-STEM professors. In fact, our estimates suggest that the harsher grading policies in STEM contribute more to the STEM gap than any of the other previously listed explanations with the exception of differences in preferences for departments. This is driven by our finding that, while both men and women value grades, women value grades significantly more than men.

Requiring all classes to have, for example, the same mean grade may professors to respond along other dimensions such as the amount of work they assign. To understand professor responses—and indeed why the grading differences are there in the first place—we estimate the supply side of the market. We posit a professor utility function that penalizes deviations from his or her (individual-specific) ideal enrollment, average grades, and workload.⁹

Estimates of professor preferences show that STEM instructors on average have lower ideal grades and higher workloads. However, differences in demand for STEM classes also contribute to the lower grades in STEM. Because higher enrollments are costly to the professor, classes that have high demand—which are disproportionately in STEM—give relatively lower grades to deter enrollment. Indeed, removing differences in professor preferences would still result in grades being 0.07 points lower in STEM as a result of the higher demand for STEM classes.

By estimating the supply side, we can examine the general equilibrium effects of policy changes. Requiring the same mean grade across classes provokes a small instructor response but the overall effect of the policy is a substantial increase in the number of STEM classes taken by females, an increase that is significantly larger than the corresponding increase for males. From a financial standpoint as well as its ability to be locally implemented at the school or department level, this

⁸Throughout this study, “ability” is shorthand for a pre-college human capital accumulation. See further discussion in section 5.3.

⁹Ferreyra et al. (2018) also endogenizes study effort (though not at the level of the course) in an equilibrium model.

would be a cheap way of reducing the STEM gender gap.¹⁰

The rest of the paper proceeds as follows. Section 2 describes the data and shows some basic patterns regarding differences in grading between STEM and non-STEM courses. In section 3, we develop a model of student demand for courses and the optimal study time associated with those courses. Section 4 describes our estimation strategy. Section 5 shows the estimates of the model and partial equilibrium counterfactuals. Section 6 explains the set up and estimation of the supply-side, as well as showing the general equilibrium counterfactuals. Section 7 concludes.

2 Data

Estimating an equilibrium model of student choices of courses and study time as well as professor preferences over grade policies requires rich data. We use administrative data coupled with course evaluations from the University of Kentucky (UK). UK, the state's flagship public post-secondary institution, has a current undergraduate enrollment of approximately 21,000. The school was ranked 119 out of approximately 200 'National Universities' by U.S. News & World Report (U.S. News & World Report 2013). This places UK in the middle of the distribution of large post-secondary institutions, and the student body serves as a good cross-section of college students nationwide.

The data set contains student demographic and course enrollment information. Each semester, the student body's course selections and grades are recorded by the Registrar's Office. This data set is particularly valuable because every student outcome in every class is captured, allowing us to estimate a rich model of student and professor interactions. We also collect information on course pre-requisites to accurately capture a student's choice set. We focus on student enrollment observations from one semester, Fall 2012.

In addition, we have access to class evaluation surveys completed by students at the end of the semester. Students do not identify themselves in the evaluation forms, and while we cannot link them to the transcript data at the individual level, we are able to match at more aggregated levels. However, coverage is not complete as some departments opt out of administering end-of-course evaluations. In addition, linking the class evaluation data to the enrollment data is complex, as rules for identifying the course (or sections within the course) and instructor (or sub-instructor - frequently a graduate student - teaching under the supervision of a head-instructor) are defined

¹⁰Such a policy may also have multiplier effects from women making up a greater share of STEM enrollments. This would be the case if the departmental preferences we estimate were influenced by the number of women choosing courses in those departments.

independently by the department. Most critically for our research, students are asked about the number of hours per week they spend on studying for this particular course. This information makes the identification of structural parameters on study effort possible, allowing us to analyze how differences in average study time relates to course and student characteristics across courses.¹¹

Our Fall 2012 selected sample yields approximately 58,000 student/class observations. There are about 16,000 unique undergraduates, implying that on average, each student enrolls and completes four courses.¹² Table 1 provides demographic summary statistics, separated by gender. Overall, women and men look similar when entering college. Women have higher high school grades and slightly lower standardized ACT scores.¹³ Women also have higher grades while in college. Sharp differences show up in major selection. While women comprise a slight majority at UK overall, the ratio between men and women in STEM majors is approximately 1.6. In contrast to students from more selective institutions (seen in many other studies of higher education outcomes), over 30 percent of students at UK are part-time students, taking less than 12 credits during the semester.

Table 2 summarizes class-level characteristics separated by STEM-status of the course. For this paper, we include Economics, Finance, Accounting, and Data Sciences courses to the commonly used definition of STEM courses (Biology, Chemistry, Engineering, Mathematics, and Physics).¹⁴ STEM classes are substantially larger and give significantly lower grades compared to non-STEM courses. As implied by Table 1, female students are the minority in STEM classes. This is despite the fact that they perform better, on average, than their male counterparts in these courses.

For these summary statistics and the reduced-form analyses that follow, we aggregate the evaluation data up to the class level and match to the transcript data. We are able to match 76 percent of classes successfully. We then restrict the data to classes with at least a 50 percent response rate and drop classes with small numbers of respondents to prevent possible identification. On

¹¹The survey asks 20 questions on the value of the course and instructor to the student on a five-point Likert scale. Each student reveals what year of school he or she is in, how valuable he or she finds the course and instructor, expected final grade, and whether the course was a major requirement.

¹²We also observe withdrawal data. Withdrawal rate of undergraduates is approximately 5.4%. Of these, approximately 45% withdraw from the course prior to the midterm examination. We speculate that many of these students were ‘shopping around’ for courses at the start of the semester and realized that they needed to drop a course they had not been attending part-way through the semester.

¹³SAT scores are converted to equivalent ACT scores.

¹⁴See Table A.12 for categorization of departments into STEM / non-STEM. All empirical results presented in the paper have been replicated using only the restrictive definition of STEM, and salient conclusions remain qualitatively identical.

average, students spend about an additional hour per week (or 40 percent more time) studying in a representative STEM course, compared to time spent on studying for an average non-STEM course. The study time difference actually understates the true gap across STEM and non-STEM courses. Students with higher academic ability will select more often into STEM courses, so each hour spent studying should yield more learning. Yet, STEM classes average much lower grades. A student attempting to generate an equivalent grade across a STEM and non-STEM course will have to invest significantly more than an extra 40 percent in study time in the STEM course.

Table 3 presents simple OLS results showing the relationship between individual and class characteristics with grades and study hours after controlling for a large number of academic background measures.¹⁵ The grades regression sample is at the student/class level, and the study hours per week regression sample is at the class level. The first column gives the results for grades. The patterns are consistent with those in Table 2: STEM classes give lower grades and females have higher grades. Classes that have a higher fraction of female students also give higher grades. This is consistent with there not being a grade curve that is common across STEM or non-STEM departments. If such a curve existed, the higher grades females receive would translate into lower grades for everyone else. Class size has a negative effect on grades. The coefficient on class size confounds two effects that work in opposite directions. On the one hand, students prefer higher grades so higher enrollments should be associated with higher grades. On the other, courses that have high intrinsic demand may have lower grades since these courses do not need to have high grades to attract students.

The second column on Table 3 shows regressions of study hours on the average characteristics of the class. STEM classes are associated with about an extra half hour of study, slightly less than what is seen in the descriptive statistics. This suggests that STEM classes are attracting students who are willing to study more, with the grading policies in the STEM classes further spurring on these students to commit more time to study. Classes that have more women also study more, consistent with the previous literature (DiPrete & Buchmann 2013). But perhaps the most interesting coefficient is that on average grades. Courses that give higher grades have

¹⁵We restrict our sample to standard classes with at least 16 students. The total number of classes in the data set is 2,026. From this we exclude nearly half of the classes from the analysis. Many of the excluded classes can be categorized into: non-academic classes (e.g. “academic orientation” or “undergraduate advising”), advanced and remedial independent student courses (including tutoring), classes in fine arts requiring individualized instruction (e.g. “voice”, “jazz ensemble”, or “art studio”), and graduate-level classes taken by very advanced undergraduate students.

Table 1: Descriptive Statistics by Gender

	Men	Women
High school GPA	3.49 (0.472)	3.62 (0.401)
ACT Reading Score	26.1 (5.13)	26.0 (4.84)
ACT Math Score	25.7 (4.65)	23.9 (4.23)
Fall 2012 GPA	2.86 (0.938)	3.12 (0.848)
Fall 2012 Credits	11.5 (3.97)	11.7 (3.85)
STEM Major	38.3%	24.6%
STEM & Econ Major	59.2%	37.6%
Black	6.86%	7.93%
Hispanic	2.48%	2.80%
Misc. Min.	2.10%	2.86%
1st Gen	13.5%	15.0%
Pell Grant	28.8%	30.2%

Note: Fall 2012 University of Kentucky undergraduate students, 7,904 men, 8,286 women. “Econ Major” includes Economics, Finance, Accounting, and Data Sciences. 1st Gen is first-generation college students. Standard deviations in parentheses.

less study time, suggesting grades should be interpreted as relative, not absolute, measures of accomplishment, as well as suggesting grade inflation may have negative consequences for learning.

3 Model

The descriptive evidence motivates our model of course choices. The key components of the model are that grading policies influence enrollment decisions directly because students value grades but also indirectly through incentivizing (costly) study effort. Individual i chooses n courses from the set $[1, \dots, J]$. Let $d_{ij} = 1$ if j is one of the n courses chosen by student i and zero otherwise.

Table 2: Descriptive Statistics by Course Type

	STEM	Non-STEM
Class Size	80.2 (99.3)	41.4 (46.0)
Average Grade	2.94 (0.45)	3.27 (0.42)
Average Grade Female	3.00 (0.56)	3.37 (0.43)
Study Hours	3.37 (1.42)	2.45 (0.81)
Percent Female	37.0%	58.3%
Percent Fem.Prof.	27.0%	46.4%

Note: Fall 2012 University of Kentucky courses with enrollments of 16 or more students, 341 STEM courses, 743 non-STEM courses. For study hours, 327 STEM courses and 652 non-STEM courses. Standard deviations in parentheses.

Following Nevo et al. (2005), we assume the payoff associated with a bundle of courses is given by the sum of the payoffs for each of the individual courses where the payoffs do not depend on the other courses in the bundle.¹⁶ We specify the payoff for a particular course j as depending on student i 's preference for the course, δ_{ij} , the amount of study effort the individual chooses to exert in the course, s_{ij} , and the expected grade conditional on study effort, $\mathbb{E}[g_{ij} | s_{ij}]$:

$$U_{ij} = \phi_i \mathbb{E}[g_{ij} | s_{ij}] - \psi_i s_{ij} + \delta_{ij} \quad (1)$$

ϕ_i then gives how much the individual values grades while ψ_i gives the individual's cost of studying. Students then solve the following maximization problem when choosing their optimal course bundle:

$$\begin{aligned} & \max_{d_{i1}, \dots, d_{iJ}} \quad \sum_{j=1}^J d_{ij} U_{ij} \\ \text{subject to: } & \quad \sum_{j=1}^J d_{ij} = n, \quad d_{ij} \in \{0, 1\} \forall j \end{aligned} \quad (2)$$

¹⁶For a model that includes complementarities in bundled choice, see Gentzkow (2007). The Gentzkow (2007) framework is not feasible in our setting because of the large number of potential course bundles.

Table 3: Regressions of Grades and Study Time on Characteristics of the Individual and/or Class

Dependent Var.	Grade	Study hours per week
STEM Class	-0.422 (0.010)	0.444 (0.094)
Female	0.213 (0.009)	
Percent Female	0.387 (0.024)	0.371 (0.172)
Average Grade		-0.688 (0.077)
ln(Class Size)	-0.057 (0.004)	-0.184 (0.051)
Observations	58,081	968

Note: Additional controls for grades regression include, minority, freshman, first-generation, STEM major, Pell grant, in-state, ACT score, HS GPA, % minority, % freshman, % first-generation. “Average Grade” is calculated using the final grades for the course. Additional controls in study hours regression include % minority, % freshmen, % first-generation, % STEM major, % Pell grant, % in-state, avg. ACT score, avg. HS GPA.

The grade student i receives in course j , g_{ij} , depends on the academic preparation of student i for course j , A_{ij} , the amount of study effort put forth by the student in the course, s_{ij} , the grading policies of the professor, and a shock that is unknown to the individual at the time of course enrollment, η_{ij} . We specify the grading process as:

$$g_{ij} = \beta_j + \gamma_j (A_{ij} + \ln(s_{ij})) + \eta_{ij} \quad (3)$$

Grading policies by the professors are then choices over an intercept, β_j , and a return to academic preparation and effort, γ_j .¹⁷ Gains from study effort enters in as a log to capture the diminishing returns to studying. Along with the linear study effort cost defined in the utility function, this ensures an interior solution for the optimal amount of study time.

Making grades desirable but effort costly allows us to consider two separate explanations for observed gender differences in study hours and conditional grade outcomes: First, women may be studying more and earning conditionally higher grades because they value grades more than men. Alternatively, women may be studying and achieving more because they find studying less costly than men. Section 4 discusses how data on study hours from course evaluations allow us to empirically distinguish these mechanisms.

Students are assumed to know the professors' grading policies.¹⁸ Substituting in for expected grades in (1) yields:

$$U_{ij} = \phi_i (\beta_j + \gamma_j [A_{ij} + \ln(s_{ij})]) - \psi_i s_{ij} + \delta_{ij} \quad (4)$$

The optimal study effort in course j can be found by differentiating U_{ij} with respect to s_{ij} :

$$\begin{aligned} 0 &= \frac{\phi_i \gamma_j}{s_{ij}} - \psi_i \\ s_{ij}^* &= \frac{\phi_i \gamma_j}{\psi_i} \end{aligned} \quad (5)$$

Substituting the optimal choice of study time into (4) yields:

$$U_{ij} = \phi_i (\beta_j + \gamma_j [A_{ij} + \ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i) - 1]) + \delta_{ij} \quad (6)$$

¹⁷For example, if there is a field-wide (or department-level) mandated/recommended grade distribution, we will be able to capture such a policy, as β_j and γ_j will have lower variance with-in vs. across department or field.

¹⁸Students have a number of formal and informal resources to learn about grading policies. Informally, they may rely on friends who have previously taken the course and other information social networks. Professors may send out preemptive signals by posting syllabi online. More formally, course evaluations, which also reveal the (anonymous) responders' own expected final course grades, are on-line and publicly available. In addition, several websites curate online "reviews" of professors and courses.

Those who have lower study costs, low ψ_i , and higher levels of academic preparation, high A_{ij} , find courses with higher γ_j 's relatively more attractive all else equal. Those who place a relatively high weight on expected grades, high ϕ_i , study more conditional on choosing the same course, but are more attracted to courses with higher grade intercepts, high β_j .

Substituting the expression for optimal study time into the grade process equation yields:

$$g_{ij} = \beta_j + \gamma_j (A_{ij} + \ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i)) + \eta_{ij} \quad (7)$$

Professors who set relatively higher values of γ_j see more study effort because higher γ_j 's induce more effort and because higher γ_j 's attract students with lower study costs.

The key equations for estimation are then given by:

- (i) the solution to the students maximization problem where (6) is substituted into (2),
- (ii) the grade production process given in (7), and
- (iii) the optimal study effort given in (5).

The next section describes the parameterizations used to estimate the model as well as the assumptions necessary to overcome the fact that our measures of study effort from the course evaluations are not linked to the individual's characteristics.

4 Estimation

We first describe our estimating strategy under the assumption that there is no student-level unobserved heterogeneity. Under this assumption, the log likelihood function is additively separable, and we are able to estimate the parameters in three stages. We then show how to adapt our estimation method to handle student-level unobserved heterogeneity.

In the first stage, we estimate a reduced form version of the grade production process (Eq. (7)). The relationship between student characteristics and grades gives estimates for the reduced form parameters. The returns to effort are also identified, up to a normalization at the department level, by how student characteristics translate into grades relative to the normalized course.¹⁹

In the second stage, we relate the optimal study effort given in Eq. (5) to the student evaluation data. The evaluations are collected for each class, and students report how many hours they spent

¹⁹Technically, we mean groupings of departments. See Table A.13 for a list of departments in each of the aggregate groupings. From hereon, we refer to these groupings as departments.

studying in that class. This helps us recover some of the study effort parameters, as well as unravel some of the normalizations on the γ 's required in the first stage. We are able to relate all γ 's across departments, but still can only identify the γ 's relative to one course.

In the last stage, we estimate the choice problem given by Eqs. (6) and (2). We use the estimates from the first stage to calculate expected grades for each student and the estimated effort from the second stage. Combining the estimates from these three stages allows us to identify all the grading policy parameters, the grade preference parameters, the effort cost parameters, and the course preference parameters.

4.1 Parameterizations

To estimate the model, we need to place some structure on course preferences, δ_{ij} , the value of grades, ϕ_i , and the cost of effort, ψ_i . Further, we must relate academic preparation, A_{ij} , to what we see in the data. Denote $w_i = 1$ if individual i is female and zero otherwise. Denote X_i as a row vector of explanatory variables such as ACT scores, high school grades, race, etc.²⁰ Denote Z_i as a row vector of explanatory variables that affect preferences for particular departments or levels of courses within departments. Hence Z_i includes gender as well as year in school, allowing women to have preferences for classes in particular departments and the attraction of upper-division versus lower-division classes to vary by department and year in school. Preference shocks for courses are represented by ϵ_{ij} . Finally, we partition courses into K departments, $K < J$, where $k(j)$ gives the department for the j th course. We then parameterize the model as follows:

$$A_{ij} = w_i \alpha_{1k(j)} + X_i \alpha_{2k(j)} \quad (8)$$

$$\delta_{ij} = \delta_{0j} + w_i \delta_{1k(j)} + Z_i \delta_{2k(j)} + \epsilon_{ij} \quad (9)$$

$$\psi_i = \exp(\psi_0 + w_i \psi_1 + X_i \psi_2) \quad (10)$$

$$\phi_i = \phi_0 + w_i \phi_1 \quad (11)$$

There is no intercept in A_{ij} as it can not be identified separately from the β_j 's. Note that the same variables enter into academic preparation, preferences, and effort costs, only with different coefficients. Preferences for courses allow for both course fixed effects as well as students with particular characteristics preferring courses in particular departments, $\delta_{1k(j)}$. Note also that the effort costs are exponential in the explanatory variables. This ensures that effort costs are positive.

²⁰The majority of students at the University of Kentucky submit ACT scores in their college applications.

Finally, preferences for grades are only allowed to vary by gender. In principle, we could allow them to vary with X_i as well, but this would substantially complicate the model.

Having separate estimates by gender across all the relevant parameters will help uncover some of the driving forces behind the gender gap in STEM. For example, if female intrinsic demand for courses in STEM departments is relatively low ($\delta_{1k(j)}$ negative) while preferences for grades and cost of effort are relatively equal across males and females (ϕ_1 and ψ_1 close to zero), then changing grading policies will have no effect on the gender gap in STEM. In this case, it would require figuring out why females are not interested in STEM fields, and policies would have to be geared more towards early education about opportunities in STEM for females, or changing cultural attitudes towards females in STEM. On the other hand, if females have significantly different preferences over grades and study effort compared to males, then altering grading policies could affect the gender distributions within classes and departments. For example, if females have higher preferences for grades (ϕ_1 positive) and lower cost of effort (ψ_1 negative) than males, then increasing γ_j and correspondingly changing β to keep enrollments in STEM courses the same would result in an increase in the fraction of females in STEM.

4.2 Estimation without Unobserved Heterogeneity

4.2.1 Grade parameters

Substituting the parameterizations for academic preparation, A_i , the value of grades, ϕ_i , and study costs, ψ_i , into (7) yields the following reduced form grade equation:

$$g_{ij} = \theta_{0j} + \gamma_j (w_i \theta_{1k(j)} + X_i \theta_{2k(j)}) + \eta_{ij} \quad (12)$$

where:

$$\theta_{0j} = \beta_j + \gamma_j (\ln(\phi_0) + \ln(\gamma_j) - \psi_0) \quad (13)$$

$$\theta_{1k(j)} = \alpha_{1k(j)} + \ln(\phi_0 + \phi_1) - \ln(\phi_0) - \psi_1 \quad (14)$$

$$\theta_{2k(j)} = \alpha_{2k(j)} - \psi_2 \quad (15)$$

We estimate the reduced form parameters $\{\theta_{0j}, \theta_1, \theta_2\}$ as well as the structural slopes, the γ_j 's, using nonlinear least squares. A normalization must be made for every department as scaling up the θ 's by some factor and scaling down the γ 's by the same factor would be observationally

equivalent. We set one γ_j equal to one for each department.²¹ Denote C_k as the normalization for department k . We then estimate γ_j^N where $\gamma_N = \gamma_j/C_{k(j)}$. Similarly, we estimate $\theta_{1k(j)}^N$ and $\theta_{2k(j)}^N$ where $\theta_{1k(j)}^N = \theta_{1k(j)}C_{k(j)}$ and $\theta_{2k(j)}^N = \theta_{2k(j)}C_{k(j)}$.

The variation in the data used to identify $\{\theta_1^N, \theta_2^N\}$ comes from the relationship between student characteristics and grades in each department. The variation in the data used to identify the γ_j^N 's is how these characteristics translate into grades relative to the normalized courses.

4.2.2 Study parameters

We next turn to recovering some of the study effort parameters as well as undoing the normalization made on all the γ 's but one. To do so, we use (5). The issue with using (5) is that we do not directly observe study effort. However, the course evaluation data give reported study hours for each individual in the classroom. This information cannot be linked to the individual data on grades, academic preparation, and course choices. But the evaluation data does provide information about the year in school of the evaluator (e.g., freshman, sophomore, junior, or senior). In contrast to the reduced form results shown in Table 3, the evaluation data is collapsed to the average values per *year in school cross course level* and linked to the enrollment data.²²

To link study hours to study effort, we assume that the relationship is log-log with measurement error ζ_{ij} :

$$\ln(h_{ij}) = \mu \ln(s_{ij}^*) + \zeta_{ij} \quad (16)$$

Substituting in for s_{ij}^* yields:

$$\ln(h_{ij}) = \mu (\ln(\mu_1) + \ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i)) + \zeta_{ij} \quad (17)$$

$$= \kappa_0 + w_i \kappa_1 - X_i \kappa_2 + \mu \ln(\gamma_j) + \zeta_{ij} \quad (18)$$

where:

$$\kappa_0 = \mu(\ln(\phi_0) - \psi_0) \quad (19)$$

$$\kappa_1 = \mu(\ln(\phi_0 + \phi_1) - \ln(\phi_0) - \psi_1) \quad (20)$$

$$\kappa_2 = \mu \psi_2 \quad (21)$$

²¹The study effort analysis allows us to recover the normalizations for all the departments but one, as we will show in section 4.2.2. The final normalization is undone in the estimation of the utility parameters, shown in section 4.2.3.

²²As with the summary statistics, we drop small classes and those with low response rates.

The coefficient on $\ln(\gamma_j)$ then gives the curvature of the relationship between study effort and hours.

Recall that we had to normalize one γ_j for every department in the grade equation. Substituting in with our estimate $\hat{\gamma}_j^N$ and making the appropriate adjustments yields:

$$\ln(h_{ij}) = \tilde{\kappa}_0 + w_i \kappa_1 - X_i \kappa_2 + \kappa_{3k(j)} + \mu \ln(\hat{\gamma}^N) + \zeta_{ij} \quad (22)$$

where $\kappa_{3k(j)} = \mu \ln(C_{k(j)}/C_1)$ and $\tilde{\kappa}_0 = \kappa_0 + \mu \ln(C_1)$. Here C_1 is the normalized course for the base department.

Since we can only link characteristics of the students to the evaluation data by year in school, the observations we use in estimating the study parameters are at the class-year level. Let l_i indicate the year in school of student i . Our estimating equation for students of level l is then:

$$\frac{\sum_i(l_i = l)d_{ij}\ln(h_{ij})}{\sum_i(l_i = l)d_{ij}} = \tilde{\kappa}_0 + w_{jl}\kappa_1 - X_{jl}\psi_2 + \kappa_{3k(j)} + \mu \ln(\hat{\gamma}^N) + \zeta_{jl} \quad (23)$$

where w_{jl} and X_{jl} are the averages of these characteristics for those of year level l enrolled in course j . We correct for potential bias due to measurement error in γ_j using instrumental variables. We use the share of freshmen, sophomores, and juniors in each class as instruments for $\log(\gamma_j)$, and estimate using limited information maximum likelihood (LIML). The instruments are strongly statistically correlated with the potential endogenous variable. F-statistic calculations on the excluded instruments pass the rule-of-thumb test (F-stat = 12.05).

Estimates of (23) allow us to recover the elasticity of hours with respect to study effort, $\hat{\mu}$, as well as an estimate of ψ_2 , how observed characteristics affect study costs, as $\hat{\psi}_2 = \hat{\kappa}_2/\hat{\mu}$. We can also partially undo the normalization on the γ 's, solving for γ 's that are normalized with respect to one course rather than one course in each department. Namely, let $\hat{\gamma}_j^P = \hat{\gamma}_j^N \exp(\hat{\kappa}_{2k(j)}/\hat{\mu})$. $\hat{\gamma}_j^P$ provides an estimate of γ_j/C_1 . The last normalization—the returns on preparation and study time in the normalized course—will be recovered in the estimation of the utility function parameters. The remaining structural parameters embedded in (23) can be recovered after estimating the parameters of the utility function, described in the next section.

4.2.3 Utility parameters

We now turn to estimation of the parameters of the utility function. Given our estimates of the grade equation, equation (12), we can calculate expected grades in each of the courses given optimal study choices:

$$E[\widehat{g_{ij}|s_{ij}^*}] = \hat{\theta}_{0j} + \hat{\gamma}_j^N \left(w_i \hat{\theta}_{1k(j)}^N + X_i \hat{\theta}_{2k(j)}^N \right) \quad (24)$$

Given the estimates of the unnormalized returns to study and ability, $\hat{\gamma}$, we can express the utility i receives from choosing course j and studying optimally as:

$$U_{ij} = \delta_{0j} + w_i \delta_{1k(j)} + Z_{ij} \delta_{2k(j)} + \left(\widehat{E[g_{ij}|s_{ij}^*]} - \hat{\gamma}_j \right) (\phi_0 + w_i \phi_1) + \epsilon_{ij} \quad (25)$$

We then substitute in for $\hat{\gamma}_j$ with $C_1 \hat{\gamma}_j^P$ which, after rearranging terms, yields:

$$U_{ij} = \delta_{0j} + w_i \delta_{1k(j)} + Z_{ij} \delta_{2k(j)} + \widehat{E[g_{ij}|s_{ij}^*]} (\phi_0 + w_i \phi_1) - C_1 \hat{\gamma}_j^P (\phi_0 + w_i \phi_1) + \epsilon_{ij} \quad (26)$$

The goal is then to recover the course fixed effects, δ_{0j} , the value women place on courses in particular departments, $\delta_{1k(j)}$, other department-specific preferences as well as preferences over instructor characteristics, $\delta_{2k(j)}$, preferences over grades, ϕ , and the returns to ability and study time in the normalized course, C_1 .

The variation in the data that identifies ϕ_0 and ϕ_1 comes from how individuals sort based on their comparative advantage in grades. Someone who is strong in mathematics will be more likely to sort into classes where the returns to ability in mathematics is high. To the extent that women are more or less likely to sort based on where their abilities are rewarded then identifies ϕ_1 .

More subtle is the identification on the returns to the normalized course. If separate course fixed effects were estimated for both men and women then C_1 would not be identified as it would be subsumed into the course fixed effects. But by allowing females preferences to vary at the department rather than the course level,²³ the extent to which sorting happens beyond the effect through grades themselves identifies C_1 .

We assume that ϵ_{ij} is distributed Type 1 extreme value. If individuals were choosing one course, estimation of the parameters in (26) would follow a multinomial logit. Students, however, choose bundles of courses. Even though the structure of the model is such that there are no complementarities for choosing particular combinations of courses, the probability of choosing a particular bundle does not reduce to the probabilities of choosing each of the courses separately.

Simulated maximum likelihood

We use simulated maximum likelihood coupled with a fixed point algorithm to estimate the choice parameters. To illustrate the approach, denote K_i as the set of courses chosen by i . Denote M_i as the highest payoff associated with any of the non-chosen courses:

$$M_i = \max_{j \notin K_i} \delta_{0j} + w_i \delta_{1k(j)} + Z_{ij} \delta_{2k(j)} + \left(\widehat{E[g_{ij}]} - \hat{\gamma}_j \right) (\phi_0 + w_i \phi_1) + \epsilon_{ij}$$

²³Note that females are also allowed to value having a female professor more or less than males.

Suppose K_i consisted of courses $\{1, 2, 3\}$ and that the values for all the preference shocks, the ϵ_{ij} 's, were known with the exception of those for $\{1, 2, 3\}$. The probability of choosing $\{1, 2, 3\}$ could then be expressed as:

$$\begin{aligned} \Pr(d_i = \{1, 2, 3\}) &= \Pr(\bar{U}_{i1} > M_i, \bar{U}_{i2} > M_i, \bar{U}_{i3} > M_i) \\ &= \Pr(\bar{U}_{i1} > M_i)\Pr(\bar{U}_{i2} > M_i)\Pr(\bar{U}_{i3} > M_i) \\ &= (1 - G(M_i - \bar{U}_{i1}))(1 - G(M_i - \bar{U}_{i2}))(1 - G(M_i - \bar{U}_{i3})) \end{aligned}$$

where $G(\cdot)$ is the extreme value cdf and \bar{U}_{ij} is the flow payoff for j net of ϵ_{ij} .

Since the ϵ_{ij} 's for the non-chosen courses are not observed, we integrate them out of the likelihood function and approximate the integral by simulating their values from the Type I extreme value distribution. Denoting M_{ir} as the value of M_i at the r th draw of the non-chosen ϵ_{ij} 's and R as the number of simulation draws, estimates of the reduced form payoffs come from solving:

$$\max_{\phi, \delta} \sum_i \ln \left(\left[\sum_{r=1}^R \prod_{j=1}^J (1 - G(M_{ir} - \bar{U}_{ij}))^{d_{ij}} \right] / R \right) \quad (27)$$

Within the simulated maximum likelihood procedure, we solve for the course-specific effects, the female cross department effects, and preferences for upper division STEM and non-STEM courses by for upperclassmen using a fixed algorithm in the spirit of Berry et al. (1995). Using the fixed point algorithm serves two purposes. First, the number of parameters the optimization routine searches over is substantially smaller, searching over four parameters rather than 1,090 parameters.²⁴ Second, it ensures that the model matches the data on the following dimensions (i) enrollment in each course, (ii) the share of courses taken by each gender in each department, and (iii) the share of upper and lower division courses in STEM and non-STEM taken by upper and lower classmen.

The fixed point algorithm works as follows. Denote the share of enrollments in course j in the data as s_j^d . Denote the share of enrollments of women in department k as s_{kw}^d and share of enrollments for upper-class students in upperclass courses of type c , $c \in \{STEM, non-STEM\}$ as s_{cu}^d . Given an initial guess of the δ 's, we can calculate the model analogs to each of these shares and update the model analogs in such a way that we eventually iterate to a fixed point. Letting

²⁴ 1,071 course fixed effects, 13 female cross department effects, and 2 STEM cross upper and lower classmen effects are solved for using the fixed point algorithm. The remaining parameters are preferences for grades for men and women, female cross female professor, and the normalizing constant.

$\tau_i \in (0, 1]$ be a dampening factor to ensure we have a contraction mapping, the m th update of the δ 's follows:²⁵

$$\begin{aligned}\delta_{0j}^{(m)} &= \delta_{0j}^{(m-1)} + \tau_0 \left(\ln \left[s_j^d \right] - \ln \left[s_j \left(\delta^{(m-1)} \right) \right] \right) \\ \delta_{1k}^{(m)} &= \delta_{1k}^{(m-1)} + \tau_1 \left(\ln \left[s_{kw}^d \right] - \ln \left[s_{kw} \left(\delta^{(m-1)} \right) \right] \right) \\ \delta_{2k}^{(m)} &= \delta_{2k}^{(m-1)} + \tau_2 \left(\ln \left[s_{ku}^d \right] - \ln \left[s_{ku} \left(\delta^{(m-1)} \right) \right] \right)\end{aligned}$$

Updating continues until the maximum of the absolute value of the change in the δ 's is sufficiently small.

Our simulated maximum likelihood approach is similar in spirit to that of Nevo et al. (2005). The estimator in Nevo et al. (2005) randomly samples rankings of chosen options, computes likelihood contributions conditional on rankings, and averages across sampled rankings to simulate a full likelihood. Our approach simulates the stochastic utility of the best non-chosen course, computes likelihood contributions conditional on this stochastic utility, and averages across simulation draws to simulate a full likelihood.

Recovering the remaining structural parameters

Given $\hat{\phi}_0$, $\hat{\phi}_1$, and \hat{C}_1 , we are now in a position to recover the remaining structural parameters. The normalizing constants for each department where $k \neq 1$ can be recovered using $\hat{C}_k = \exp(\hat{\kappa}_{2k(j)}/\hat{\mu}_2)\hat{C}_1$. Estimates of the unnormalized γ_j 's are given by $\hat{\gamma}_j = \hat{\gamma}_j^N \hat{C}_{k(j)}$.

The remaining structural parameters from the study effort estimation, equation (23), are the study cost intercept, ψ_0 , and the (relative) female study costs, ψ_1 . These can be recovered using:

$$\begin{aligned}\hat{\psi}_0 &= \ln(\hat{C}_1) + \ln(\hat{\phi}_0) - \frac{\hat{\kappa}_0}{\hat{\mu}} \\ \hat{\psi}_1 &= \ln(\hat{\phi}_0 + \hat{\phi}_1) - \ln(\hat{\phi}_0) - \frac{\hat{\kappa}_1}{\hat{\mu}}\end{aligned}$$

The remaining structural parameters of the grade equation, equation (12), are the course intercepts,

²⁵If we were only updating one of these sets of δ 's no dampening would be needed. We have found that setting τ_0 and τ_1 to 1 and setting τ_3 to 0.5 is sufficient for the problem to be a contraction mapping.

β_j , and the returns to observed abilities $\alpha_{1k(j)}$ and $\alpha_{2k(j)}$. These can be recovered using:

$$\begin{aligned}\hat{\beta}_j &= \hat{\theta}_{0j} - \hat{\gamma}_j(\ln(\hat{\phi}_0) + \ln(\hat{\gamma}_j) - \hat{\psi}_0) \\ \hat{\alpha}_{1k(j)} &= \frac{\hat{\theta}_{1j}^N}{\hat{C}_{k(j)}} - \ln(\hat{\phi}_0 + \hat{\phi}_1) + \ln(\hat{\phi}_0) + \hat{\psi}_1 \\ \hat{\alpha}_{2k(j)} &= \frac{\hat{\theta}_{1j}^N}{\hat{C}_{k(j)}} + \hat{\psi}_2\end{aligned}$$

4.3 Estimation with Unobserved Heterogeneity

We now consider the case when one of the components of X_i is unknown to take into account correlation across outcomes for the same individual. We assume that this missing component takes on S values where π_s is the unconditional probability of the s th value. Let X_{is} be the set of covariates under the assumption that individual i is of type s . The components of the unobserved heterogeneity are identified through the correlation of grades in each of the courses as well as the probabilities of choosing different course combinations.

Integrating out over this missing component destroys the additive separability of the log likelihood function suggesting that the estimation of the three sets of parameters (grades, course choices, and study time) can no longer be estimated in stages. However, using the insights of Arcidiacono & Jones (2003) and Arcidiacono & Miller (2011), it is possible to estimate some of the parameters in a first stage.

In particular, note that the selection problem occurs because students choose courses. By focusing just on the grade estimation as well as a reduced form of the choice problem, we can greatly simplify estimation, recovering the grade parameters as well as the conditional probabilities of being each of the types. These conditional type probabilities can then be used as weights in the estimation of the choice and study parameters.

First consider the parameters of the grade process and the course choices. With unobserved heterogeneity, we now need to make an assumption on the distribution of η_{ij} , the residual in the grade equation. We assume the error is distributed $N(0, \sigma_\eta)$. We then specify a flexible choice process over courses that depends on an parameter vector φ . The integrated log likelihood is:

$$\sum_i \ln \left(\sum_{s=1}^S \pi_s \mathcal{L}_{igs}(\theta, \gamma) \mathcal{L}_{ics}(\varphi) \right) \quad (28)$$

where $\mathcal{L}_{igs}(\theta, \gamma)$ and $\mathcal{L}_{ics}(\varphi)$ are the grade and choice (of courses) likelihoods respectively conditional on i being of type s .

We apply the EM algorithm to then estimate the grade parameters and course choice parameters in stages. We iterate on the following steps until convergence, where the m th step follows:

1. Given the parameters of the grade equation and choice process at step $m-1$, $\{\theta^{(m-1)}, \gamma^{(m-1)}\}$ and $\{\varphi\}$ and the estimate of $\pi^{(m-1)}$, calculate the conditional probability of i being of type s using Bayes rule:

$$q_{is}^{(m)} = \frac{\pi_s^{(m)} \mathcal{L}_{igs}(\theta^{(m-1)}, \gamma^{(m-1)}) \mathcal{L}_{ics}(\varphi^{(m-1)})}{\sum_{s'} \pi_{s'}^{(m)} \mathcal{L}_{igs'}(\theta^{(m-1)}, \gamma^{(m-1)}) \mathcal{L}_{ics'}(\varphi^{(m-1)})} \quad (29)$$

2. Update $\pi_s^{(m)}$ using $\left(\sum_{i=1}^N q_{is}^{(m)}\right) / N$.
3. Using the $q_{is}^{(m)}$'s as weights, obtain $\{\theta^{(m)}, \gamma^{(m)}\}$ by maximizing:

$$\sum_i \sum_s q_{is}^{(m)} \ln [\mathcal{L}_{igs}(\theta, \gamma)] \quad (30)$$

4. Using the $q_{is}^{(m)}$'s as weights, obtain $\varphi^{(m)}$ by maximizing:

$$\sum_i \sum_s q_{is}^{(m)} \ln [\mathcal{L}_{ics}(\varphi)] \quad (31)$$

Once the algorithm has converged, we have consistent estimates of $\{\theta, \gamma, \varphi\}$ as well as the conditional probabilities of being in each type. We can use the estimates of q_{is} as weights to form the average type probabilities of students of year in school l in class j to then estimate the parameters in (23). Finally, we use the estimates of q_{is} as weights in estimating the structural choice parameters using (27).

4.4 Implications from the Demand-Side Estimation

Even without estimating professor preferences, much can be learned from the demand-side estimates. First, we can explain some of the persistent gender gap in STEM majors. Demand-side estimates allow us to decompose differences in course choices, grades, and study effort between males and females into parts due to:

- (i) differences in preferences (δ_{ij}),
- (ii) differences in value of grades (ϕ_i),
- (iii) differences in study costs (ψ_{ij}).

The estimates of the model can also be used to see how enrollment in STEM courses by both men and women would be affected by changes in grading practices. First, we can adjust the intercepts in the grading equation such that the average student's expected grade is the same across courses, isolating the role of the level of the grade from the differences in the slopes, and therefore return to effort. Second, we can forecast course choices if all professors were to have the same grading practices.

5 Results

5.1 Preference estimates

Table 4 presents the preference parameters with the exception of the study costs, the class-specific intercepts, and the coefficients on year in school cross department cross level of the course. Recall that the parameter on expected grades is identified from variation in how abilities are rewarded in different classes. While both men and women value grades, women derive substantively higher utility from higher grades. One additional grade point yields about 62 percent higher utility for females, compared to males. The estimate of female preferences for female professors is positive, with the estimate suggesting that women are indifferent between a class that had a female professor and one that had a male professor who gives grades that are about 0.41 points higher. This coefficient is likely biased upward due to the aggregation of departments. To the extent that female professors are more likely to be in departments that females have a preference for and there is variation within our aggregated groups, we may be picking up within-group preferences for departments.

The second set of rows of Table 4 shows female preferences (relative to male preferences) for different departments. The omitted category is Agriculture. The largest difference in preferences is between Engineering and English: 1.47, which translates into over 3.7 grade points. This helps to account for the severe under subscription to Engineering courses by women, where female share of enrollment is at 18 percent. No other department has female shares under 37 percent. Engineering is thus an outlier, with all the other gaps at 0.73 or smaller (≤ 1.82 grade points).

Table 4: Estimates of Preference Parameters

Preference for:	Coeff.	Std. Error
Expected grades (ϕ)	0.246	(0.010)
Female \times expected grades	0.153	(0.010)
Female \times female professor	0.162	(0.010)
C_1 (normalizing constant)	0.156	(0.014)
<i>Female preferences for Departments</i>		
English	0.345	
Education & Health	0.332	
Psychology	0.245	
Biology	0.202	
Mgmt. & Mkting.	0.151	
Regional Studies	0.147	
Language	-0.123	
Mathematics	-0.148	
Communications	-0.159	
Chem & Physics	-0.198	
Social Sciences	-0.279	
Econ., Fin., Acct.	-0.380	
Engineering	-1.128	

Note: *Female preferences for Departments* are female non-grade preference for departments, compared to males.
STEM departments are bolded.

5.2 Study effort estimates

Estimates of the study effort parameters are presented in Table 5.²⁶ Women have almost 10% higher studying costs compared to men. Overall, women study 33% more than men, but we estimate that they would study over 42% more because of preferences for grades.²⁷ The effort elasticity, μ , is

²⁶Because of measurement error in the γ 's that is compounded by it entering as a log in the study effort equation, we drop classes in the bottom 5% of the γ distribution. Parameters of the study effort equation stabilize after this point.

²⁷This number comes from the difference in the log of the preferences for grades and the translation between effort and study hours: $\mu(\ln(\phi_0 + \phi_1) - \ln(\phi_0) - \psi_1) = 0.861 \cdot (\ln(0.399) - \ln(0.246))$.

0.861 implying 1 percent increase in the returns to effort (γ), translates to a 0.86 percent increase in study hours.

The second set of columns shows how the returns to study effort vary across classes, taking the median γ class for each course grouping. The heterogeneity is quite large. Doubling study effort would translate into an almost 0.4 grade point increase in Engineering but would translate into a bit more than a tenth of a grade point in Management & Marketing and Education & Health.

Table 5: Estimates of Study Effort and Departmental Returns to Studying

	Study Effort			Median γ
	Coeff. ($-\psi$)	Std. Error	Department	Coeff.
Female	-0.097	(0.117)	Engineering	0.373
ACT read	0.025	(0.027)	Mathematics	0.293
ACT math	0.048	(0.029)	Econ., Fin., Acct.	0.284
HS GPA	-0.191	(0.115)	English	0.277
Unobs. Type	-0.674	(0.205)	Chem & Physics	0.268
			Communication	0.226
			Languages	0.222
	Effort Elasticity		Regional Studies	0.220
ln(γ)	0.861	(0.299)	Biology	0.219
			Psychology	0.200
			Agriculture	0.187
			Social Sciences	0.179
			Education & Health	0.150
			Mgmt. & Mkting	0.133

Note: Additional controls include minority indicator variables (Black, Hispanic, Other Minority) and first generation college student status. STEM departments are bolded.

5.3 Grade estimates

The estimated α 's (the department-specific ability weights) are given in Table 6. These are calculated by taking the reduced-form θ 's, undoing the normalization on the γ 's, and subtracting off the part of the reduced form that θ 's that reflects study time (taken from ψ). The departments are sorted such that those with the lowest female estimate are listed first. The female-specific ability

parameter is a measure of pre-college human capital accumulation. Differences between males and females can arise due to individual preferences and/or conscious/sub-conscious bias from teachers or parents, which then lead to different choices in course-work, extra-curricular activities, and hobbies that lead to human capital accumulation along different dimensions.²⁸

Note that four of the five STEM departments have the lowest female estimates. This occurs because females study substantially more than males yet receive only slightly higher grades in STEM courses. Given that sorting into universities takes place on both cognitive and non-cognitive skills and that women have a comparative advantage in non-cognitive skills, males at UK have higher STEM-specific cognitive skills than their female counterpart even though in the population cognitive skills are similar between men and women. The returns to the different components of the ACT score are intuitive. Higher returns to ACT math are seen in STEM classes, with higher returns for ACT reading in English and Psychology.

With the estimates of the grading equation, we can calculate expected grades for an average student in each department. We do this for freshmen, separately by gender, both unconditionally and conditional on taking courses in the department. Results are presented in Table 7. Four patterns stand out. First, there is positive selection into STEM courses: generally those who take STEM classes perform better than the average student. This is not the case for all departments. Indeed, the second pattern is that negative selection is more likely to occur in departments with higher grades. Third, women are disproportionately represented in departments that give higher grades for the average student. Of the six departments that give the highest grades for the average student (female or male), all have a larger fraction female than the overall population. In contrast, of the five departments that give the lowest grades—all of which are in the STEM umbrella—, females are under-represented relative to the overall population in all but one (Biology). Finally, and consistent with Table 6, women have a comparative advantage in non-STEM courses. In almost all cases, the unconditional expected grades for women are higher than for men, in part because women study more. The two exceptions are i) Chemistry & Physics and ii) Economics, Finance, & Accounting.

²⁸Of course, *all* of the data used in the education production function are projections of human capital. The large estimates on female across almost all departments when we control for ACT scores and HS GPA shows that these numerical pre-college academic measures fail to capture substantial portions of human capital.

Table 6: Estimates of Department-Specific Ability Weights (α)

	Female	ACT read	ACT math	HS GPA	Type
Education & Health	0.360	0.001	0.038	1.424	7.694
Regional Studies	0.334	-0.001	0.087	1.903	7.588
Psychology	0.279	0.082	0.083	2.385	7.676
Communication	0.222	0.008	-0.018	1.545	6.330
Mathematics	0.030	-0.024	0.227	1.714	5.772
Languages	0.026	0.022	0.044	1.613	6.716
Social Science	-0.054	0.080	0.035	2.284	8.126
Mgmt. & Mkting	-0.114	0.036	0.013	2.736	10.316
English	-0.150	0.048	0.001	1.353	6.513
Agriculture	-0.219	0.040	0.069	2.371	4.359
Biology	-0.257	0.043	0.107	2.045	7.580
Engineering	-0.264	-0.005	0.059	0.745	4.429
Econ., Fin., Acct.	-0.475	0.016	0.122	1.432	5.120
Chem. & Physics	-0.664	0.000	0.180	1.939	6.214

Note: Additional controls include minority indicator variables (Black, Hispanic, Other Minority) and first generation college student status. STEM departments are bolded.

5.4 Drivers of the STEM gap

Given the estimates of the student's choices over classes and effort and the estimates of the grading process, we now turn to examining sources of the male-female gap in choice of STEM classes. Table 8 shows share of STEM classes taken for males and females as well as how that share changes for women as we change different characteristics. We also report the difference between the male and female shares as a measure of the gender gap in STEM participation. The baseline shares of STEM classes for men and women are 50.7% and 34.6%, respectively implying there is a baseline participation gap of 16 percentage points. The first counterfactual changes female preferences for grades to be the same as male preferences for grades. This increases the share of STEM courses for women by 2.0 percentage points, which is a 12% reduction in the gap. Because professors do not respond to changes in female demand for STEM courses, male STEM enrollment remains

Table 7: Expected Freshmen GPA for Median Classes By Department, Unconditional and Conditional on Taking Courses in that Department

	EGPA Females Unconditional	EGPA Females Conditional	EGPA Males Unconditional	EGPA Males Conditional	Share Female
Education & Health	3.57	3.62	3.40	3.30	0.70
Agriculture	3.46	3.26	3.38	2.97	0.56
Communication	3.43	3.37	3.19	3.17	0.56
Mgmt. & Mkting.	3.34	3.41	3.20	3.28	0.52
Languages	3.28	3.29	3.12	3.12	0.55
Regional Studies	3.22	3.30	3.01	3.12	0.66
Social Sciences	3.12	3.10	2.97	2.90	0.51
English	3.07	3.13	2.91	2.99	0.65
Psychology	2.99	2.98	2.76	2.73	0.67
Engineering	2.86	3.02	2.79	3.04	0.18
Econ., Fin., Acct.	2.79	2.93	2.80	2.90	0.37
Mathematics	2.69	2.75	2.58	2.70	0.47
Biology	2.66	2.81	2.56	2.76	0.60
Chem & Physics	2.43	2.61	2.48	2.73	0.47
Overall					0.51

Note: “Share Female” is % of enrollment in courses offered in the department that is female. “Unconditional” represents the avg. grade outcome assuming that the entire student population enrolls in the course. STEM departments are bolded.

unchanged in this partial equilibrium scenario.²⁹

Turning off observed ability differences such as differences in ACT scores and high school grades has smaller effects on the gap (counterfactual (2)), though larger effects are found for unexplained gender differences in ability (counterfactual (3)). Note that these effects are not driven by women being weaker academically per se, but in part due to women being relatively stronger in non-STEM

²⁹In deriving the counterfactuals, ten classes (out of 1,084) with the estimated $\gamma < 0.01$ have their enrollments fixed. A class with such a flat returns to effort implies that grades are exclusively dependent on ability, which leads to unstable outcomes where these courses are either empty or extremely over-populated. Administrative constraints such as enrollment caps and cancellation of classes with no demand would prevent such occurrences in the real world.

Table 8: STEM Enrollment Shares in Counterfactual Scenarios (Partial Equilibrium)

	STEM Enrollment Share		
	Female	Male	Change in STEM gap [†]
Baseline	34.6%	50.7%	16.0
(1) Equalize Grade Prefs	36.6%		-2.0
(2) Shift obs ability	35.2%		-0.5
(3) Shift unobs ability	35.5%		-0.9
(4) Equalize Non-grade Prefs	39.4%		-4.8
(5) Fem prof effect off	35.0%		-0.4
(6) Grade around 3°	38.5%	52.9%	-1.7

Note: †: “Change in STEM gap” for the Baseline is the difference between the male and female STEM enrollment shares. Values beneath this number are deviations from the baseline for the corresponding counterfactual simulation. Female preference and ability parameters are adjusted to be identical to male preferences and abilities.
 ◊: “Grade around 3” adjusts mean grade in all courses to a B, affecting both males and females. Professor responses remain fixed in partial equilibrium analysis.

courses.

Counterfactuals (4) and (5) look at differences in tastes. Counterfactual (4) turns off non-grade taste differences for departments, which increases the share of women to 39.4%, closing the STEM gap by 4.8 percentage points or 30%. These taste differences may be a mixture of pre-college experiences and the culture of different departments. Hence anything the university can do to close the STEM gap on this end is likely bounded above by this number. Counterfactual (5) turns off female preferences for female professors. One way of closing the gender gap in STEM would be to hire more female professors. However, our estimates suggest that even equivalent representation of female faculty in STEM and non-STEM fields would have a very small effect on the gender gap in STEM participation.

Finally, in counterfactual (6), we examine how changing expected grades across departments affects the gender gap. Namely, we equalize mean grades across courses by increasing (or decreasing) the course-specific intercepts. However, there is still heterogeneity in grades due to the relative difference in γ 's and α 's, the former being especially important as it dictates the returns to studying. This counterfactual raises the share of STEM courses taken by females to about 38.5%, higher than most of the other counterfactuals. The reason the effects are larger than in the first counterfactual

is that the returns to studying are much higher in STEM courses and women are willing to study more than men, due to valuing grades more. Note that while equalizing grades increases female participation by 3.9 percentage points, it also increases male participation by 2.3 percentage points. In aggregate, this decreases the gender gap in STEM participation by 1.7 percentage points or 10%.

While the patterns here suggest a potentially inexpensive way of closing the gender gap is to equalize average grades across fields, professors are likely to strategically respond to restrictions on grading policies. Professors may respond to restrictions on average grades by changing workloads (γ). In the following sections, we explicitly model professor responses and incorporate these responses in our counterfactual predictions.

6 Equilibrium Grading Policies

In Section 5, we showed grading policy parameters β_j and γ_j differ significantly across departments. In particular, STEM courses generally have lower grading intercepts β_j but higher returns on effort γ_j compared to non-STEM courses. One principle goal of this paper is to analyze how these grading differences influence course choices and the implications for the gender gap in STEM.

However, this finding also prompts an additional question: Why do grading policies vary across courses? In particular, why do STEM courses have lower average grades but higher returns on effort than non-STEM courses? Understanding how professors choose grading policies is crucial to anticipate equilibrium responses to changes in the environment. For example, increasing STEM preparation in the hopes of increasing the number of STEM majors may be partially undone by how professors change their grading policies in response to the new environment.

The model we develop allows for grading policies to arise from differences in intrinsic demand of students. Heterogeneity in non-grade preferences δ_{ij} and abilities A_{ij} imply that some courses will be more popular than others even with homogenous grading policies. These differences in intrinsic demand imply that the relationship between grading policies and the composition and outcomes of enrolled students differs across courses. A professor teaching an intrinsically popular course will need to grade especially harshly to achieve the same class size as a less popular course with average grading standards.

Because grading policies in all courses affect the choices of students, the composition of students in each course depends on the grading policies of all professors. This general equilibrium feature means that each professor's optimal grading policy depends on the grading policies of all other

professors. We assume professors do not collude when choosing grading policies implying policies are set in a non-cooperative game between professors.

To estimate professor preference parameters, we solve for parameter values which explain why observed grading policies were optimal for professors. First, we estimate grading policy parameters and student preference parameters using the methods described in Section 4. Second, we derive the set of first order conditions which describe a pure-strategy equilibrium to the non-cooperative grade policy setting game. This system of first order conditions describes how professor preference parameters, grading policy parameters, and student parameters relate to one another when all professors are setting grading policy parameters optimally. Finally, we solve for professor preference parameters which satisfy the set of first order conditions given estimates of grading policy parameters and student preference parameters.

6.1 The Professor's Problem

We assume professors choose grading policy parameters β_j and γ_j to maximize an objective function which depends on both the number and outcomes of students who take their class. We set up the professor's objective function to depend on (i) the number of students in their class, (ii) grades given in the course, and (iii) the cost of assigning work. In particular, we specify the professor's objective function to penalize deviations from the professor's ideal log enrollment, ideal grade for the average student in the class, and ideal workload (as given by γ).

Denote $\bar{G}_j(\beta, \gamma)$ as the expected average grade in class j given the vector of grading policies for all courses β and γ . The dependence on β and γ comes through the composition of the students that take the course. Denoting $P_{ij}(\beta, \gamma)$ as the probability i takes course j given the vector of grading policies, $\bar{G}_j(\beta, \gamma)$ is given by:

$$\bar{G}_j(\beta, \gamma) = \beta_j + \gamma_j \left[\frac{\sum_i^N P_{ij}(\beta, \gamma) [A_{ij} + \ln(\phi_i) - \ln(\psi_i)]}{\sum_i^N P_{ij}(\beta, \gamma)} + \ln(\gamma_j) \right]$$

Then the objective function professor j maximizes is:

$$V_j(\beta, \gamma) = - \left(\ln \left[\sum_i P_{ij}(\beta, \gamma) \right] - e_0 \right)^2 - \lambda_1 (\bar{G}_j(\beta, \gamma) - e_{1j})^2 - \lambda_2 (\gamma_j - e_{2j})^2 \quad (32)$$

Specifying the objective function in this way allows for three sources of heterogeneity across departments to affect grading policies. First is through innate demand for the courses. Courses differ in their innate demand and systematic differences exist across departments. These innate differences in demand affect both log enrollment—which affects the first term—but also affects the

probability particular students take particular classes which can then affect the average grade given in the class. Second is through preferences for ideal grades. This is captured by e_{1j} where ideal grades may on average be different across courses of particular types. Third is through ideal optimal amount of work to assign, e_{2j} . The ideal amount of work to assign may differ across departments in part due to differences in the grading technology: it may be easier to grade, for example, math problems than essays.

6.2 Estimation

Denote W_j as a vector that includes indicators for whether the course is in each of the fourteen departments³⁰, whether the course is an upper division, and whether the course is upper division and in STEM. We parameterize e_{1j} and e_{2j} such that they vary with W_j as well as unobserved preference terms e_{1j}^* and e_{2j}^* according to:

$$\begin{aligned} e_{1j} &= W_j \Psi_1 + e_{1j}^* \\ e_{2j} &= W_j \Psi_2 + e_{2j}^* \end{aligned}$$

e_{1j}^* and e_{2j}^* are assumed to be orthogonal to the department and the level of the course as well as innate demand (the δ_j 's). These unobserved preference terms are not, however, orthogonal to the grading policies. To account for the endogeneity of the grading policies, we use simulated method of moments to estimate the weights on grades and workload (λ_1 and λ_2) as well as the parameters governing ideal grades and workloads (Ψ_1 and Ψ_2).

Because of the complexity of jointly solving for the optimal grading policies as functions of the parameters, we instead set up a grid of possible grading combinations. These combinations include the full range of γ_j 's as well as β_j 's that produce expected grades on the support of the data. Fixing the other professor's grading policies as what we have estimated in the previous step (the estimates of the β_j 's and γ_j 's), we calculate expected grades and enrollment for the different combinations of β and γ on the grid.³¹ Using the log expected enrollments and grades for each combination of β and γ as dependent variables, we then regress them on the a spline in β and γ , doing this separately for each class. These functions predict expected enrollments and grades extremely well and we use them when we calculate the professor's objective function.

³⁰Recall that classes are only counted as belonging to one department

³¹We use 50 β 's and 50 γ 's where the grid points are equally spaced and the support covers the range of expected grades and γ 's seen in the data.

The method of simulated moments then works as follows. Given draws of e_{1j}^* and e_{2j}^* , we maximize the professor's objective function with respect to β and γ where some of the components of the objective function (expected grades and log expected enrollments) are approximated using the previously mentioned spline. Given the optimal grading policies for particular draws of the unobserved preferences, we then create moments that we can match from the data.³² Examples of moments we match are, for each department, average grades, estimated γ 's, and enrollment. The full list of moments are included in the Appendix Table A.14.

6.3 Professor preference estimates

Table 9 shows estimates of the professor parameters. The first row shows ideal log enrollment. The number is close to zero—which is outside of the support—implying that enrollment is always costly. Given that enrollment is costly, instructors deviate from their optimal grades and workloads by giving lower grades and higher workloads than they would like.

The first column of Table 9 shows how ideal grades vary by department. The sorting of departments by ideal grades shows that STEM instructors generally prefer having lower average grades than their non-STEM counterparts. Note that for the baseline category—agriculture—the ideal grade is between an A and a B (3.5). The actual average grade is closer to a B (3.2). This suggests instructors prefer to give high grades but do not do so in part because it would encourage student demand. The penultimate entry of the first column shows the standard deviation of the unobserved preference over ideal grades. Most departments are within one standard deviation of the unobserved preference with Education & Health an outlier on the high side and Engineering a near-outlier on the low side.

The second column shows the results for the ideal workload. Departments with low ideal grades tend to have high ideal workloads. For example, Engineering, Economics, and Mathematics have the lowest ideal grades, yet are in the group of the four departments with the highest ideal workloads. The correlation between the two sets of departmental coefficients is -0.76. The estimated parameters imply that upper-level classes in some non-STEM departments (such as Education & Health and Management & Marketing), the ideal grade is an A and the ideal γ is near zero.

To summarize, non-STEM professors generally prefer higher grades and require lighter workloads compared to their STEM colleagues.³³ And both STEM and non-STEM professors prefer to

³²We use 64 moments to estimate 37 parameters.

³³Note that our model ignores spillovers across classes. Higher grades in introductory classes results in higher

give out higher grades with lower workloads in upper-division classes. But student demand also affects grading policies as instructors would prefer smaller class sizes. Biology and Chemistry & Physics departments offer an example of this dynamic. Professors in these departments prefer giving students relatively high grades and low workloads. Indeed, their grading and effort preferences track closer to professors in non-STEM departments. Yet, as we can observe in Table 7, these two departments have the lowest average grades for freshmen, at around 2.7. The connection between these seeming contradictory results lies in the high level of demand for courses in these two departments. Average enrollment size in classes in these two departments are over 100, more than double the size of the average class in other departments.

6.4 Supply-side counterfactuals

To better understand the role of professor preferences in grading policies and STEM enrollment by gender, Table 10 revisits the counterfactual analyses of Table 8 in a general equilibrium framework in which instructors are allowed to re-optimize their grading policies in response to changes in student characteristics or other factors. To do this, we first take the first order conditions of professor j 's maximization problem with respect to β_j and γ_j . Following this, we rearrange these first order conditions so that optimal grading policies of instructor j , β_j and γ_j , are defined as a function of student characteristics, professor and student preference parameters, and equilibrium grading policies of all instructors including instructor j . We then iteratively generate vectors of grading policies until we obtain policies which satisfy the rearranged first order conditions. When this condition holds, every instructor's grading policies are a best response to all other instructor grading policies.

Table 10 conducts the same counterfactuals as those in Table 8 with instructors now allowed to update their grading practices in response to changes in student preferences and characteristics. With one exception—equalizing grade preferences—, adding in responses by professors slightly lowers the counterfactual decreases in the STEM gender gap. For example, in partial equilibrium, shifting unobserved ability for females to that of males reduces the difference between the share of STEM courses taken by females and the share of STEM courses taken by males by 0.9 percentage points.³⁴ But Table 10, counterfactual (3) shows that taking into account professor responses means that the

demand for more advanced classes. Hence departmental pressures may explain part of the preferences we observe.

³⁴From Table 8, the gender gap falls from 16 percentage points to 15.1 percentage points, for a difference of 0.9 percentage points.

Table 9: Estimates of Professor Preferences

Ideal log enrollment (000's)	0.109		
		<i>Ideal Grades</i>	<i>Ideal γ</i>
Constant	3.536	Constant	0.112
Econ., Fin., Acct.	-0.431	Engineering	0.181
Engineering	-0.271	Communication	0.109
Mathematics	-0.076	Econ., Fin., Acct.	0.103
English	-0.061	Regional Studies	0.099
Chem & Physics	-0.002	English	0.092
Biology	0.014	Mathematics	0.050
Social Sciences	0.038	Language	0.016
Language	0.061	Chem & Physics	0.001
Psychology	0.101	Biology	0.000
Regional Studies	0.203	Education & Health	-0.009
Mgmt. & Mktng.	0.219	Psychology	-0.015
Communication	0.244	Mgmt. & Mktng.	-0.034
Education & Health	0.430	Social Sciences	-0.036
Upper division	0.262	Upper division	-0.088
STEM upper division	-0.001	STEM upper division	0.055
standard dev of ideal \bar{G} pref	0.449	standard dev of ideal γ pref	0.118
weight on ideal \bar{G}	1.866	weight on ideal γ	16.748

Note: "Upper division" are courses with numerical designations 300 and above, typically reserved for classes with pre-requisites where the majority of students enrolled have opted into the major. STEM departments are bolded.

gender gap only falls by 0.7 percentage points. The lower general equilibrium difference is driven by the gap in grades given between STEM and non-STEM professor growing. Namely, shifting female unobserved ability to that of males results in STEM (non-STEM) professors responding by giving grades that are 0.04 points lower (0.2 points higher) than in partial equilibrium. Average grades in STEM classes then move from being 0.38 points lower than average grades in non-STEM classes to 0.43 point lower. Since females respond more to grade differences than males, professor responses work to reduce the change in the STEM gender gap.

Equalizing grade preferences (counterfactual 1) shows a different pattern. In partial equilibrium,

Table 10: STEM Enrollment Shares in Counterfactual Scenarios (General Equilibrium)

	STEM Enrollment Share			Fem PE Effect*
	Female	Male	Change in STEM gap [†]	/ GE Effect
Baseline	34.6%	50.7%	16.0	
(1) Equalize Grade Prefs	36.8%	50.6%	-2.2	0.90
(2) Shift obs ability	34.9%	50.5%	-0.4	1.86
(3) Shift unobs ability	35.1%	50.4%	-0.7	1.88
(4) Equalize Non-grade Prefs	39.1%	50.4%	-4.6	1.08
(5) Fem prof effect off	35.0%	50.7%	-0.3	1.10
(6) Grade around 3°	38.4%	52.9%	-1.6	1.02

Note: †: “Change in STEM gap” for the Baseline is the difference between the male and female STEM enrollment shares. Values beneath this number are deviations from the baseline for the corresponding counterfactual simulation. Female preference and ability parameters are adjusted to be identical to male preferences and abilities.

◊: “Grade around 3” adjusts mean grade in all courses to a B, affecting both males and females.

*: Female PE effect / GE effect reports the PE counterfactual change in the share of courses that women take in STEM divided by the corresponding GE counterfactual change. Professors change grading strategies based on student responses to changes in preferences and abilities for general equilibrium analysis.

the STEM gap shrinks by 2.0 percentage points when we shift female preferences for grades to be the same as male preferences for grades. In general equilibrium, the STEM gap shrinks by even more at 2.3 percentage points. The reason for this is that grades are no longer as effective at deterring enrollment. Since instructors are grading more harshly than they would prefer, all departments raise the grades they offer. But this increase is larger for departments that are furthest away from their ideal grades; that is, those departments who are grading harshly because their class sizes are especially large. Biology and Chemistry & Physics see especially large increases in grades and are the two STEM fields where women have preferences most similar to men.

But considering the STEM gap alone masks the extent of the equilibrium responses as these responses often move men and women in the same direction. Table 10 compares how large the partial equilibrium effects are relative to the general equilibrium effects by dividing the partial equilibrium change in the share of classes taken by women that are in STEM by the corresponding general equilibrium change. Doing so shows the extent to which the partial equilibrium change

understates or overstates the general equilibrium change.

As with the changes in the STEM gap, the partial equilibrium changes are larger than the general equilibrium changes in all the counterfactuals except the one that equalizes grading preferences. Changing female abilities (either observed or unobserved) results in the partial equilibrium change being over 86% higher than the general equilibrium change. The differences between the partial equilibrium and general equilibrium are smaller for changes in preferences for department or shifting the fraction of female professors to be the same across departments (8% to 10% higher in partial equilibrium), and smaller still for curving around a 3.0 (2% higher in partial equilibrium). This latter result is especially important, implying that, even taking into account professor responses, curving around a 3.0 would increase the fraction of courses women take in STEM by 3.8 percentage points, an 11% increase.

Finally, in our framework, grading policies vary for three reasons: first, because of differences in intrinsic student demand; second, because of differences in professor preferences for ideal grades; and third, because of differences in ideal workloads. If the same instructor is assigned to teach both a course with low intrinsic demand and a course with high intrinsic demand, she will grade more harshly in the course with high demand to reduce enrollment.

Table 11: STEM Enrollment Shares and Grades After Eliminating Differences in Professor Preferences

	STEM Share			Avg Grade	
	Female	Male	Overall	STEM	Non-STEM
Baseline	34.6%	50.7%	42.4%	2.79	3.22
Equalize professor preferences	38.0%	53.4%	45.5%	2.95	3.02

Note: Equalizing professor preferences entails setting ideal average grades and ideal γ 's for all courses to the mean across courses.

In Table 11 we investigate how grading practices and enrollments would change if we removed preferences instructor differences in preferences for grades and workloads. We do this by setting all instructor ideal grades and γ 's to the mean. The differences between average grades given in STEM and non-STEM classes is markedly reduced, though not eliminated, when differences in instructor preferences are turned off. At baseline, STEM classes gave grades that were 0.43 points lower than their non-STEM counterparts. Removing differences in instructor preferences shrinks the difference

to 0.07 points. This remaining difference is the result of higher demand for STEM courses.³⁵

Removing differences in instructor preferences increases enrollment in STEM courses by 3.1 percentage points. These effects are larger for women: the share of courses women take in STEM rises by 3.4 percentage points versus 2.7 percentage points for men. Comparing the effects of equalizing professor preferences to the effects of grading around a 3.0 in Table 10 shows less (more) STEM participation by women (men). Grading around a 3.0 eliminates the grading differences across STEM and non-STEM courses but leaves substantial differences in workloads. Eliminating differences in professor preferences reduces, but does eliminate, grading differences across departments as well as reducing differences in workloads. Since women value grades more than men, the elimination of the grading differences benefits them more than their male counterparts.

7 Conclusion

The number of STEM graduates—especially for under-represented groups—has been an ongoing concern. At the same time, STEM courses are on average associated with lower grades and higher study times, both factors that may deter enrollment. Using administrative data from the University of Kentucky, we estimate a model of course choices to understand what influences STEM enrollment and how those influences differentially affect men and women. While we show that a variety of factors influence how students choose courses, we find that differences in grading policies play an important role in suppressing STEM demand and this is particularly true for female students.

Specifically, we show that STEM courses assign significantly lower grades than non-STEM courses and that a policy of curving all courses around a B would increase overall STEM participation by 7.2% and female STEM participation by 11.3%. This effect on female participation is only slightly less than the effect of giving female students the same course content preferences as male students (13.9%). Removing gender differences in content preferences would require large changes in social attitudes and behaviors and might take decades. Comparatively, eliminating grading differences across fields should be relatively straightforward and affordable.

One issue with policies aimed at reducing grading differences is that instructors may respond to these policies by changing other aspects of their courses. To capture these responses—and to understand the source of grading differences more generally—our analysis treats grading policies

³⁵Note that this is the difference due to demand for these particular courses. A potential reason why STEM classes as a whole have lower ideal grades may be overall demand for these majors resulting in pressure to restrict enrollment in STEM classes above and beyond that due to the demand for a specific STEM course.

as equilibrium objects chosen by instructors in competition with one another. Higher demand for STEM courses explain part, but not all, of the differences in grading between STEM and non-STEM courses. Accounting for the endogenous response by professors to having to curve around a B, however, has a limited impact on the overall effects of the policy: large gains in STEM enrollment would still occur and the gains would be especially large for women.

While it is beyond the scope of this analysis, we speculate that long run effects of grading regulations could also exceed our estimates due to positive feedback loops generated by other mechanisms. For example, additional female STEM students attracted by more balanced grading may induce culture changes that make STEM courses more hospitable to women. These culture changes may in turn attract additional women creating a positive feedback loop that supplements the initial effects of grading regulations. We leave analyses of these potential positive feedbacks to future work.

References

- Altonji, J. G., Arcidiacono, P. & Maurel, A. (2016), ‘The analysis of field choice in college and graduate school: Determinants and wage effects’, *Handbook of Labor Economics* **5**.
- Altonji, J. G., Blom, E. & Meghir, C. (2012), ‘Heterogeneity in human capital investments: High school curriculum, college major, and careers’, *Annual Review of Economics* **4**(1), 185–223.
- Altonji, J. G., Kahn, L. B. & Speer, J. D. (2014), ‘Trends in earnings differentials across college majors and the changing task composition of jobs’, *The American Economic Review* **104**(5), 387–393.
- Altonji, J. G., Kahn, L. B. & Speer, J. D. (2016), ‘Cashier or consultant? entry labor market conditions, field of study, and career success’, *Journal of Labor Economics* **S1**(34), 361–401.
- Arcidiacono, P. (2004), ‘Ability sorting and the returns to college major’, *Journal of Econometrics* **121**(1), 343–375.
- Arcidiacono, P., Aucejo, E. M. & Hotz, V. J. (2016), ‘University differences in the graduation of minorities in STEM fields: Evidence from California’, *American Economic Review* **106**(3), 525–562.
- Arcidiacono, P., Aucejo, E. M. & Spenner, K. (2012), ‘What happens after enrollment? an analysis of the time path of racial differences in GPA and major choice’, *IZA Journal of Labor Economics* **1**(1), 1–24.
- Arcidiacono, P. & Jones, J. B. (2003), ‘Finite mixture distributions, sequential likelihood and the EM algorithm’, *Econometrica* **71**(3), 933–946.
- Arcidiacono, P. & Miller, R. A. (2011), ‘Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity’, *Econometrica* **7**(6), 1823–1868.
- Astorne-Figari, C. & Speer, J. D. (2018), ‘Drop out, switch majors, or persist? The contrasting gender gaps’, *Economics Letters* pp. 82–85.
- Astorne-Figari, C. & Speer, J. D. (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.

Berry, S., Levinsohn, J. & Pakes, A. (1995), ‘Automobile prices in market equilibrium’, *Econometrica* **63**(4), 841–890.

Brint, S., Cantwell, A. M. & Saxena, P. (2012), ‘Disciplinary categories, majors, and undergraduate academic experiences: Rethinking Bok’s “underachieving colleges” thesis’, *Research in Higher Education* **53**(1), 1–25.

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

Chapman, B. (2014), ‘Top-performing high school seniors can get free ride to state colleges for science studies’, *New York Daily News May*.

DiPrete, T. A. & Buchmann, C. (2013), *Rise of Women, The: The Growing Gender Gap in Education and What it Means for American Schools*, Russell Sage Foundation.

Ferreyra, M., Garriga, C. & Diaz, A. S. (2018), ‘A general equilibrium analysis of college enrollment, completion, and labor market outcomes’, Working paper.

Gemici, A. & Wiswall, M. (2014), ‘Evolution of gender differences in post-secondary human capital investments: College majors’, *International Economic Review* **55**(1), 23–56.

Gentzkow, M. (2007), ‘Valuing new goods in a model with complementarity: Online newspapers’, *American Economic Review* **97**(3), 713–744.

Graeml, Y. (2019), ‘Senators introduce bill seeking to help international stem students work in us’, *The Daily Californian* .

URL: www.dailycal.org/2019/06/12/senators-introduce-bill-seeking-to-help-international-stem-students-to-work-in-us/

Griffith, A. L. (2010), ‘Persistence of women and minorities in STEM field majors: Is it the school that matters?’, *Economics of Education Review* **29**(6), 911–922.

Gross, N. (2019), ‘Your GI bill benefits: Everything you need to know’, *Military Times* .

URL: www.militarytimes.com/education-transition/2019/07/20/gi-bill-benefits-guide/

Hinz, G. (2019), ‘STEM students would get free tuition under Il. bill’, *Crain’s Chicago Business* .

URL: www.chicagobusiness.com/greg-hinz-politics/stem-students-would-get-free-tuition-under-illinois-bill

- Johnson, V. E. (2003), *Grade inflation: A crisis in college education*, Springer.
- Kugler, A. D., Tinsley, C. H. & Ukhaneva, O. (2017), ‘Choice of majors: Are women really different from men?’, Working paper.
- Latek, T. (2019), ‘Lawmakers drawing up proposals for 2020 general assembly’, *Kentucky Today* .
URL: kentuckytoday.com/stories/lawmakers-drawing-up-proposals-for-2020-general-assembly
- Minaya, V. (2018), ‘Do differential grading standards across fields matter for major choice? Evidence from a policy change in florida’, Working paper.
- Nat (2019), *Maintaining U.S. Leadership in Science and Technology*.
- Nevo, A., Rubinfeld, D. L. & McCabe, M. (2005), ‘Academic journal pricing and the demand of libraries’, *American Economic Review* **95**(2), 447–452.
- Ost, B. (2010), ‘The role of peers and grades in determining major persistence in the sciences’, *Economics of Education Review* **29**(6), 923–934.
- Price, J. (2010), ‘The effect of instructor race and gender on student persistence in STEM fields’, *Economics of Education Review* **29**(6), 901–910.
- Report to the President: Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics* (2012).
- Sabot, R. & Wakeman-Linn, J. (1991), ‘Grade inflation and course choice’, *The Journal of Economic Perspectives* pp. 159–170.
- Saltiel, F. (2019), ‘What’s math got to do with it? multidimensional ability and the gender gap in STEM’, Working paper.
- Stinebrickner, R. & Stinebrickner, T. (2014), ‘A major in science? Initial beliefs and final outcomes for college major and dropout’, *The Review of Economic Studies* **81**(1), 426–472.
- Stinebrickner, R. & Stinebrickner, T. R. (2004), ‘Time-use and college outcomes’, *Journal of Econometrics* **121**(1), 243–269.
- Xue, Y. & Larson, R. C. (2015), ‘Stem crisis or stem surplus: Yes and yes’, *ILR Review* **1**(138).
- Zafar, B. (2013), ‘College major choice and the gender gap’, *Journal of Human Resources* **48**(3), 545–595.

Appendix

Table A.12: STEM Classification by Department

	Non-STEM	STEM
Aerospace Studies	Education	Merch, Apparel & Textile
Agr Economics	Ed Curriculum & Instr	Mkt Sci & Leadership
Agricultural Biotechnology	Ed Pol Studies & Evaluation	Modern & Classical Lang
Agricultural Education	Ed, Schl & Counseling Psych	Nursing
Agriculture General	English	Philosophy
Allied Health Ed & Rsrch	Environmental Studies	Plant Pathology
Animal & Food Sciences	Family Sciences	Plant & Soil Sciences
Anthropology	Fine Arts - Music	Political Science
Appalachian Studies	Fine Arts - Theatre Arts	Psychology
Arts Administration	Forestry	Public Health
Biosystems & Agr Engineering	Gender & Women's Studies	STEM Education
Communication	Geography	Schl of Journalism & Telecomm
Communication Disorders	Health Sciences Education	Schl of Art and Visual Studies
Communication & Info Studies	Hispanic Studies	Schl of Human Env Sci
Community & Leadership Dev	History	Schl of Interior Design
Dept of Management	Kinesiology - Hlth Promotion	Social Work
Dept Of Gerontology	Landscape Architecture	Sociology
Dept of Mkt & Supply Chain	Latin American Studies	Sustainable Agriculture
Dietetics & Nutrition	Library & Info Science	School of Architecture
Early Childhd, Spec Ed, Rehab	Linguistics	Statistics

Table A.13: Aggregation of Departments

Categories	Departments
Agriculture	Agricultural Biotechnology, Agricultural Economics, Agricultural Ed, Agriculture General, Animal & Food Sciences, Biosystems & Agr Engineering, Environmental Studies, Forestry, Landscape Architecture, Plant Pathology, Plant & Soil Sciences, Sustainable Agriculture
Regional Studies	Appalachian Studies, Family Sciences, Gender & Women's Studies, Hispanic Studies, Latin American Studies
Communication	Arts Admin, Communication, Communication & Info Studies, Fine Arts - Music, Fine Arts - Theatre Arts, Schl Of Journalism & Telecomm, Schl of Art & Visual Studies, Schl of Interior Design
Ed & Health	Allied Health Ed & Research, Comm Disorders, Community & Leader Dev, Dept of Gerontology, Dietetics & Nutrition, Early Child, Spec Ed, Rehab, Ed, Ed Curriculum & Instr, Ed Policy Studies & Eval, Ed, Schl & Counsel Psych, Health Sci Ed, Kinesiology- Health Promotion, Lib & Info Sci, Nursing, Public Health, STEM Ed, Social Work
Engineering	Chemical & Materials Engineering, Civil Engineering, Computer Science, Electrical & Computer Engineering, Engineering, Mechanical Engineering, Mining Engineering, Schl of Architecture
Languages	Linguistics, Modern & Classical Languages, Philosophy
English	English
Biology	Biology, Entomology
Mathematics	Mathematics, Statistics
Chem & Physics	Chemistry, Earth & Environmental Sciences, Physics & Astronomy
Psychology	Psychology
Social Sciences	Anthropology, Geography, History, Political Science, Schl of Human Environmental Sciences, Sociology
Mgmt. & Mktng.	Aerospace Studies, Department of Mgmt, Dept of Mkt & Supply Chain, Merchand, Apparel & Textiles, Mil Sci & Leadership
Econ., Fin., Acct.	Accountancy, Economics, Dept of Finance & Quantitative Methods

Table A.14: Simulated and Actual Moments

	Av. log Enrollment		Av. \bar{G}		Av. γ	
	Actual	Simulated	Actual	Simulated	Actual	Simulated
Agriculture	3.25	3.33	3.26	3.42	0.135	0.175
Regional Studies	3.32	3.37	3.25	3.25	0.292	0.206
Communication	3.46	3.50	3.31	3.32	0.287	0.233
Education & Health	3.48	3.50	3.54	3.57	0.157	0.165
Engineering	3.56	3.58	3.01	3.08	0.331	0.356
Languages	3.40	3.41	3.12	3.20	0.194	0.228
English	3.31	3.35	2.97	3.02	0.278	0.283
Biology	4.58	4.58	2.93	2.84	0.239	0.215
Mathematics	4.04	4.05	2.81	2.80	0.239	0.215
Chem & Physics	4.36	4.34	2.88	2.79	0.232	0.253
Psychology	4.10	4.09	3.11	3.08	0.172	0.189
Social Sciences	3.68	3.67	3.07	3.04	0.161	0.180
Mgmt. & Mkting.	3.62	3.63	3.41	3.39	0.135	0.135
Econ., Fin., Acct.	3.90	3.95	2.69	2.84	0.276	0.276
non-STEM upper division	3.41	3.37	3.45	3.34		
STEM upper division	3.59	3.58	3.07	3.00		
Std dev lower non-stem	0.862	0.856	0.474	0.471	0.136	0.101
Std dev lower stem	0.587	0.590	0.478	0.400	0.132	0.126
Std dev upper non-stem	0.668	0.681	0.447	0.370	0.138	0.095
Std dev upper stem	0.552	0.530	0.458	0.441	0.115	0.092
non-STEM covariance GPA with log enrollment			-0.030	-0.039		
STEM covariance GPA with log enrollment			-0.098	-0.099		
non-STEM covariance γ with log enrollment			0.004	0.002		
STEM covariance γ with log enrollment			0.002	0.006		
non-STEM covariance GPA with γ			-0.016	-0.016		
STEM covariance GPA with γ			-0.016	-0.016		