Submitted to *Econometrica*

EQUILIBRIUM GRADING POLICIES WITH IMPLICATIONS FOR FEMALE INTEREST IN STEM COURSES

TOM AHN
Naval Postgraduate School

PETER ARCIDIACONO
Duke University, NBER, & IZA

AMY HOPSON
Bureau of Labor Statistics

JAMES THOMAS
Federal Trade Commission

We show that stricter grading policies in STEM courses reduce STEM enrollment, especially for women. We estimate a model of student demand for courses and optimal effort choices given professor grading policies. Grading policies are treated as equilibrium objects that in part depend on student demand for courses. Differences in demand for STEM and non-STEM courses explain much of why STEM classes give lower grades. Restrictions on grading policies that equalize average grades across classes reduce the STEM gender gap and increase overall enrollment in STEM classes.

1. INTRODUCTION

The effect of college on human capital is heterogeneous based in part on the courses students take. Human capital in science, technology, engineering, and mathematics (STEM) is

---

Tom Ahn: sahn1@nps.edu
Peter Arcidiacono: psarcidi@econ.duke.edu
Amy Hopson: akhopson@gmail.com
James Thomas: jthomas2@ftc.gov

Previously titled "Equilibrium Grade Inflation with Implications for Female Interest in STEM Majors." The views expressed in this article are those of the authors and do not necessarily reflect those of the US Navy, BLS, or FTC. Ahn thanks Eugenia Toma for her invaluable support in acquiring the data.

perceived to be in short supply, even though jobs in these fields pay substantially more than those in other fields and are more robust to recessions (Altonji et al., 2016).[1] Significant government activity centers on promoting STEM skills to increase their supply.[2] Of particular concern is the lack of female representation in STEM jobs: Women comprise 47% of the workforce but just 30% of STEM workers (U.S. Bureau of Labor Statistics, 2019).

The lack of female representation in STEM occupations is in part due to the lack of female representation in STEM college courses, which in turn may depend on the university environment. At the University of Kentucky (UK), the data source for this study, men take 43% more STEM classes than women. STEM classes are characterized by enrollments that are almost twice as high, grades that are 0.36 points lower, and study times that are 43% higher than non-STEM classes.[3] Furthermore, given existing evidence that women value grades more than men, lower STEM grades may disproportionately deter women from choosing STEM classes (Rask and Bailey, 2002).

In this paper, we examine the effects of grading policies on student demand for STEM courses, with a particular focus on female students. To do this, we estimate a two-sided model of professors choosing linear grading policies in part to influence demand for their courses. The slope of these grading policies dictates the returns to student ability and study time. On the demand side, we specify a model where students choose courses and study time in response to the grading policies of the professors.[4] Modeling both sides of the market allows us to examine the equilibrium effects of policies such as restricting average grades across classes, where professors may respond by changing the returns to studying and where students may respond by changing their course bundles and study times.

We estimate our model using transcript and course evaluation data from UK. Our detailed data and rich utility structure allow us to be quite flexible over student preferences for

---

[1] There is evidence that heterogeneity across fields is increasing over time. See Gemici and Wiswall (2014).

[2] See, for example, Hinz (2019), Stevens (2021).

[3] Other institutions show similar patterns. See Bagues et al. (2008), and Rask (2010) on grading differences and Brint et al. (2012) and Stinebrickner and Stinebrickner (2014) on differences in study time.

[4] Given our focus on course choices, we take the choice of major as given for juniors and seniors and then allow the payoff for a course to depend on whether it fulfills a major requirement. As a result, any policy simulations that affect the choices of juniors or seniors are best interpreted as short-run effects; in the long run, the choice of major may respond to the policy change.

courses in particular departments while still separately identifying how professor grading polices affect student decisions. For example, high ACT math scores may lead to higher grades in engineering classes, but students who have high ACT math scores may also find engineering classes more attractive for reasons not related to grades. The key identifying assumption needed to disentangle preferences for courses and grades is that higher ACT math scores increase the non-grade-related preferences for all engineering classes equally. However, because of the heterogeneity in the slopes of professor grading policies, sorting of students into courses *within* the same department can be used to recover how important grades are in determining student course choices and whether the importance of grades differs by gender.

However, grading policy slopes also affect the returns to studying, potentially confounding whether students value grades more or simply have lower costs of studying. We use self-reported study hours from course evaluations to estimate a model of study effort decisions. The course evaluation data cannot be linked to student transcripts, but they do include student cohort information. This allows us to examine differences in study hours across courses and within courses across cohorts to identify heterogeneity in the costs of study effort by gender and other observed characteristics.

We use the estimates from the model of course choices and study effort to examine the drivers behind the gender gap in STEM course enrollment. Conditional on professor grading policies, two key drivers emerge. The first is that women value grades significantly more than men, suggesting that the lower grades given in STEM courses lower female enrollment in particular. The second, consistent with Griffith (2010), is comparative advantage: Women at UK have higher high school GPAs but lower ACT math scores than their male counterparts. Higher high school GPAs and lower ACT math scores are associated with relatively higher grades in non-STEM courses as well as higher non-grade preferences for non-STEM courses. Other factors, such as female-specific preferences for departments and differences in study costs, play more modest roles.

As STEM courses have grading policies with lower average grades and steeper slopes than those of non-STEM courses, grading policies contribute substantially to the STEM gender gap. Holding fixed the slopes of the professor grading policies but imposing that all courses must have an average grade of a B substantially increases enrollment in STEM courses by both men and women. However, because women value grades more than men, the enrollment increase is markedly higher for women, shrinking the STEM gender gap. Requiring all courses to have both the same average grades and the same slopes decreases the gender gap even more because steeper slopes in STEM amplify the effects of women's comparative advantage in non-STEM courses.

However, grading policies are not fixed parameters—they are choices made by professors in competition with one another. Thus, restricting average course grades may elicit professor responses, mediating the effectiveness of the policy. To incorporate these responses, we explicitly model how instructors choose grading policies to influence demand for their courses in equilibrium. We specify an objective function where professors have individual specific preferences over grades, enrollments, and workloads. Professors that face low innate demand may then raise their grades in an effort to attract more students. Indeed, we find that differences in innate demand account for over 38% of the difference in average grades between STEM and non-STEM courses. For example, biology professors and language professors have, on average, similar preferences over average grades. But because demand for biology courses is so much higher, biology courses have grades that are more than 0.4 points lower than those of language courses.

Given the estimates of both sides of the model, we can then examine how enrollments would change in equilibrium should all courses be required to have a B average. We find that even when we allow professors to change their grading slopes, requiring the same average grade across courses would increase female STEM enrollment by 30.3% and male STEM enrollment by 15.6%.[5] Since our simulations hold the effects of declared majors on upperclassman utility fixed, our estimates are best interpreted as short-run predictions. In the long run, when major choices can vary in response to grading policies, the effects may be substantially larger.

That equalizing average grades across courses would have such large effects may seem surprising. If students and employers have complete information, differences in grading policies should be largely inconsequential. However, this is not the case in limited-information settings (Piopiunk et al., 2020, Chan et al., 2007). Nominal grades may also matter to students for reasons beyond those underscored by traditional economic explanations, such as parental pressure or the psychological assessment of self-worth, a hypothesis

[5] The corresponding numbers when holding grading slopes fixed are 34.2% and 17.9%.

supported by findings in the sociology and psychology literature (Rosenberg et al., 1995, Crocker et al., 2003).

Our paper relates to the literature examining how college students make education decisions. One strand of this literature focuses on the effects of grades on these decisions. It generally finds that low grades reduce persistence in STEM (Astorne-Figari and Speer, 2019, McEwan et al., 2021) and that grades affect educational choices for female students more than for male students (Ost, 2010, Rask and Bailey, 2002, Rask and Tiefenthaler, 2008, Zafar, 2013).[6] Closest to our paper, Butcher et al. (2014) show that imposing a cap on average class grades at Wellesley College increased enrollment in science classes; however, as Wellesley is a women's college, the authors cannot estimate differential effects by gender. In this context, one of our contributions is showing how grading restrictions affect the gender gap in STEM. Moreover, we estimate how much students value grades, how this differs by gender, and how important grades are relative to factors such as instructor gender and gender-specific preferences for departments.

Finally, our paper relates to a growing literature that empirically analyzes supply-side decision-making in higher education. For example, Epple et al. (2006) and Fu (2014) analyze how universities admit students and set tuition, while Thomas (forthcoming) examines the determinants of university course offerings. Our paper contributes to this literature by providing an empirical analysis of how grading policies are set in equilibrium. We thus build on descriptive evidence on the heterogeneity of grading policies over time and across departments (Johnson, 2003, Sabot and Wakeman-Linn, 1991), policy experiments to reduce grading differences (Butcher et al., 2014, Bar et al., 2009), and theoretical work on grading policies (Chan et al., 2007, Zubrickas, 2015).

2. DATA AND DESCRIPTIVE EVIDENCE

In this section, we describe our data and the descriptive analysis that motivates our structural model. We show that STEM courses have higher enrollments, lower grades, and higher study times than non-STEM courses, suggesting that student demand may influence instructors' grading policies. We show that women perform better than men in STEM and

[6]Notable exceptions are Kaganovich et al. (2021) and Kugler et al. (2021), who find that women have stronger responses than men to low grades only in certain academic departments.

non-STEM courses. However, women enroll in substantially fewer STEM courses and sort into classes with higher average grades. These facts motivate an equilibrium model where students choose courses and study times in response to the grading policies, with differential responses to these policies by male and female students.

We focus on undergraduate students at the University of Kentucky (UK) in the fall of 2012. We use three types of data. The first is student-level data on demographics, precollege academic measures, course enrollment, and grade outcomes. The second is course-level data on instructor gender and rank, enrollment caps, prerequisites, and whether the course satisfies other prerequisites or major and university requirements. The third is end-of-semester class evaluation surveys on students' course quality perceptions, expected grades, and the number of hours spent per week studying for the class. The evaluation forms provide no information identifying the students other than their cohort (freshman, sophomore, etc.), so we use class–cohort averages when we match these data to the transcript data.[7]

We focus on classes with at least fifteen students. We aggregate departments into fourteen categories and further aggregate these categories into STEM/non-STEM. We include economics and related fields as part of STEM because courses in these departments exhibit grading patterns and study times similar to those of traditional STEM departments. Our sample includes 55,701 student/course observations from 16,079 students and 1,003 courses. Online Appendix B provides additional details on how we split departments into STEM/non-STEM and select courses for the analysis.

We show summary statistics by gender in Table I. Women at UK arrive with higher high school grades but lower ACT math scores. In the semester we analyze, women have significantly higher grades in both STEM and non-STEM courses, though the gender gap in grades is smaller in STEM courses. They also take fewer courses in STEM: while men take half their courses in STEM departments, the share is a just over a third for women.

Part of the gender gap in STEM course enrollment may be due to different expectations and environments in STEM courses. Table II summarizes course characteristics by STEM status. STEM courses are almost double the size of non-STEM courses and have grades that

[7]We focus on cohort–classes with response rates between 70% and 100%. Response rates can be higher than 100% due to incorrect matching between the transcript and evaluation data. The evaluation data are also incomplete because some courses opt out of participating in the evaluation.

TABLE I

DESCRIPTIVE STATISTICS BY GENDER

	Men	Women	Diff. Men & Women
High school GPA	3.55 (0.56)	3.68 (0.50)	-0.14 (0.01)
ACT Reading Score	26.1 (5.1)	26.0 (4.8)	0.1 (0.1)
ACT Math Score	25.7 (4.6)	23.9 (4.2)	1.7 (0.1)
Fall 2012 GPA	2.86 (0.95)	3.12 (0.86)	-0.26 (0.01)
Fall 2012 GPA|STEM	2.68 (1.06)	2.82 (1.03)	-0.14 (0.02)
Fall 2012 GPA|non-STEM	3.07 (0.96)	3.31 (0.83)	-0.25 (0.02)
Fall 2012 Share Courses in STEM	0.50 (0.35)	0.35 (0.32)	0.15 (0.01)

Note: Fall 2012 University of Kentucky undergraduate students; 7,833 men and 8,246 women. SAT scores are converted to equivalent ACT scores. Standard deviations in parentheses for columns 1 and 2; standard errors in parentheses for column 3. See Online Appendix B.2 for sample selection.

TABLE II

DESCRIPTIVE STATISTICS BY COURSE TYPE

	STEM	Non-STEM	Diff. STEM & Non-STEM
Class Size	82.6 (92.6)	45.0 (59.1)	37.6 (5.9)
Average Grade	2.90 (0.46)	3.26 (0.42)	-0.36 (0.03)
Average Grade | Female	2.98 (0.56)	3.35 (0.43)	-0.38 (0.04)
Average Grade | Male	2.84 (0.47)	3.13 (0.52)	-0.28 (0.03)
Study Hours	3.42 (1.62)	2.39 (0.92)	1.03 (0.14)
Percent Female	0.37 (0.20)	0.58 (0.19)	-0.22 (0.01)
Percent Fem. Prof.	0.28 (0.45)	0.46 (0.50)	-0.18 (0.03)

Note: Fall 2012 University of Kentucky courses; 282 STEM courses and 721 non-STEM courses (for study hours, 139 STEM courses and 402 non-STEM courses). Standard deviations in parentheses for columns 1 and 2; standard errors in parentheses for column 3. See Online Appendix B.2 for sample selection.

are 0.36 points lower.[8] Students spend an extra hour per week (43% more time) studying in STEM courses. STEM courses also have a lower share of female professors.

[8]Classes that have lower grades are associated with other characteristics as well such as lower course evaluations and disproportionately male instructors (see Online Appendix Table B.3). In order to separate out how grading practices themselves drive course decisions separately from these other factors, we use variation in how student characteristics are rewarded across classes in the same department through their grading policies.

To motivate our structural analysis, we run a series of descriptive regressions where grades and study hours are the outcome variables to better understand the driving forces behind the patterns in Table II. Table III presents the results for grades across all classes and students in columns (1) and (2), across only upper level classes in columns (3) and (4), and separately for STEM and non-STEM classes in columns (5) and (6). Columns (1) and (3) control for whether the class is in STEM; remaining columns control for our fourteen department categories.

Across all columns of Table III, the following patterns emerge. First, even with department fixed effects and controls for student preparation, larger class sizes are associated with lower grades. All else equal, we would expect students to prefer classes where they receive higher grades. The fact that larger classes give lower grades could reflect supply considerations. Specifically, classes with low demand could be inflating grades to attract students and classes with high demand could be giving lower grades to deter enrollment.[9] Second, women outperform men across the board, and this is especially true in non-STEM classes. Third, controlling for student gender and baseline preparation, students earn higher grades in classes with a higher fraction of female students. This suggests women disproportionately sort into classes with more lenient grading. Finally, even controlling for academic background and course enrollment, STEM courses give lower grades than non-STEM courses.

Looking at STEM and non-STEM courses separately in columns 5 and 6 of Table III, we see two key patterns. First is that, while women outperform men in STEM classes, the gap is much smaller than in non-STEM classes, potentially reflecting comparative advantage. The second key pattern is test scores—especially ACT math—are much more important in STEM classes than non-STEM classes. In contrast, high school grades have a similar importance across fields.

The descriptive regressions for study hours are limited because we can only use variation at the cohort-class level. We show in Table IV the results of regressions of class–cohort study hours on the average characteristics of the class. We do this for all classes and then separately for elective classes, STEM classes, and non-STEM classes. Both overall and for

[9] Alternatively, if students with lower unobserved ability are more attracted to popular courses—perhaps because of higher labor market returns—these courses might have both higher enrollments and lower grades.

TABLE III

REGRESSIONS OF GRADES ON STUDENT AND CLASS CHARACTERISTICS

	All Classes (1)	(2)	Upper Level (3)	(4)	STEM Classes (5)	Non-STEM Classes (6)
Female	0.102	0.101	0.103	0.099	0.028	0.152
	(0.009)	(0.009)	(0.016)	(0.016)	(0.015)	(0.011)
Percent Female	0.354	0.461	0.434	0.379	1.024	0.176
	(0.025)	(0.028)	(0.042)	(0.047)	(0.049)	(0.034)
ACT Reading*	0.037	0.049	0.044	0.069	0.055	0.043
	(0.005)	(0.005)	(0.009)	(0.009)	(0.008)	(0.006)
ACT Math*	0.133	0.132	0.085	0.080	0.208	0.090
	(0.006)	(0.006)	(0.010)	(0.010)	(0.009)	(0.007)
High School GPA*	0.266	0.268	0.183	0.187	0.275	0.264
	(0.005)	(0.005)	(0.009)	(0.009)	(0.008)	(0.006)
ln(Class Size)	-0.059	-0.037	-0.054	-0.106	-0.067	-0.037
	(0.004)	(0.005)	(0.008)	(0.009)	(0.008)	(0.006)
STEM Class	-0.420		-0.278			
	(0.010)		(0.020)			
Dept. FE	No	Yes	No	Yes	Yes	Yes
Observations	55,701	55,701	15,458	15,458	23,280	32,421

Note: ∗ indicates the variable is z-scored. Additional controls are indicators for upper-level classes and minority and first-generation college students. Regressions (2) and (4)-(6) split STEM/non-STEM into 14 department categories.

each subgroup, the share of the class–cohort that is female is positively correlated with study time. That women study more than men is consistent with survey evidence from Arcidiacono et al. (2012) and Stinebrickner and Stinebrickner (2012).[10] STEM classes have higher study times, though not in elective courses where enrollments are lower.

Perhaps most interesting is the coefficient on average course grade. Courses with higher grades are associated with *less* study time even after conditioning on department fixed effects. This suggests that grades are relative measures of accomplishment and, in conjunction with workload, may be used by professors to influence demand for their courses.

[10]Given women's higher reported study times, there may be a concern that women are also more likely to fill out course evaluation forms. Regressing response rates at the class–cohort level on share female, however, yields a small and insignificant coefficient both with and without course fixed effects.

TABLE IV

REGRESSIONS OF STUDY TIME ON AVERAGE STUDENT AND CLASS CHARACTERISTICS

	All Classes (1)	All Classes (2)	Elective Classes (3)	Elective Classes (4)	STEM Classes (5)	Non-STEM Classes (6)
Female	0.175	0.245	0.109	0.159	0.401	0.226
	(0.071)	(0.070)	(0.099)	(0.097)	(0.194)	(0.073)
ACT Reading*	0.049	0.038	0.183	0.173	-0.308	0.105
	(0.041)	(0.041)	(0.059)	(0.060)	(0.114)	(0.044)
ACT Math*	0.065	0.044	0.016	0.014	0.237	0.034
	(0.048)	(0.048)	(0.063)	(0.065)	(0.139)	(0.050)
High School GPA*	-0.064	-0.070	-0.056	-0.098	-0.048	-0.090
	(0.046)	(0.045)	(0.064)	(0.062)	(0.124)	(0.048)
Average Grade	-0.308	-0.265	-0.276	-0.215	-0.401	-0.240
	(0.044)	(0.046)	(0.070)	(0.071)	(0.112)	(0.050)
ln(Class Size)	-0.124	-0.083	-0.146	-0.158	-0.054	-0.106
	(0.027)	(0.028)	(0.049)	(0.051)	(0.061)	(0.031)
STEM Class	0.362		-0.058			
	(0.050)		(0.084)			
Dept. FE	No	Yes	No	Yes	Yes	Yes
Observations	866	866	346	346	204	662

Note: $*$ indicates that variable is z-scored. Observations are at the class-cohort level. Additional controls are an indicator for upper-level classes, % minority, and % first-generation. Regressions (2) and (4) - (6) split classes into 14 department categories.

3. DEMAND-SIDE MODEL

The descriptive results in Section 2 revealed significant differences in grading and study times across departments. Motivated by these patterns, we next develop a model of how students make course choices and study effort decisions. These decisions are made in part in response to professor grading policies. From the perspective of the student, these grading policies are taken as given. How grading policies are chosen is described in Section 6.

The demand-side model produces three estimating equations. The first is the optimal choice of study effort, which depends on the cost of studying, the extent to which the student values grades, and the incentives provided by professor through their grading policies. The second is the grade production process, which depends on the student's preparation, the optimal choice of study effort, and the professor's grading policies. The final estimating equation comes from the solution to the student's problem of choosing a bundle of courses given his or her preferences over grades, expected optimal study times, and non-grade preferences for particular courses.

3.1. *Choice set*

Student i chooses n_i courses from a subset of all courses $\mathcal{J}_i \subset [1,\dots,J]$, where J is the total number of courses and $\mathcal{J}_i$ is the set of courses i is eligible to take. While our data set contains over 1,000 classes, students are precluded from registering for a substantial fraction of these courses. To account for the restrictions to a student's choice set that arise due to academic and administrative considerations, we use information on course prerequisites, class enrollment capacity constraints, students' course histories from past semesters, and their AP exam results. Accounting for these factors results in students having on average 700 courses in their choice set. See Online Appendix B.3 for a description of the supplemental data that we collected and how we utilized this data to form the choice sets.

3.2. *Course payoffs*

We specify the payoff for a particular course j as dependent on student i's non-grade preference for the course, δ_{ij}, the amount of study effort that he or she chooses to exert in the course, s_{ij}, and the course grade conditional on study effort, $g_{ij}(s_{ij})$. Following Nevo et al. (2005), we assume that 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 those of the other courses in the bundle.[11] The individual's realized utility from choosing course j and exerting s_{ij} units of effort is given by:

$$U_{ij}(s_{ij}) = \phi_i g_{ij}(s_{ij}) - \psi_{ij} s_{ij} + \delta_{ij} \tag{1}$$

[11] For 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. A natural concern with our setup is that students may balance hard courses with easier ones. In Appendix A.2, we show that our estimated model matches both the within-student distribution of high-workload courses and the within-student distribution of STEM courses.

We parameterize ϕ_i as dependent on the student's gender. Denoting w_i as an indicator for whether the student is female, $\phi_i = \phi_0 + \phi_1 w_i$. The costs of studying in course j, ψ_{ij}, are specified to depend on w_i and a set of characteristics X_i, defined in Table V. The costs of studying also depend on a shock ζ_{ij} that is revealed after the student has chosen his or her courses, implying that students form expectations over the realizations of ζ_{ij} when making their course choices. In particular, we specify ψ_{ij} as:

$$\psi_{ij} = \zeta_{ij}\psi_i = \zeta_{ij}\exp\left(\psi_0 + w_i\psi_1 + X_i\psi_2\right) \tag{2}$$

where ζ_{ij} is log-normally distributed.

Preferences for courses net of grades and study costs, δ_{ij}, depend on the characteristics of the student and the course. Each course belongs to some department $k \in [1, \ldots, K]$, where $k(j)$ gives the department for the jth course. Denote as Z_{1i} the set of variables that affect preferences for courses in particular departments. For example, students with high ACT math scores may prefer courses in physics for reasons above and beyond how ACT math scores affect physics grades. Denote as Z_{2ij} the set of variables that affect the match between the student and the course. This includes factors such as whether a course satisfies prerequisites for freshmen or major requirements for upperclassmen. The full listing of what is included in Z_{1i} and Z_{2ij} is shown in Table V.

We then parameterize δ_{ij} as:

$$\delta_{ij} = \delta_{0j} + w_i\delta_{1k(j)} + Z_{1i}\delta_{2k(j)} + Z_{2ij}\delta_3 + \epsilon_{ij} \tag{3}$$

where δ_{0j} are course fixed effects and ϵ_{ij} is i.i.d. type 1 extreme value. Women's preferences for the course material and the climate in particular departments are captured by $\delta_{1k(j)}$.

Substituting the parameterizations of ϕ_i, ψ_{ij} and δ_{ij} into (1) yields:

$$\begin{aligned} U_{ij}(s_{ij}) = {} & (\phi_0 + \phi_1 w_i)g_{ij}(s_{ij}) - \zeta_{ij}\exp\left(\psi_0 + w_i\psi_1 + X_i\psi_2\right)s_{ij} \\ & + \delta_{0j} + w_i\delta_{1k(j)} + Z_{1i}\delta_{2k(j)} + Z_{2ij}\delta_3 + \epsilon_{ij} \end{aligned} \tag{4}$$

Estimating department preferences, preferences for grades, and study costs separately by student gender helps us uncover some of the driving forces behind the gender gap in STEM. For example, if women's intrinsic demand for courses in STEM departments is relatively

TABLE V

LIST OF CONTROLS

Gender (w_i): affects department-specific academic preparation and preferences, study costs, & grade preferences
Covariates for academic preparation and cost of study effort (X_i): ACT reading & math, high school grades, minority, first generation, unobserved type
Covariates for preferences that vary by department (Z_{1i}): ACT reading & math, high school grades, unobs. type
Covariates for preferences that vary by class match (Z_{2ij}): female $\times$ female professor; freshman and sophomore $\times$ STEM $\times$ female; (juniors and seniors) whether the course is required for the major, whether it is one of two or more courses that would fill a major requirement, whether the course is upper division; (sophomores) log number of courses opened up by taking the course, STA210; (freshmen) log number of courses opened up by taking the course, CIS/WRD110

Note: Opened-up courses are ones where the particular course is a prerequisite; STA210 and CIS/WRD110 are university core requirements typically taken during the sophomore and freshman years respectively.

low ($\delta_{1k(j)}$ negative) while the preferences for grades and cost of effort are relatively equal across men and women (ϕ_1 and ψ_1 close to zero), then changing grading policies would have little effect on the gender gap in STEM. On the other hand, if women have significantly different preferences over grades and study effort than men, then altering grading policies could affect the gender composition of classes and departments.

Note that one of the components of both X_i and Z_{1i} is the student's unobserved type. This is time-invariant private information (not observed by the econometrician) that the student has regarding (i) their department-specific preferences, (ii) their department-specific abilities, and (iii) their study costs. It is through these unobserved types that we account for selection on unobservables. Identification and estimation in the presence of this unobserved heterogeneity is discussed in Section 4.4.

3.3. *Grades*

The grade that student i receives in course j, g_{ij}, depends in part on student i's academic preparation for course j. We allow academic preparation to vary across departments. For example, ACT math scores may be important for math classes but less so for English classes. Academic preparation for class j in department k is then given by a department-specific weighted average of the student's characteristics, X_i. Note that these are the same characteristics that affect study costs and are given in Table V.

In addition to academic preparation for the course, g_{ij} depends on the student's study effort, s_{ij}, and a mean-zero shock that is unknown to the individual at the time of course enrollment, η_{ij}. Given study effort s_{ij}, g_{ij} is specified as:

$$g_{ij}(s_{ij}) = \beta_j + \gamma_j \left(w_i \alpha_{1k(j)} + X_i \alpha_{2k(j)} + \ln(s_{ij}) \right) + \eta_{ij} \tag{5}$$

Professors' grading policies are then choices over an intercept, β_j, and a slope, γ_j, that dictate the returns to academic preparation and effort. Gains from study effort enter as a log to capture the diminishing returns to studying.

3.4. *Study effort*

Students are assumed to know professors' grading policies.[12] After we substitute the grading process (5) into the utility function (4), the optimal study effort given a realization of ζ_{ij} can be found by differentiating $U_{ij}(s_{ij})$ with respect to s_{ij}:

$$s^*_{ij} = \frac{\phi_i \gamma_j}{\psi_{ij}} = \frac{(\phi_0 + \phi_1 w_i)\gamma_j}{\zeta_{ij} \exp\left(\psi_0 + w_i \psi_1 + X_i \psi_2\right)} \tag{6}$$

A higher γ_j increases the incentives to study. Those who value grades more (have higher values of ϕ_i) and have lower study costs (lower values of ψ_{ij}) also exert more effort.

Equation (6) gives us our first estimating equation, linking grading policies and student characteristics to study effort. Substituting the optimal choice of study effort given (6) into (5) yields our second estimating equation, which is the grade production process:

$$\begin{aligned} g_{ij} = \beta_j + \gamma_j \left[w_i(\alpha_{1k(j)} - \psi_1) + X_i(\alpha_{2k(j)} - \psi_2) + \ln(\phi_0 + \phi_1 w_i) + \ln(\gamma_j) - \psi_0 \right] \\ + \eta_{ij} - \gamma_j \ln(\zeta_{ij}) \end{aligned} \tag{7}$$

[12]Students can learn about grading policies from friends, course syllabi, or from publicly available course evaluations that include average expected grades. See also Ferreyra et al. (2021) for a model of student effort as a function of college policies.

3.5. *Course choices*

An important assumption is that study time is chosen optimally after the realization of the shock to study costs, ζ_{ij}, but that ζ_{ij} is unknown at the time of course selection. The shock to grades, η_{ij}, is also unknown at the time when courses are chosen. Hence, individuals maximize the expected utility of their course bundle taking into account their optimal response to the realizations of the ζ_{ij}s. Taking expectations over ζ_{ij} and η_{ij} in (7) gives the expected grades that individuals use when forming their expectations over course payoffs:

$$E(g_{ij}) = \beta_j + \gamma_j \left[w_i(\alpha_{1k(j)} - \psi_1) + X_i(\alpha_{2k(j)} - \psi_2) + \ln(\phi_0 + \phi_1 w_i) + \ln(\gamma_j) - \psi_0 \right] \tag{8}$$

After substituting in the optimal study effort responses from (6) and the corresponding expected grades given in (8) into (1) and taking expectations, the expected utility of course j can be written as:

$$\mathbb{E}(U_{ij}) = (\phi_0 + \phi_1 w_i)\left(E(g_{ij}) - \gamma_j)\right] + \delta_{0j} + w_i\delta_{1k(j)} + Z_{1i}\delta_{2k(j)} + Z_{2ij}\delta_3 + \epsilon_{ij} \tag{9}$$

Let $d_{ij} = 1$ if j is one of the n_i courses chosen by student i and zero otherwise. Students then solve the following maximization problem when choosing their optimal course bundle:

$$\max_{d_{i1},\ldots,d_{iJ}} \sum_{j=1}^{J} d_{ij}\mathbb{E}(U_{ij}) \quad \text{subject to:} \sum_{j=1}^{J} d_{ij} = n_i \tag{10}$$

where n_i is taken as given.[13] We then obtain our third estimating equation by solving the maximization problem in (10).

4. DEMAND-SIDE ESTIMATION

The model in the previous section was characterized by three sets of equations governing (i) the grade production process, (ii) the optimal choice of study effort, and (iii) student class choices. We now describe the estimation of the model and the identification assumptions.

[13] Although some students may adjust the number of courses they take in counterfactual scenarios, n_i is only weakly related to other observed variables. For example, the correlation between the number of classes and each of female, HS gpa, ACT reading, and ACT math all lie between -0.015 and 0.03.

For expositional clarity, we begin with the case where there are no unobserved types. Key to the identification arguments is the sequential revelation of new information. In particular, new information on the costs of studying are revealed after the course choice decisions are made. Further, new information on grade realizations is revealed after course choices and study decisions. With each piece of new information assumed to be uncorrelated with the others and absent unobserved heterogeneity, the model reduces to one of selection on observables.

We then describe identification and estimation in the case with unobserved types in Section 4.4. Finally, given the strong assumptions made in the model and estimation, we discuss the implications for our results should these assumptions be violated and develop tests for whether certain violations would lead us to miss key data moments.

4.1. *Reduced-form grade equation*

We begin with the estimation of the grade process. Equation (7) yields the following reduced form:

$$g_{ij} = \theta_{0j} + \gamma_j \left(w_i \theta_{1k(j)} + X_i \theta_{2k(j)} \right) + \eta^*_{ij} \tag{11}$$

where

$$\theta_{0j} = \beta_j + \gamma_j (\ln(\phi_0) + \ln(\gamma_j) - \psi_0) \tag{12}$$

$$\theta_{1k(j)} = \alpha_{1k(j)} + \ln(\phi_0 + \phi_1) - \ln(\phi_0) - \psi_1 \tag{13}$$

$$\theta_{2k(j)} = \alpha_{2k(j)} - \psi_2 \tag{14}$$

$$\eta^*_{ij} = \eta_{ij} - \gamma_j \ln(\zeta_{ij}) \tag{15}$$

The reduced-form parameters $\{\theta_{0j}, \theta_{1k(j)}, \theta_{2k(j)}\}$ and the structural slopes, the γ_js—both relative to a normalization—can be estimated by means of nonlinear least squares. A normalization must be made for every department, as scaling up the θs by some factor and scaling down the γs by the same factor would be observationally equivalent. We set one γ_j equal to one for each department, a normalization that will be undone in Section 4.2. Denote as C_k the normalization for department k. We then estimate γ_j^N, where $\gamma_j^N = \gamma_j / C_{k(j)}$. Similarly, we estimate $\theta^N_{1k(j)}$ and $\theta^N_{2k(j)}$, where $\theta^N_{1k(j)} = \theta_{1k(j)} C_{k(j)}$ and $\theta^N_{2k(j)} = \theta_{2k(j)} C_{k(j)}$.

4.2. *Reduced-form study equation*

The course evaluation data give reported study hours for each individual in the classroom, and we use these study hours as our measure of effort, s^*_{ij}. Taking logs of (6) yields:

$$\ln(s^*_{ij}) = \kappa_0 + w_i\kappa_1 - X_i\psi_2 + \ln(\gamma_j) - \ln(\zeta_{ij}) \tag{16}$$

where

$$\kappa_0 = \ln(\phi_0) - \psi_0 \tag{17}$$

$$\kappa_1 = \ln(\phi_0 + \phi_1) - \ln(\phi_0) - \psi_1 \tag{18}$$

Recall that one γ_j for every department was normalized in the grade equation. Substituting in $\hat{\gamma}^N_j C_{k(j)}$ for γ_j in (16) and rearranging yields:

$$\ln(s^*_{ij}) - \ln(\hat{\gamma}^N_j) = \tilde{\kappa}_0 + w_i\kappa_1 - X_i\psi_2 + \kappa_{2k(j)} - \ln(\zeta_{ij}) \tag{19}$$

where $\kappa_{2k(j)} = \ln(C_{k(j)}/C_1)$ and $\tilde{\kappa}_0 = \kappa_0 + \ln(C_1)$. Here, C_1 is the normalized course for the base department.

The course evaluation data cannot be linked to the individual data on grades and academic preparation. However, the evaluation data do provide information about the cohort of the evaluator (i.e., freshman, sophomore, junior, or senior). The observations that we use in estimating the choice of study effort are then at the class–cohort level. Letting l_i indicate the cohort of student i and averaging (19) at the class-cohort level yields our estimating equation:

$$\frac{\sum_i (l_i = l) d_{ij} \ln(s^*_{ij})}{\sum_i (l_i = l) d_{ij}} - \ln(\hat{\gamma}^N_j) = \tilde{\kappa}_0 + w_{jl}\kappa_1 - X_{jl}\psi_2 + \kappa_{2k(j)} - \ln(\zeta_{jl}) \tag{20}$$

where w_{jl}, X_{jl}, and $\ln(\zeta_{jl})$ are the averages of w_i, X_i, and $\ln(\zeta_{ij})$ for students of cohort l enrolled in course j. $\ln(\zeta_{jl})$ is unobserved and assumed to be uncorrelated with w_{jl} and X_{jl}.

The estimates of (20) allow us to recover how observed characteristics (other than gender) affect study costs, $\hat{\psi}_2$. We can also partially undo the normalization on the γs, solving for γs that are normalized with respect to one course rather than to one course in each department. Namely, let $\gamma_j^P = \gamma_j^N \exp(\kappa_{2k(j)})$. $\hat{\gamma}_j^P$ then provides an estimate of γ_j / C_1.

The last normalization—the returns to preparation and study time in the only remaining normalized course—can be recovered from the heteroskedasticity in the grading residuals that result from the study cost shocks, $\ln(\zeta_{ij})$, mattering more in classes where γ_j is high. In particular, the residual from Equation (11) can be written as:

$$g_{ij} - E(g_{ij}) = -C_1 \gamma_j^P \ln(\zeta_{ij}) + \eta_{ij} \tag{21}$$

Since $\ln(\zeta_{ij})$ is assumed to be independent of η_{ij}, we can express the variance of the residuals for class j, σ_j^2, as:

$$\sigma_j^2 = (\gamma_j^P)^2 \kappa_3 + \sigma_\eta^2 \tag{22}$$

where $\kappa_3 = C_1^2 \sigma_{\ln(\zeta)}^2$. Regressing σ_j^2 on $(\gamma_j^P)^2$ then gives us an estimate of C_1^2 up to the variance of the study cost shock.

We can recover an estimate of $\sigma_{\ln(\zeta)}^2$ using the residuals of Equation (20) for cohort–course combinations with one student. Let this estimate be given by $\hat{\sigma}_{\ln(\zeta)}^2$. However, because we observed the study times only in grouped intervals, there will be measurement error in these residuals. We simulate data from the continuous study time process (so that the hours are not lumped into bins), using the cohort–class data employed and the corresponding parameter estimates from Equation (20) and where the $\ln(\zeta_{ij})$s are drawn from a normal distribution with mean zero and variance calibrated so that, after censoring into bins, the censored residuals have variance equal to $\hat{\sigma}_{\ln(\zeta)}^2$. This calibrated variance then allows us to recover C_1.

4.3. *Estimation of utility parameters*

We now turn to estimating the utility function parameters. Recall that the expected utility from taking course j in Equation (9) was given by:

$$\mathbb{E}(U_{ij}) = (\phi_0 + \phi_1 w_i)\left(E(g_{ij}) - \gamma_j)\right] + \delta_{0j} + w_i \delta_{1k(j)} + Z_{1i}\delta_{2k(j)} + Z_{2ij}\delta_3 + \epsilon_{ij}$$

The estimates of expected grades and γs follow from the previous steps. With these taken as given, the variation in the data that identifies ϕ_0 and ϕ_1 comes from how individuals sort within departments based on their comparative advantage in grades. For example, the extent to which men with high math ability sort into math classes where math ability is especially rewarded identifies ϕ_0; the difference in this sorting behavior between men and women identifies ϕ_1.

We use simulated maximum likelihood coupled with a nested fixed-point algorithm to estimate the choice parameters.[14] To illustrate the approach, denote as K_i the set of courses chosen by i. Denote $M_i = \max_{j \notin K_i}\{\mathbb{E}(U_{ij})\}$ as the highest payoff among the unchosen courses. Suppose that K_i consisted of courses $\{1,2,3\}$ and that the values for all the preference shocks ϵ_{ij} were known with the exception of those for $\{1,2,3\}$. The probability of choosing $\{1,2,3\}$ could then be expressed as:

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

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

Since the ϵ_{ij}s for the unchosen 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 as M_{ir} the value of M_i at the rth draw of the unchosen ϵ_{ij}s and as R the number of simulation draws, the full log-likelihood function is given by[15]:

$$\ln \mathcal{L} = \sum_i \ln\left(\left[\sum_{r=1}^{R}\prod_{j=1}^{J}\left(1 - G(M_{ir} - \overline{U}_{ij})\right)^{d_{ij}}\right]/R\right) \tag{23}$$

While in theory one could estimate all of the choice parameters δ_{0j}, $\delta_{1k(j)}$, $\delta_{2k(j)}$, δ_3, ϕ_0, and ϕ_1 by solving for the parameter values that maximize Equation (23), the large number of courses makes doing so computationally challenging. To circumvent this issue,

[14] Nesting a fixed-point algorithm within a maximum likelihood routine follows Berry et al. (1995).

[15] Our setup is similar to Nevo et al. (2005). Nevo et al. (2005) randomly samples rankings of chosen options, computes likelihood contributions conditional on rankings, and averages across the sampled rankings to simulate the full likelihood. We simulate the stochastic utility of the best unchosen course, compute the likelihood contributions conditional on this stochastic utility, and average across simulation draws to simulate the full likelihood.

in the spirit of Berry et al. (1995), we nest a fixed-point algorithm within the maximization routine that matches estimates of the course-specific intercepts δ_{0j} directly to data on enrollment shares. Details of the algorithm can be found in Online Appendix C.2. Once the parameters of Equation (23) have been estimated, it is straightforward to recover the remaining structural parameters ψ_0, ψ_1, β_j, and $\alpha_{1k(j)}$ from combinations of previously identified parameters.

4.4. *Estimation with unobserved heterogeneity*

We now consider the case when one of the components of X_i and Z_{1i} 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 sth value. In practice, we set S to three.

Identification of the unobserved types comes from the within-student correlation across grades and choices after conditioning on observables. Unobserved types affect study costs, department-specific preferences, and department-specific abilities. Differences between observed grade outcomes and those predicted by the observables help pin down department-specific abilities. For example, a student may perform better than expected based on observables in all his or her STEM classes, implying a comparative advantage in STEM. Students may also sort into STEM classes as a whole more than what his or her expected grades and observed characteristics would predict, implying an unobserved preference for these classes as well. Coupled with the structure of the model where type classification is also determined by grades and course choices, differences in study times at the course-cohort level help pin down the effect of the unobserved types on study costs.

Integrating out over this missing component removes 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 and Jones (2003) and Arcidiacono and Miller (2011), we can restore separability via a modified expectations maximization (EM) algorithm. Our full estimation procedure is described in Online Appendix C.1.

4.5. *Discussion*

While we view our model of course level choices with flexible utility, flexible grade production, and endogenous study effort as a substantial innovation relative to existing literature, our model still requires a number of important assumptions that may affect our conclusions. In Appendix A, we discuss some of these assumptions in detail and perform specification tests to assess robustness to these assumptions. Our general approach is to predict how violations of certain assumptions would lead to poor model fit for particular moments of the data. We then assess model fit for these moments to examine whether our assumptions lead us to poorly approximate student behavior.

For example, our model does not allow students to balance difficult courses with easy courses when choosing course bundles. This suggests we might over-predict the number of students choosing mostly difficult courses or mostly easy courses. However, when we simulate choices for all students, calculate the share of difficult classes that each student takes, and compute the standard deviation in this share across students, we find that this standard deviation is actually slightly lower than the standard deviation in the data. This implies that our model closely approximates the course balancing behavior observed in our data. Details of our specification tests are provided in Appendix A.

5. DEMAND-SIDE ESTIMATES

5.1. *Preference estimates*

Table VI presents a subset of the preference parameters. While both men and women value grades, women derive substantively higher utility from higher grades. The results show that women value grades 28% more than men. Consistent with Carrell et al. (2010), female students prefer classes with female professors, with the estimate suggesting that women would have to be compensated by a little over one-tenth of a grade point to enroll in the same class taught by a male professor.[16]

In the first column of the second panel of Table VI, we show women's preferences (relative to men's) for different departments, with the omitted category being Agriculture. The

[16]The coefficient may be biased upward due to aggregation of departments. If female professors are more likely to be in departments that women prefer and that variation exists in aggregated groups, we may be capturing within-group department preferences.

largest difference in preferences is between Engineering and Biology, at 1.4, which for women translates to 1.18 grade points. Engineering and Biology are outliers, with the difference in the penultimate categories (Psychology and Communications) translating to 0.47 grade points.

These departmental preferences of women emerge after we account for grade considerations and sorting on other ability factors.[17] Education & Health, where women make up almost 70% of course enrollments, are shown to be preferred to a similar extent as Chemistry & Physics, where women make up less than half of enrollments. The primary driver of women into Education & Health over Chemistry & Physics is the difference in grades and the matching of observed characteristics to characteristics of the department.

We show in the second to fourth columns of Table VI how non-grade preferences for classes in particular departments vary by academic characteristics. The most salient result is the strong positive correlation between ACT math scores and preferences for STEM departments.[18] Since men at UK on average have higher ACT math scores (and lower high school grades) than their female counterparts, this too contributes to more men choosing STEM classes above and beyond the fact that higher ACT math scores are especially rewarded in the grading policies of STEM classes.

5.2. *Study effort estimates*

Estimates of the study cost parameters are presented in Table VII. Women have 5.4% lower study costs than men, though the estimate is not statistically significant. Conditional on having the same observed characteristics and taking the same class, women study 30% more than men; however, our estimates of ϕ_0 and ϕ_1 imply that over 80% of their increased studying is due to preferences for grades.

We show in the second set of columns how the returns to study effort vary across classes, taking the median γ class for each course grouping. The heterogeneity is quite large, with classes in STEM departments having the highest returns to studying. A doubling of study

[17] See Jacob et al. (2018), Kaganovich et al. (2021), Wiswall and Zafar (2015), and Zafar (2013) for other examples in the literature exploring non-grade preferences for departments or majors.

[18] The one exception is mathematics, which may be due in part to students with lower mathematics skills being required to take additional remedial classes to satisfy general university requirements.

TABLE VI

ESTIMATES OF SELECTED STUDENT PREFERENCE PARAMETERS

Preference for:		Coef.	Std. Error	
Expected grades (ϕ_0)		0.927	(0.006)	
Female x expected grade (ϕ_1)		0.257	(0.009)	
Female x female professor		0.141	(0.007)	
Departments	Female	ACT read*	ACT math*	HS GPA*
Biology	0.504	-0.238	0.033	-0.136
Psychology	0.162	-0.375	0.047	-0.202
Education & Health	0.139	-0.336	0.010	0.075
English	0.127	0.024	-0.208	0.051
Chem. & Physics	0.069	-0.177	0.143	-0.178
Mgmt. & Mkting	-0.049	-0.243	0.138	-0.081
Regional Studies	-0.091	-0.231	-0.062	-0.057
Math	-0.133	-0.154	-0.193	-0.200
Languages	-0.159	-0.116	0.042	-0.154
Social Sciences	-0.350	-0.175	-0.010	-0.182
Econ., Fin., Acct.	-0.378	-0.295	0.131	-0.040
Communications	-0.401	-0.247	0.036	-0.090
Engineering	-0.897	-0.275	0.494	0.048

Note: See Online Appendix Table B.5 for complete results. $*$ indicates that the variable is z-scored. *Female* is women's non-grade preference for departments relative to men's. Department preferences are relative to agriculture. STEM departments are in bold.

effort would translate into an increase of 0.44 grade points in Engineering but would be less than one-third as effective in increasing grades in Education & Health.

5.3. *Grade estimates*

The estimated department-specific ability weights, the αs, are given in Table VIII. The departments are sorted such that those with the highest *Female* estimate are listed first.

The coefficients on *Female* in the first column suggest that women have a comparative advantage in non-STEM departments after differences in test scores and high school grades are accounted for. This result makes sense in the context of the descriptive statistics presented in Table II: Women have higher grades than men in both STEM and non-STEM

TABLE VII

ESTIMATES OF STUDY COSTS AND DEPARTMENTAL RETURNS TO STUDYING

	Study Cost Coef. (ψ)	Std. Error	Department	Median γ Coef.
Female	-0.054	(0.085)	**Engineering**	0.442
ACT Reading Score*	-0.056	(0.050)	**Biology**	0.308
ACT Math Score*	0.064	(0.058)	**Math**	0.271
High School GPA*	0.098	(0.056)	**Econ., Fin., Acct.**	0.259
			Psychology	0.253
			Chem. & Physics	0.241
			Regional Studies	0.233
			English	0.224
			Languages	0.194
			Communications	0.190
			Social Sciences	0.187
			Agriculture	0.168
			Mgmt. & Mkting	0.146
			Education & Health	0.142

Note: $*$ indicates that the variable is z-scored. Study costs also depend on minority and first-generation status and unobserved type. STEM departments are in bold. Departments are sorted by their median value of γ.

classes, but the gap is smaller in STEM classes. Given that the returns to studying are higher in STEM classes and that women study more than men, we would expect women to substantially outperform men in STEM classes should women not have a comparative advantage in non-STEM courses.

We show in the second through fourth columns the ability weights on the two components of the ACT and high school grades. The returns to the different components of the ACT score are intuitive. The five STEM categories have five of the six highest returns to the ACT math score, with the highest return found in Math classes. Higher returns to ACT reading are found in Social Sciences, Psychology, English, and Languages.

5.4. *Drivers of the STEM gap*

Given the estimates of the grading process and students' choices over classes and study time, we now examine the sources of the male–female gap in STEM enrollment. We focus

TABLE VIII

ESTIMATES OF DEPARTMENT-SPECIFIC ABILITY WEIGHTS (α)

	Female		ACT read*		ACT math*		HS GPA*	
Education & Health	0.522	(0.164)	0.164	(0.064)	0.407	(0.077)	0.750	(0.119)
Communications	0.372	(0.170)	0.167	(0.067)	0.167	(0.051)	0.960	(0.210)
Regional Studies	0.361	(0.153)	0.071	(0.072)	0.660	(0.105)	1.020	(0.143)
Agriculture	0.172	(0.196)	0.223	(0.106)	0.613	(0.130)	1.121	(0.176)
Psychology	-0.055	(0.100)	0.415	(0.064)	0.517	(0.065)	0.940	(0.078)
English	-0.072	(0.160)	0.296	(0.103)	0.453	(0.110)	1.006	(0.141)
Languages	-0.140	(0.105)	0.311	(0.085)	0.530	(0.102)	0.929	(0.153)
Social Sciences	-0.199	(0.089)	0.480	(0.088)	0.408	(0.074)	1.082	(0.141)
Math	-0.244	(0.063)	-0.074	(0.036)	1.594	(0.133)	0.984	(0.084)
Mgmt. & Mkting	-0.309	(0.124)	0.166	(0.097)	0.440	(0.127)	0.985	(0.246)
Biology	-0.439	(0.078)	0.166	(0.047)	0.648	(0.076)	0.827	(0.084)
Engineering	-0.442	(0.070)	-0.018	(0.029)	0.631	(0.067)	0.362	(0.040)
Econ., Fin., Acct.	-0.547	(0.087)	0.146	(0.053)	0.980	(0.117)	0.917	(0.098)
Chem. & Physics	-0.708	(0.080)	0.042	(0.044)	1.286	(0.075)	1.165	(0.066)

Note: $*$ indicates that the variable is z-scored. STEM departments are in bold. Departments are sorted by women's α. Standard errors in parentheses.

our attention on freshmen and sophomores because junior and seniors have already chosen their majors.[19] In all simulations, we change the parameters or characteristics for women to match the parameters or characteristics for men.[20]

[19] Juniors and seniors change their choices in these partial equilibrium counterfactual scenarios because counterfactual choices by freshmen and sophomores alter which courses are capacity constrained. However, these changes are generally very small because most juniors and seniors register before freshmen and sophomores and thus are not exposed to the effects of freshmen's and sophomores' choices on capacity constraints.

[20] Similar to male juniors and seniors, male freshmen and sophomores change their choices in these counterfactual scenarios only because the counterfactual choices by female freshmen and sophomores alter which course are capacity constrained. These effects are small, so we omit them for the sake of brevity.

We show in Table IX how the share of classes taken in STEM for women and men change as different characteristics are equalized across genders.[21] We also report the difference between the male and female shares as a measure of the gender gap in STEM participation. The first two rows of Table IX show that our model matches the data well. The model-predicted shares of STEM classes for men and women are 53.3% and 41.0%, respectively. The 12.3-percentage-point gap between the two model-predicted shares is what we use as our baseline when comparing the drivers of the STEM gender gap.

TABLE IX

STEM ENROLLMENT FOR FRESHMEN AND SOPHOMORES IN COUNTERFACTUAL SCENARIOS (PARTIAL EQUILIBRIUM)

		STEM Enrollment Share		
		Female	Male	STEM gap
(1)	Data	40.9%	53.3%	
(2)	Baseline model	41.0%	53.3%	12.3
(3)	Equalize grade preferences	45.0%		8.3
(4)	Shift obs. abil. incl. abil. tastes	43.8%		9.5
(5)	Shift unobs. abil. in grades	45.2%		8.1
(6)	Equalize unobs. pref. for depts.	40.7%		12.5
(7)	Female professor effect turned off	41.3%		12.0
(8)	Grade around a B: $\gamma_j = 0$	62.3%	65.5%	3.2
(9)	Grade around a B: $\gamma_j = 0.21$	60.7%	67.9%	7.1
(10)	Grade around a B: $\gamma_j = \hat{\gamma}_j$	55.6%	63.9%	8.2

Note: Women's preference and ability parameters are adjusted to be identical to men's preferences and abilities. Counterfactuals are partial equilibrium, as the grading policies of professors are held fixed.

The predicted outcomes when women's preferences for grades are changed to be the same as men's (ϕ_1 is set to zero) are shown in the third row. Equalizing grade preferences increases the share of classes that women take in STEM to 45.0%. This reduces the gender

[21]Our counterfactual simulations hold the utilities of courses with $\gamma < 0.01$ fixed. There were twenty-five such courses. See Online Appendix C.3 for how the counterfactual choice probabilities are calculated in the presence of capacity constraints.

gap in STEM by almost a third. This reduction arises both because STEM courses have lower grades and because women have a comparative advantage in non-STEM courses; lowering the value of grades weakens the importance of this comparative advantage.

The fourth and fifth rows change the observed and unobserved abilities so that the distribution is the same for men and women. The observed abilities affect both grades and the non-grade department preferences. Because men have higher math ACT scores and this makes STEM classes more attractive both through grades and through the non-grade department preferences, equalizing observed abilities reduces the gender gap by 2.8 percentage points. Even stronger effects from equalizing unobserved ability are presented in the fifth row, where the gender gap is reduced by 4.2 percentage points. We find that women have a comparative advantage in non-STEM courses beyond what is associated with observable characteristics such as test scores. Because women value grades more, removing these relative advantages makes STEM courses significantly more attractive to women and reduces the gender gap accordingly.

The next two counterfactuals (rows six and seven) equalize women's unobserved preferences for departments and remove women's preferences for female instructors, respectively. Both changes barely move the gender gap (0.3 percentage points in either direction). Overall, we find that the non-grade preferences not already accounted for through other background measures are relatively unimportant to the STEM gap. However, the small effect of equalizing unobserved preferences for departments masks larger movements within STEM, lowering female participation in Biology and raising it in Engineering and Economics.

The final set of rows in Table IX examines how a standardized curve policy would affect both overall STEM enrollment and the gender gap in STEM. We consider a policy in which the average grade in each course must be a B; however, at this point, we cannot predict how professors would adjust both β_j and γ_j to comply with this policy. As such, we perform simulations that fix γ_j at particular values and adjust β_j to satisfy the curve. Simulations that allow professors to endogenously adjust both β_j and γ_j will be discussed in Section 6.

We perform three simulations using different sets of values for the γ_js. First, we set γ_j to zero for all classes. This is equivalent to removing grades from the utility function altogether. Second, we set γ_j to the median value across all courses (0.21). Finally, we fix γ_j at their estimated values. In all scenarios, we find that a standardized curve substantially increases STEM enrollment and substantially decreases the gender gap in STEM. The effects are generally larger in the more standardized scenarios where γ_j are equated across classes. For example, with γ_j set to zero, the curve policy would increase female STEM enrollment by 52% and decrease the gender gap in STEM to 3.2 percentage points (around a quarter of the size of the original gap).

In sum, we find three primary drivers of the gender gap in STEM. First, women have a comparative advantage in non-STEM courses. Second, women value grades more than men, exacerbating the effects of this comparative advantage. Finally, lower grades in STEM courses play a substantial role in limiting STEM enrollment, and this is especially true for women. This last finding suggests that policies that lead to more uniform grading may work to close the gender gap in STEM. We examine how professors may respond to restrictions on grading policies in the next section.

6. EQUILIBRIUM GRADING POLICIES

Section 5 revealed that differences in grading policies across departments influence course choices and contribute to the gender gap in STEM. It also revealed large differences in grades and workloads across departments. In this section, we develop and estimate a model of how professors set their grading policies in equilibrium. Doing so serves two purposes. First, it allows us to show the role that differences in demand for courses plays in differences in grading policies. Second, it shows the scope that professors have to undo the effects of policy changes by changing their behavior along other dimensions. The particular policy change that we consider is a policy restricting average grades to be the same across courses or subsets of courses.

6.1. *Reduced-form evidence of the effect of enrollment on grading policies*

We begin by providing reduced-form evidence that higher course enrollments result in professors both giving lower grades and assigning more work (higher γs). Consider regressions of average courses grades, $\overline{G}_j$, and workloads, γ_j, on log enrollment, $\ln(E_j)$, and course and faculty characteristics, W_j:

$$\overline{G}_j = W_j \vartheta_{G1} + \ln(E_j)\vartheta_{G2} + \varepsilon_{Gj} \tag{24}$$

$$\gamma_j = W_j \vartheta_{\gamma 1} + \ln(E_j)\vartheta_{\gamma 2} + \varepsilon_{\gamma j} \tag{25}$$

Log enrollment is endogenous, and the OLS estimate of its coefficient captures both how enrollment affects grades through grading policies and how grading policies affect enrollment. To account for the endogeneity of log enrollment, we instrument for it by using predicted log enrollment when all classes have the same grading policies. In practice, we set β_j and γ_j to the median values across all classes and then use our structural model to predict course enrollments. Clearly, this instrument satisfies the relevance requirement, as classes with large course fixed effects (δ_j) will have higher enrollments. The exogeneity assumption requires that innate instructor leniency be uncorrelated with innate course demand (as captured by δ_j but also the other demand determinants) after the characteristics of the course given in W_j are accounted for. While this assumption is not testable, we see similar results from instrumenting instead with predicted enrollment when (i) grades are the same in all courses ($\beta_j, \gamma_j = 0$) or (ii) all coefficients in the utility function are turned off with the exception of the course fixed effects (δ_j).

We estimate three versions of Equations (24) and (25): (i) one without the control for log enrollment, (ii) one with the control for log enrollment, and (iii) one instrumenting for log enrollment. We restrict our analysis to those classes where the capacity constraint is not met and where $\gamma_j > 0$.[22] We show in Table X the coefficients on log enrollment in each regression and how removing the effects of log enrollment affects the average gap between STEM and non-STEM courses in the outcome.[23] Averaging across courses shows that STEM classes have grades that are 0.34 points lower than the grades of their non-STEM counterparts (column 1). Accounting for the endogeneity of log enrollment using our instrument (column 3) reduces this gap to 0.16 points and leads to increases in the magnitude of the effect of log enrollment by a factor of over 5 relative to the OLS estimate.

We show in the second set of columns of Table X that differences in enrollment (after reverse causality is removed) are part of the explanation for why STEM classes have higher workloads. The standard deviation of γ_j is 0.12, implying that the average value of γ for STEM courses is over one standard deviation larger than that of non-STEM courses (column 4). Accounting for the endogeneity of log enrollment using our instrument (column 6) reduces this gap to 0.089 points. Together, these results suggest that higher demand

[22]Recall that twenty-five courses had γ_js that hit the zero constraint.

[23]The remaining parameters are shown in Online Appendix B.8.

for courses leads to lower grades and higher workloads, and that differences in demand partially explain differences in grading policies between STEM and non-STEM courses.

TABLE X

RELATING COURSE DEMAND TO GRADES AND WORKLOADS

	Average Grades			γ		
	Baseline	OLS	IV	Baseline	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
non-STEM	0.340	0.307	0.164	-0.128	-0.124	-0.089
Ln Enroll		-0.066	-0.352		0.007	0.077
		(0.020)	(0.023)		(0.004)	(0.005)

Note: The analysis is at the course level. Estimates are from 951 courses where $\gamma_j > 0$ and the course capacity constraint does not bind. Additional controls include department fixed effects, rank of the instructor, female instructor interacted with STEM and upper-level class interacted with STEM. See Online Appendix Table B.7 for the coefficients on the additional controls.

6.2. *The professor's problem*

We next develop a model of professor choices that is designed to produce estimating equations similar to those in Equations (24) and (25). We assume that professors choose grading policy parameters β_j and γ_j to maximize an objective function that depends on (i) the number of students in their class, (ii) the grades given in the course, and (iii) the cost of assigning work (γ).[24] In particular, we specify the professor's objective function to penalize deviations from their ideal log enrollment, e_{0j}, their ideal average grade in the class, e_{1j}, and their ideal workload, e_{2j}. These ideals depend on observed and unobserved characteristics of the professor.

We specify the objective function this way in part because these are the measures that we observe in the data. Preferences over enrollments and workloads may relate to learning outcomes that the professor values but also impose time costs on the professor through

[24]We also estimate models where professors exert effort to directly affect demand for courses. Incorporating professor effort has little effect on our counterfactual results, though measuring professor effort is difficult. See Online Appendix D for the model with professor effort, the description of how professor effort is measured, and the model results with this channel incorporated.

increased student interaction and development and grading (or supervision of teaching assistants in the administration and grading) of assignments and exams. Having the professor directly value class grades (beyond their impact on enrollment) may reflect departmental norms and the desire to avoid student complaints.

Denote as $\overline{G}_j(\beta,\gamma)$ the expected average grade in class j given the grading policies for all courses (β and γ). Denote as $P_{ij}(\beta,\gamma)$ the probability that i takes course j given the grading policies. $\overline{G}_j(\beta,\gamma)$ and log enrollment in course j are given by:

$$\overline{G}_j(\beta,\gamma) = \beta_j + \gamma_j \left[\frac{\sum_i^N P_{ij}(\beta,\gamma)\left[A_{ij} + \ln(\phi_i) - \ln(\psi_i)\right]}{\sum_i^N P_{ij}(\beta,\gamma)} + \ln(\gamma_j) \right] \tag{26}$$

$$\ln\left[E_j(\beta,\gamma)\right] = \ln\left[\sum_i^N P_{ij}(\beta,\gamma)\right] \tag{27}$$

The objective function that professor j maximizes is then:

$$V_j(\beta,\gamma) = -\left(\ln\left[E_j(\beta,\gamma)\right] - e_{0j}\right)^2 - \lambda_1\left(\overline{G}_j(\beta,\gamma) - e_{1j}\right)^2 - \lambda_2(\gamma_j - e_{2j})^2 \tag{28}$$

where the coefficient on ideal log enrollment is normalized to one.[25]

Professors choose grading policies given different innate demand for their courses. Absent the first term, the professor of course j would set γ_j to e_{2j}. Given γ_j, the professor would then set β_j such that expected grades would equal e_{1j}. However, with the first term, professors deviate from their ideal grades and workloads to mitigate the costs associated with having classes that are not the ideal size. If demand for a course would be above (below) e_{0j} when grades and effort were set to their ideal levels, professors adjust grades (workloads) downward (upward) to move enrollment closer to the ideal.

[25] The coefficient on one of the squared terms must be normalized to identify the model. As normalizing one of these coefficients to one is a monotonic transformation of the underlying utility function, the normalization has no implications for the counterfactual policy analysis.

6.3. *Estimation*

We use the first-order conditions of the professor's objective function to form our estimating equations. After dividing by two, these are given by:[26]

$$0 = -\left(\ln\left[E_j(\beta,\gamma)\right] - e_{0j}\right)\frac{\partial \ln E_j}{\partial \beta_j} - \lambda_1\left(\overline{G}_j(\beta,\gamma) - e_{1j}\right)\frac{\partial \overline{G}_j}{\partial \beta_j} \tag{29}$$

$$0 = -\left(\ln\left[E_j(\beta,\gamma)\right] - e_{0j}\right)\frac{\partial \ln E_j}{\partial \gamma_j} - \lambda_1\left(\overline{G}_j(\beta,\gamma) - e_{1j}\right)\frac{\partial \overline{G}_j}{\partial \gamma_j} - \lambda_2(\gamma_j - e_{2j}) \tag{30}$$

We allow for heterogeneity across professors in their preferences through the e_{lj}s, specifying e_{lj} as:

$$e_{lj} = W_{lj}\Psi_l + \varepsilon_{lj} \tag{31}$$

where ε_{lj} is unobserved professor-specific tastes for the lth outcome. Given that there are two first-order conditions and three ε terms and that we cannot recover three unobservables from two equations, we normalize ε_{0j} to zero. When l refers to enrollment, we specify W_{lj} to include a constant term and an indicator for whether the course is upper division, with the latter allowing instructors to prefer lower enrollments in upper-division courses. When l refers to grades or workload, we specify W_{lj} to include course category fixed effects, rank of the instructor, whether the instructor is female, and whether the course is upper division. The indicators for whether the instructor is female and whether the course is upper division are also interacted with STEM.

The unobserved preferences ε_{1j} and ε_{2j} in part determine the optimal choice of β_j and γ_j. The rest of this section shows how we obtain estimates of the parameters given the endogeneity of the grading policies. As in Section 6.1, the key identification assumption is that the unobserved professor preferences for grades and workload are uncorrelated with innate demand for the courses that they teach after we condition on W_{lj}.

---

[26] Note that when capacity constraints bind, the first term in each of the expressions is zero. In this case, professors set their grades, workloads, and effort to their ideal levels. As a result, we do not use courses where capacity constraints bind in the estimation. Given the parameter estimates, we can, however, back out the corresponding unobserved preference terms using Equations (29) and (30) with the first terms of each set to zero.

6.3.1. *Recovering* Ψ_0, Ψ_2, *and* λ_2

We estimate the professor parameters in two steps. In the first step, we estimate Ψ_0 (preferences for ideal enrollment), Ψ_2 (preferences for ideal workload), and λ_2 (the weight on the ideal workload). Rearranging and differencing the first-order conditions given in Equations (29) and (30) to eliminate the λ_1 term and solving for γ_j, we obtain:

$$\gamma_j = (1/\lambda_2)\ln[E_j]\,A_j - \Psi_0 A_j + W_{2j}\Psi_2 + \varepsilon_{2j} \tag{32}$$

where A_j is given by:

$$A_j = \left[\frac{\partial \ln[E_j]}{\partial \beta_j}\frac{\partial \overline{G}_j}{\partial \gamma_j} \Big/ \frac{\partial \overline{G}_j}{\partial \beta_j}\right] - \frac{\partial \ln[E_j]}{\partial \gamma_j} \tag{33}$$

Both the first and second terms of Equation (32) are correlated with ε_{2j}, as ε_{2j} affects enrollment through γ_j and the corresponding derivatives of enrollment and grades with respect to γ_j. We create instruments for these two terms by using the course choice model to evaluate A_j and $A_j \ln[E_j]$ at common values of β^0 and γ^0. In practice, we set β^0 and γ^0 to the median values across all courses. The variation across classes is then driven by the innate demand for courses given fixed grading policies.

6.3.2. *Recovering* Ψ_1 *and* λ_1

In the second step, we recover estimates of Ψ_1 (preferences for ideal grades) and λ_1 (the weight on ideal grades). Equation (29), the first-order condition with respect to β_j, can be rewritten as:

$$\overline{G}_j(\beta,\gamma) = -(1/\lambda_1)B_j\left(\ln[E_j] - \Psi_0\right) + W_{1j}\Psi_1 + \varepsilon_{1j} \tag{34}$$

where Ψ_0 is known from step 1 and B_j is given by:

$$B_j = \left[\frac{\partial \ln[E_j]}{\partial \beta_j} / \frac{\partial \overline{G}_j}{\partial \beta_j}\right]$$

We then instrument for $B_j\left(\ln[E_j] - \Psi_0\right)$ following the procedure in step 1, evaluating B_j and $\ln[E_j]$ with the policy parameters set to β^0 and γ^0 in each course.

6.4. *Supply-side results*

We show in Table XI the estimates of the professor preference parameters. Professors who teach upper-level classes prefer higher grades, lower workloads, and smaller class sizes relative to those teaching lower-level classes, perhaps reflecting the more specialized nature of these courses. The estimates imply that it is more often the case that professors raise grades to attract students than lower grades to deter students. Among lower-division classes, 9.6% are above the ideal size of 146. Among upper-division classes, 25.9% are above the ideal size of 40.

There is also heterogeneity in the grading practices based on instructor rank and gender. Tenured and tenure-track faculty prefer lower grades than lecturers. Instructors who are not on the tenure track may have an incentive to offer higher grades, as their contracts may depend on teaching evaluations, which in turn rise with expected grades (see Online Appendix Table D.2). Female professors have higher ideal grades than their male counterparts, though the differences are smaller in STEM departments.

Department-specific parameters are listed from highest ideal grades to lowest. Professors in the Management & Marketing and Education & Health departments on average have the highest ideal grades, while professors in the English and Mathematics departments have the lowest. Higher ideal grades are also generally associated with lower ideal workloads. Note that these ideal grades and workloads are not driven by direct student demand for courses; they may be set by norms in the department, perhaps following the lead of senior faculty or influenced by instructors' own experiences as undergraduates.

While Table XI reveals how professors would prefer to assign grades and workloads, student demand for courses induces deviations from these ideals to achieve enrollments closer to e_{0j}. One can see these demand adjustments directly in the rearranged first-order conditions given in Equations (32) and (34). The first terms of each equation show how deviations from ideal enrollments (demand adjustments) impact professor choices.

Table XII reports averages of these demand adjustment terms by department relative to the average across all courses. Demand adjustments are sorted from lowest to highest based on the adjustment to grades. In response to higher student demand, STEM departments (along with the Psychology and Management & Marketing departments) give lower grades and assign higher workloads than non-STEM departments. The difference in grades

TABLE XI

ESTIMATES OF PROFESSOR PREFERENCES

	Ideal grade		Ideal workload		Ideal log enrl	
λ (Preference Weight)	2.969	(0.374)	46.068	(3.988)	1.000	—
Constant	2.595	(0.102)	0.273	(0.035)	4.981	(0.592)
Upper-Level Class	0.417	(0.056)	-0.069	(0.035)	-1.283	(0.491)
Upper-Level X STEM	-0.050	(0.066)	0.076	(0.019)		
Grad. Student	-0.028	(0.043)	0.007	(0.011)		
Lecturer	0.104	(0.045)	-0.013	(0.011)		
Asst. Prof.	-0.060	(0.049)	0.027	(0.013)		
Tenured Prof.	-0.068	(0.040)	0.002	(0.010)		
Female Prof.	0.091	(0.030)	0.002	(0.008)		
Female Prof. X STEM	-0.040	(0.060)	-0.009	(0.016)		
Management & Marketing	0.344	(0.086)	-0.048	(0.022)		
Education & Health	0.297	(0.066)	-0.020	(0.017)		
Communications	0.128	(0.060)	0.009	(0.016)		
Regional Studies	0.051	(0.073)	0.045	(0.020)		
Biology	0.004	(0.114)	0.061	(0.027)		
Language	-0.003	(0.066)	0.035	(0.017)		
Engineering	-0.003	(0.080)	0.209	(0.023)		
Psychology	-0.009	(0.099)	0.048	(0.025)		
Econ., Fin., Acct.	-0.038	(0.096)	0.004	(0.024)		
Chemistry & Physics	-0.099	(0.090)	0.011	(0.022)		
Social Science	-0.109	(0.063)	0.013	(0.017)		
English	-0.177	(0.080)	0.090	(0.022)		
Math	-0.235	(0.077)	0.071	(0.019)		

Note: The weight on ideal log enrollment, λ_0, is normalized to 1. The base professor rank category is adjunct instructors contracted by the course. Lecturers are offered longer-term contracts and are salaried. See Online Appendix Table D.3 for a specification that includes professor effort. STEM departments are in bold. The baseline department is Agriculture.

between Biology and English (the two extremes) due to demand factors is about 0.47 grade points. English has the lowest ideal grades among all departments except Math (Table XI) yet offers grades around the median in equilibrium due to the relatively low demand for English courses. In contrast, Biology is close to the median on ideal grades yet gives substantially lower grades due to the high demand for Biology courses.

TABLE XII

DEMAND ADJUSTMENTS RELATIVE TO THE MEAN

	Grades	Workload
Biology	-0.293	0.052
Psychology	-0.258	0.051
Econ., Fin., Acct.	-0.213	0.039
Management & Marketing	-0.198	0.038
Chemistry & Physics	-0.105	0.018
Engineering	-0.056	0.026
Math	-0.036	0.003
Education & Health	-0.017	0.005
Social Science	-0.010	-0.009
Agriculture	0.035	-0.005
Communications	0.079	-0.012
Language	0.133	-0.026
Regional Studies	0.161	-0.013
English	0.182	-0.057

Note: Demand adjustments are calculated from the first terms of Equations (32) and (34).

Using the results in Tables XI and XII, we can decompose the gaps in grades and workloads between STEM and non-STEM courses into the contributions from student demand and professor preferences. In particular, we examine the share of the gap in grades and workloads that is due to differences in i) demand (Table XII), ii) level of course offerings (Table XI rows 3–4), iii) rank of the instructor (Table XI rows 5–8), iv) female professor representation (Table XI rows 9–10), and v) departmental effects (Table XI rows 11–23). The results are presented in Table XIII.

We show in the first column of Table XIII how demand factors vary across STEM and non-STEM courses. These are calculated by weighting the department demand adjustments in Table XII by the number of courses in each STEM and non-STEM department. As we show in the first panel, differences in demand result in STEM grades being 0.15 grade points lower than non-STEM grades. This represents 38% of the average difference between STEM and non-STEM course grades. Differences in demand account for a similar share of the differences in workloads across STEM and non-STEM departments, as seen in the second panel.

The results in the next set of columns come from calculating how the components of ideal grades and workloads given in Table XI vary by department. More upper-division classes are offered in non-STEM departments. This, coupled with upper-level classes having higher ideal grades, accounts for 11.5% of the difference between STEM and non-STEM grades. Despite the substantial heterogeneity in ideal grades across instructor rank, this accounts for very little of the differences between STEM and non-STEM courses. Differences in female representation, coupled with women in non-STEM fields having higher ideal grades, accounts for an additional 4.4%. Department-specific intercepts account for the remaining difference, representing a slightly higher share than student demand.

The patterns for workloads, shown in the second panel, also highlight the importance of student demand, which accounts for over 24% of the STEM/non-STEM gap. Upper-level non-STEM classes have lower ideal workloads. This, coupled with the greater upper-level courses offerings in non-STEM departments, accounts for 23% of the gap. More important are department norms, captured by the department-specific intercepts, which account for over 50% of the gap. Engineering is the primary driver of this last result, as it is an outlier on the ideal workload (see Table XI).

TABLE XIII

DECOMPOSING STEM/NON-STEM DIFFERENCES IN GRADES AND WORKLOADS

	Demand Adjust	Upper-Level Class	Faculty Rank	Female Faculty	Dept. Prefs.	Total Effect
STEM grade	-0.1084	-0.0326	-0.0079	-0.0123	-0.1129	-0.2832
Non-STEM grade	0.0434	0.0131	0.0031	0.0049	0.0452	0.1134
Diff.	0.1519	0.0457	0.0110	0.0173	0.1582	0.3966
Shares	38.30%	11.53%	2.77%	4.35%	39.88%	
STEM workload	0.0222	0.0212	-0.0003	-0.0003	0.0472	0.0911
Non-STEM workload	-0.0089	-0.0085	0.0001	0.0001	-0.0189	-0.0365
Diff.	-0.0310	-0.0296	0.0004	0.0005	-0.0662	-0.1276
Shares	24.31%	23.21%	-0.33%	-0.37%	51.85%	

Note: The decompositions of STEM/non-STEM differences in grades and workloads (γ) are calculated by averaging across courses the Ψ estimates of department-category intercepts, instructor rank, upper-level class, and female professor from Table XI and the demand-side adjustments calculated in Table XII.

6.5. *General equilibrium counterfactuals*

With estimates of professor preferences, we can examine how equilibrium grading practices would change in counterfactual scenarios. We focus on two counterfactual scenarios: one where the average grade is set to a B in all courses and another where only lower division courses are subject to this grading policy. The former mirrors the partial equilibrium case considered in the last row of Table IX but now allows professors to respond to the policy by adjusting workloads. Note that all counterfactuals hold the choice of major by juniors and seniors fixed. They should therefore be interpreted as short-run results, with larger long-run impacts likely to occur as students adjust their majors.[27]

The counterfactual results are presented in Table XIV. The first row shows the data. Row 2 shows the partial equilibrium effects of setting average grades in each course to a B but fixing the γ_js at their estimated values. These results are equivalent to those in the last row of Table IX but now including juniors and seniors. The share of STEM classes increases for men and women by 8.9 and 11.8 percentage points, respectively. We then show in row 3 what happens when professors are able to partially undo the effects of the policy by changing their workloads (γ_js). Professor responses to the policy lower the increases for men and women to 7.7 and 10.5 percentage points. However, the effects on STEM enrollment—especially for women—remain large. Finally, row 4 contains results equivalent to those in row 3 but where the curve applies only to lower-division courses. Since over 80% (65%) of STEM (non-STEM) enrollment is in lower-division courses, the effects remain large, with STEM enrollment dropping by only 1.3 percentage points for women and 0.9 percentage points for men relative to STEM enrollment under a curve that affects all courses.

7. CONCLUSION

The number of STEM graduates—especially from underrepresented groups—has been an ongoing concern. At the same time, STEM courses are on average associated with lower grades and higher study times, both of which may deter enrollment. Using administrative data from the University of Kentucky, we estimate a model of course choices to understand

[27] The number of previous courses taken in a department also affects the payoffs of taking courses offered in that department. Hence we may expect larger long run responses from our counterfactual policies from changes in sophomore course histories as well.

TABLE XIV

COUNTERFACTUALS WITH ENDOGENOUS PROFESSOR RESPONSES

	Class Size		STEM Enroll. Share		Weighted Avg. Grade	
	STEM	Non-STEM	Female	Male	STEM	Non-STEM
(1) Baseline	82.6	45.0	34.6%	49.5%	2.756	3.207
(2) Grade around a B (PE)	103.1	36.9	46.4%	58.4%	3.000	3.000
(3) Grade around a B (GE)	100.6	37.9	45.1%	57.2%	3.000	3.000
(4) Grade around a B (GE, lower div)	98.4	38.8	43.8%	56.3%	2.996	3.096

Note: "Weighted" means weighting by class enrollment. "PE" indicates partial equilibrium and is equivalent to the last row in Table IX but for all students. "GE" indicates general equilibrium and allows professors to change their grading policies. While in Table IX we show the STEM shares for freshmen and sophomores, the results here are for all students. See Online Appendix Table D.4 for the changes to γ and counterfactuals evaluated when professors directly influence enrollment.

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 that this is particularly true for female students.

One issue with policies aimed at reducing grading differences is that instructors may respond 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 as equilibrium objects chosen by instructors in competition with one another. Taking into account these equilibrium responses, we show that a policy of curving all courses around a B would increase male STEM participation by 7.7 percentage points (15.7% increase) but would increase female STEM participation by 10.5 percentage points (30.3% increase). Changing grading practices to mitigate large departmental differences in average grades then results in substantial increases in STEM enrollment and a shrinking of the gender gap.

There are at least two reasons why our estimates likely understate the long-run effects of equalizing average grades across classes. First, our counterfactual holds the choice of major fixed for juniors and seniors; later cohorts will also be able to respond to the policy by shifting into STEM majors. Second, the shifting composition of STEM classes toward more women may have a positive feedback effect by changing the climate of the classes.

Weighed against these positive effects, increases in the supply of STEM majors may result in lower wage premiums for STEM majors, partially undoing the effects of the policy.

REFERENCES

ALTONJI, JOSEPH G, LISA B KAHN, AND JAMIN D SPEER (2016): "Cashier or consultant? Entry labor market conditions, field of study, and career success," *Journal of Labor Economics*, 34 (S1), 361–401. [2]

ARCIDIACONO, PETER, ESTEBAN M AUCEJO, AND KEN SPENNER (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. [9]

ARCIDIACONO, PETER AND JOHN BAILEY JONES (2003): "Finite mixture distributions, sequential likelihood and the EM algorithm," *Econometrica*, 71 (3), 933–946. [20]

ARCIDIACONO, PETER AND ROBERT A. MILLER (2011): "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, 79 (6), 1823–1868. [20]

ASTORNE-FIGARI, CARMEN AND JAMIN D SPEER (2019): "Are changes of major, major changes? The roles of grades, gender, and preferences in college major switching," *Economics of Education Review*, 70, 75–93. [5]

BAGUES, MANUEL, MAURO SYLOS LABINI, AND NATALIA ZINOVYEVA (2008): "Differential Grading Standards and University Funding: Evidence from Italy," *CESifo Economics Studies*, 52 (2), 149–176. [2]

BAR, TALIA, VRINDA KADIYALI, AND ASAF ZUSSMAN (2009): "Grade information and grade inflation: The Cornell experiment," *Journal of Economic Perspectives*, 23 (3), 93–108. [5]

BERRY, STEVEN, JAMES LEVINSOHN, AND ARIEL PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63 (4), 841–890. [19, 20]

BRINT, STEVEN, ALLISON M CANTWELL, AND PREETA SAXENA (2012): "Disciplinary categories, majors, and undergraduate academic experiences: Rethinking Bok's "Underachieving Colleges" thesis," *Research in Higher Education*, 53 (1), 1–25. [2]

BUTCHER, KRISTIN F, PATRICK J MCEWAN, AND AKILA WEERAPANA (2014): "The Effects of an Anti-Grade-Inflation Policy at Wellesley College," *Journal of Economic Perspectives*, 8 (3), 189–204. [5]

CARRELL, SCOTT E, MARIANNE E PAGE, AND JAMES E WEST (2010): "Sex and science: How professor gender perpetuates the gender gap," *Quarterly Journal of Economics*, 125 (3), 1101–1144. [21]

CHAN, WILLIAM, LI HAO, AND WING SUEN (2007): "A Signaling Theory of Grade Inflation," *International Economic Review*, 48 (3), 1065–1090. [4, 5]

CROCKER, JENNIFER, RIIA K. LUHTANEN, M. LYNNE COOPER, AND ALEXANDRA BOUVRETTE (2003): "Contingencies of Self-Worth in College Students: Theory and Measurement," *Journal of Personality and Social Psychology*, 85 (6), 894–908. [5]

EPPLE, DENNIS, RICHARD ROMANO, AND HOLGER SIEG (2006): "Admission, tuition, and financial aid policies in the market for higher education," *Econometrica*, 74 (4), 885–928. [5]

FERREYRA, MARIA, CARLOS GARRIGA, AND ANGELICA SANCHEZ DIAZ (2021): "Raising College Access and Completion: How Much Can Free College Help?" Borradores de Economia 1155. [14]

FU, CHAO (2014): "Equilibrium tuition, applications, admissions, and enrollment in the college market," *Journal of Political Economy*, 122 (2), 225–281. [5]

GEMICI, AHU AND MATTHEW WISWALL (2014): "Evolution Of Gender Differences In Post-Secondary Human Capital Investments: College Majors," *International Economic Review*, 1 (55), 23–56. [2]

GENTZKOW, MATTHEW (2007): "Valuing new goods in a model with complementarity: Online newspapers," *American Economic Review*, 97 (3), 713–744. [11]

GRIFFITH, AMANDA 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. [3]

HINZ, GREG (2019): "STEM students would get free tuition under Il. bill," *Crain's Chicago Business*. [2]

INSLER, MICHAEL, ALEXANDER F. MCQUOID, AHMED RAHMAN, AND KATHERINE SMITH (2021): "Fear and Loathing in the Classroom: Why Does Teacher Quality Matter?" IZA Discussion Papers 14036. [60]

JACOB, BRIAN, BRIAN MCCALL, AND KEVIN M STANGE (2018): "College as Country Club: Do Colleges Cater to Students' Preferences for Consumption?" *Journal of Labor Economics*, 4 (1), 309–348. [22]

JOHNSON, VALEN E (2003): *Grade inflation: A crisis in college education*, Springer. [5]

KAGANOVICH, M., M. TAYLOR, AND R. XIAO (2021): "Gender Differences in Persistence in a Field of Study," CESifo Working Paper No. 9087. [5, 22]

KUGLER, A.D., C.H. TINSLEY, AND O. UKHANEVA (2021): "Choice of Majors: Are Women Really Different from Men?" *Economics of Education Review*, 81, 1–19. [5]

MCEWAN, PATRICK J., SHERIDAN ROGERS, AND AKILA WEERAPANA (2021): "Grade Sensitivity and the Economics Major at a Women's College," *AEA Papers and Proceedings*, 11, 102–106. [5]

NELSON, JON P AND KATHLEEN A LYNCH (1984): "Grade inflation, real income, simultaneity, and teaching evaluations," *Journal of Economic Education*, 21–37. [60]

NEVO, AVIV, DANIEL L RUBINFELD, AND MARK MCCABE (2005): "Academic journal pricing and the demand of libraries," *American Economic Review*, 95 (2), 447–452. [11, 19]

OST, BEN (2010): "The role of peers and grades in determining major persistence in the sciences," *Economics of Education Review*, 29 (6), 923–934. [5]

PIOPIUNK, M., G. SCHWERDT, L. SIMON, AND L. WOESSMANN (2020): "Skills, Signals, and Employability: An Experimental Investigation," *European Economic Review*, 123, 1–25. [4]

RASK, KEVIN (2010): "Attrition in STEM fields at a liberal arts college: The importance of grades and pre-collegiate preferences," *Economics of Education Review*, 29 (6), 892–900. [2]

RASK, KEVIN AND JILL TIEFENTHALER (2008): "The role of grade sensitivity in explaining the gender imbalance in undergraduate economics," *Economics of Education Review*, 27 (6), 676–687. [5]

RASK, KEVIN N AND ELIZABETH M BAILEY (2002): "Are faculty role models? Evidence from major choice in an undergraduate institution," *Journal of Economic Education*, 33 (2), 99–124. [2, 5]

ROSENBERG, MORRIS, CARMI SCHOOLER, CARRIE SCHOENBACH, AND FLORENCE ROSENBERG (1995): "Global Self-Esteem and Specific Self-Esteem: Different Concepts, Different Outcomes," *American Sociological Review*, 60 (1), 141–156. [5]

SABOT, RICHARD AND JOHN WAKEMAN-LINN (1991): "Grade inflation and course choice," *Journal of Economic Perspectives*, 159–170. [5]

STEVENS, REP. HALEY (2021): "How Congress can help get more women and girls involved in STEM education," *The Hill*. [2]

STINEBRICKNER, RALPH AND TODD STINEBRICKNER (2014): "A major in science? Initial beliefs and final outcomes for college major and dropout," *Review of Economic Studies*, 81 (1), 426–472. [2]

STINEBRICKNER, RALPH AND TODD R STINEBRICKNER (2012): "Learning about Academic Ability and the College Dropout Decision," *Journal of Labor Economics*, 30 (4), 707–748. [9]

THOMAS, JAMES (forthcoming): "What do course offerings imply about university preferences?" *Journal of Labor Economics*. [5]

U.S. BUREAU OF LABOR STATISTICS (2019): "Employed persons by detailed occupation, sex, race, and Hispanic or Latino ethnicity - Labor Force Statistics from the Current Population Survey, Table 11," . [2]

WISWALL, M. AND B. ZAFAR (2015): "Determinants of College Major Choice: Identification using an Information Experiment," *Review of Economic Studies*, 82 (2), 791–824. [22]

ZAFAR, BASIT (2013): "College major choice and the gender gap," *Journal of Human Resources*, 48 (3), 545–595. [5, 22]

ZANGENEHZADEH, HAMID (1988): "Grade inflation: A way out," *Journal of Economic Education*, 217–226. [60]

ZUBRICKAS, ROBERTAS (2015): "Optimal grading," *International Economic Review*, 3 (56), 751–776. [5]

APPENDIX A: APPENDIX – FOR INCLUSION IN TEXT

We discuss some assumptions made to estimate the demand side and provide validity tests by comparing the simulated vs. actual course choices. To simulate course choices,

1. we simulate preference shocks ϵ_{ij} and compute choice utilities U_{ij} for all student class combinations,
2. beginning with the student with the earliest timestamp, we take the n_i courses with the highest choice utility,
3. we repeat step 2 for each student based on the ordering of timestamps and removing courses from students' subsequent choice sets once the capacity constraint is reached.

A.1. *Knowledge of the grading process*

We assume that students know the grading policies for every class (γ_j and β_j). In reality, they may be uncertain about the true parameters. The model can incorporate limited forms of uncertainty. For example, if a student were equally overoptimistic in all courses, then the overoptimism would cancel out in the choice problem. If some students are better informed about policies in particular classes than others, this may have implications for what the model predicts vs. what is in the data. For instance, upperclassmen may sort better than lowerclassmen, and women may be more informed about classes with more women.

If upperclassmen are better informed about grading policies, we would over- (under-) predict how well lower- (upper-) classmen sort into classes matching their abilities. To test this, we simulate course choices using model estimates and calculate expected grades using the simulated and actual choices. If expected grades in the simulated courses were on average higher than those in actual courses chosen by lowerclassmen, this would imply upperclassmen are better informed. However, the difference between the actual and simulated expected grades for freshmen, sophomores, and upperclassmen are all less than 0.01.[28]

As a central finding of our model is that women value grades more than men, women may have better information in courses where women comprise a greater share of enrollment. We may expect women to better match their abilities for these classes than what the model predicts. To test this, we calculate average expected grades for women at the class level based on their actual and simulated choices. We then difference at the class level women's actual and model-simulated grades and regress this measure on the share of women in the class. The coefficient on share women is small and negative (-0.019). If women were better informed about grading processes in classes that more women take, we would expect this relationship to be positive.

A.2. *Balancing effort across classes*

Key to the tractability of our model is the assumption that the utility from a course does not depend on the other chosen courses. This rules out students balancing workloads (γ) with a mix of easy and hard classes. To test this assumption, we simulate the model to predict the share of classes that each student takes above the median γ class and calculate the SD across students. If the model overpredicts this SD, it would be evidence of balancing of workloads: the model would be overpredicting the number of students who take all hard or all easy classes. The model-predicted SD (0.308) is actually slightly lower than SD in

[28]The actual minus simulated differences for freshmen, sophomores, and upperclassmen are 0.003, -0.001, and -0.009, respectively.

the data (0.320). We also predict the share of classes that each student takes in STEM, with STEM serving as a proxy for classes that require more work. As in the previous test, we find no evidence that the model misses students balancing STEM/non-STEM. For the share of STEM, the model-predicted SD is 0.328 versus 0.343 in the data.

A.3. *Flexibility of gender effects*

Although our model includes a variety of mechanisms to explain the gender gap in STEM, there may be features not captured by the model for key STEM courses that deter women from enrolling. To investigate this possibility, we simulate course choices and examine how the share of women varies across different subsets of simulated and actual courses. We find that across a range of specifications, the simulated female share is within one percentage point of the actual share.[29] This suggests that our model is capturing how women and men are distributed across courses in different departments.

Finally, women may perform better in classes taught by women, a feature not allowed for in our model. We estimate a version of our model which includes a female student time female professor term in grade production. While the coefficient is positive, it is also small at 0.03. Allowing for this interaction has little effect on our counterfactuals, though it does slightly reduce the estimated preference that females have for female instructors because part of this effect is now transmitted through grades.

[29] We compare simulated vs. actual enrollment in lower-division classes, large classes ($\geq$ 100 students), STEM classes minus biology, underclassmen sample only, and all combinations.

APPENDIX B: DATA AND RESULTS APPENDIX – ONLINE

In this section, we describe how we processed the data and show additional results for the various estimation stages. The appendix covers:

1. our method of aggregating the departments into our fourteen categories,
2. our sample selection procedures for the various estimation stages,
3. construction of student choice sets,
4. descriptives on how course grading is related to other outcomes,
5. additional parameters from the motivating regressions on grades and hours studied (Tables III and IV)
6. additional structural parameters (expanding on Table VI),
7. the share of courses taken in each department by unobserved type, and
8. additional parameters from the regressions of average grades and workload on enrollment and instructor characteristics (expanding on Table X).

B.1. *Aggregation of departments*

In Table B.1 we show the aggregation of departments into our fourteen categories. We partitioned departments into these categories by first grouping departments by their school organization. UK consists of the Colleges of Agriculture, Arts and Sciences, Business and Economics, Communication and Information, Design, Education, Engineering, and Fine Arts. Within the colleges, departments were further grouped based partly on shared core requirements and cross-listed coursework. Finally, some departments were manually extracted (e.g., Psychology has its own category) or inserted into a category (e.g., all fine arts departments were subsumed under Communications), mostly due to department size.

B.2. *Sample selection*

We now describe our sample selection rules for the various stages of estimation. We restrict courses to those that have enrollment of at least 15 undergraduates. This cuts the 2,026 classes observed in the population to 1,084. The total number of individual–course observations resulting from this cut is 58,081 with 16,190 unique students. We then remove specialized classes that would result from taking a second course in a sequence, as the decision process is very different for these courses. The restriction we impose is that at

TABLE B.1

AGGREGATION OF DEPARTMENTS

Category	Departments
Agriculture	Ag. Biotechnology, Ag. Economics, Ag. Ed, Ag. General, Animal & Food Sciences, Biosystems & Ag. Engineering, Environmental Studies, Forestry, Landscape Architecture, Plant Pathology, Plant & Soil Sciences, Sustainable Ag.
Regional Studies	Appalachian Studies, Family Sciences, Gender & Women's Studies, Hispanic Studies, Latin American Studies
Communications	Arts Admin, Comm., Comm. & Info Studies, Fine Arts – Music, Theatre Arts, Schl of Journalism & Telecomm, Schl of Art & Visual Studies, Schl of Interior Design
Education & 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	Chem. & Materials Engineering, Civil Engineering, Com. Sci., Electrical & Com. Engineering, Engineering, Mech. 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, Pol. Sci., Schl of Human Env. Sci., Sociology
Mgmt. & Mkting	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

Note: STEM departments are in bold.

least 99 students had the course in their choice set and that less than 50% of those who had the course in their choice set took the course. Imposing this restriction results in 1,003 courses chosen by 16,079 unique students. This represents our baseline data for the choices of courses and grades.

There is an additional restriction imposed in the grade estimation. Namely, there are 18 courses where all students received the exact same grade, accounting for 518 individual–course observations or less than 1% of enrollments in the baseline data. The courses are still part of our course choice problem but are not used in the estimation of grades. Instead, the expected grades for students in these courses is set to what it is in the data, 4.0, with γ for these courses set to 0.

Estimates of the grade parameters show an additional 7 courses where the estimate of γ is less than 0.01 (estimates of γ are constrained to be greater than zero). For these courses, the factors yielding high grades are fundamentally different from those of other courses in the same department. These courses account for 224 individual–course observations or less than 0.5% of enrollments in the baseline data. For the purposes of estimating study times (where one of the inputs is $\ln(\gamma)$) and the professor estimation (where γ is a choice), we do not use these courses. For our counterfactuals, we fix the grading policies of these courses to what we observe in the data.

For the study effort analysis, observations are at the course–cohort level with an initial sample of 2,395 course–cohort evaluations. In principle, there could have been 4,012 observations if there were a student from each cohort in the class who also filled out a course evaluation. The 2,395 is then the result of some courses either not having students in a particular cohort or having students in a particular cohort where none filled out the course evaluation.

We implement a number of additional restrictions on the sample for the study effort analysis. The cohort of the student in the evaluation data is based on the students' self reports, while in the administrative data, it is based on our calculations given the academic records of the student. We define the response rate for the course–cohort as the number of course–cohort observations in the evaluation data divided by course–cohort enrollment in the registrar data. Because we want the average characteristics for a particular course–cohort from the registrar data to match the characteristics of those who filled out the evaluations, we restrict our analysis to course–cohorts where the response rate on the evaluations is between

70% and 101%. Imposing this restriction reduces our number of course-cohort observations to 866. Further removing courses with γ_js less that 0.01 results in a final sample of 850 course-cohort observations.

For the professor estimation, we do not use courses where γ is less than 0.01. We also do not use courses that hit their capacity constraint, as the professor maximization problem is different when the capacity constraint binds. This reduces our number of courses to 951.

B.3. *Construction of students' choice sets*

We account for administrative and academic rules and students' academic histories to construct accurate class choice sets for students:

1. Academic history: We drop classes that the student completed over the prior seven semesters (fall 2008–spring 2011) unless he or she is in the class in fall 2012.
2. Class prerequisites: We compile lists of prerequisite classes (from the UK Undergraduate Bulletin) for every course. We use the student's academic history and close the choice set unless all prerequisites are met. If a student is in a class without having completed all requirements, we assume an exemption was granted by the instructor.
3. AP exams: Students can bypass introductory courses in some subjects (from the UK Undergraduate Bulletin) with a score of 3 or above on the corresponding AP exam.
4. Room capacity: We have timestamps for all classes that students registered for. Using data on room capacity, we find the timestamp of when/if the class reaches capacity. We compare this time stamp to the first observed time stamp for the student. If the student's first timestamp is after the class's timestamp, the class is not in the choice set.

Table B.2 shows the average share of STEM and non-STEM classes available by cohort after the imposition of each restriction. Restriction 1 implies that for seniors, almost 10% (5%) of courses in STEM (non-STEM) are closed, which is reflective of more demand for STEM courses. Restriction 2 substantively restricts the choice set, especially for STEM. Overall, almost 40% (20%) of STEM (non-STEM) courses are closed due to students either not meeting prerequisites or having already completed the course. The changes to the choice set from AP exams (Restriction 3) or capacity constraints (Restriction 4) are marginal.

TABLE B.2

SHARE OF COURSES AVAILABLE BY COHORT/STEM CLASSIFICATION UNDER CHOICE SET RESTRICTIONS

Restriction	Freshmen	Sophomores	Juniors	Seniors	Overall
STEM departments					
(1)	1.000	0.953	0.921	0.906	0.946
(2)	0.550	0.541	0.552	0.555	0.550
(3)	0.562	0.547	0.556	0.558	0.556
(4)	0.556	0.546	0.555	0.557	0.554
non-STEM departments					
(1)	1.000	0.974	0.962	0.952	0.972
(2)	0.802	0.785	0.790	0.789	0.792
(3)	0.804	0.787	0.791	0.790	0.793
(4)	0.800	0.786	0.790	0.789	0.792

Note: (1) removes courses already taken. (2) removes courses where prerequisites are not met based on transcripts. (3) adds courses for which the prerequisites were met by AP exams. (4) removes courses where capacity constraints are met and adds courses where the student enrolled in the course despite not meeting the prerequisites.

B.4. *Characteristics of classes with above and below median grades*

Table B.3 compares classes with above median average grades to classes with below median average grades. Higher grades are associated with being a non-tenure track instructor as well as with being female. The latter could be due in part to STEM classes giving lower grades. Courses with high grades are also somewhat more likely to receive positive student evaluations. In our structural model, we handle the fact that higher grades may be correlated with other factors that drive student demand by directly including many of these variables in our estimating equations, modeling the professor behavior explicitly, and including course-level fixed effects to account for other student non-grade preferences.

B.5. *Additional parameters from the motivating regressions*

In Table B.4 we show estimates of the department indicator variables from Tables III and IV. The grade regression results show that the coefficients are lowest for STEM classes plus English and Psychology. For example, in the first column with the entire sample, there is a gap of over 0.8 grade points between the highest-grading department (Education & Health)

TABLE B.3

CHARACTERISTICS OF HIGH VS. LOW GRADE CLASSES

		High Grades	Low Grades
Faculty Rank	Full / Assoc. / Assist. / Lecturer	0.13 / 0.16 / 0.13 / 0.58	0.20 / 0.20 / 0.15 / 0.45
	Female Professor	0.48	0.34
Class Eval	Presents effectively	3.44 (0.40)	3.23 (0.46)
	Stimulates interest	3.37 (0.39)	3.11 (0.43)
	Stimulates further reading	3.25 (0.40)	2.97 (0.50)
	STEM Dept.	0.20	0.47

Note: Fall 2012 University of Kentucky courses with enrollments of 15 or more students; Classes divided at the median grade: 3.16. Standard deviations in parentheses. Class evaluation questions use a 5-point Likert scale.

and lowest-grading department (Chemistry & Physics). The second set of columns shows that the Engineering and Chemistry & Physics departments have the highest coefficients for hours of study.

B.6. *Additional structural parameters and standard errors*

In Table B.5 we show the full set of student preference parameters (see Table VI for a subset of the parameters). The parameters not discussed in the body of text also follow the expected patterns. The more courses opened up by a class (ln Open Class), the more appealing the class is for sophomores and even more so for freshmen. For junior and seniors, courses that fill requirements for their declared majors are associated with higher utilities, as are upper-level classes in general.

TABLE B.4

REDUCED-FORM REGRESSIONS OF DEPARTMENT GRADE AND STUDY HOURS PARAMETERS

	Grades				Study Hours			
Department	All Classes	Upper Level	STEM	non-STEM	All Classes	Electives	STEM	non-STEM
Regional Std.	0.014(0.031)	0.099(0.054)		0.024(0.029)	0.192(0.084)	0.650(0.134)		0.188(0.079)
Communication	0.244(0.025)	0.210(0.035)		0.223(0.023)	0.182(0.085)	0.422(0.151)		0.194(0.080)
Ed&Health	0.341(0.026)	0.421(0.034)		0.368(0.025)	-0.100(0.082)	0.382(0.177)		-0.093(0.077)
Engineering	-0.161(0.029)	-0.156(0.042)			0.767(0.092)	0.812(0.196)		
Language	0.104(0.027)	0.036(0.042)		0.085(0.025)	0.094(0.074)	0.512(0.115)		0.085(0.070)
English	-0.102(0.034)	-0.170(0.053)		-0.096(0.032)	0.269(0.090)	0.621(0.132)		0.269(0.085)
Biology	-0.408(0.028)	-0.213(0.043)	-0.370(0.031)		0.263(0.137)	0.372(0.201)	-0.437(0.177)	
Math	-0.353(0.026)	-0.350(0.058)	-0.277(0.025)		0.487(0.108)	0.544(0.280)	-0.305(0.125)	
Chem.&Phys.	-0.487(0.027)	-0.286(0.092)	-0.424(0.026)		0.364(0.110)	0.167(0.184)	-0.434(0.148)	
Psychology	-0.212(0.030)	-0.286(0.092)		-0.203(0.029)	0.310(0.154)	0.781(0.206)		0.315(0.146)
Social Science	-0.022(0.025)	0.057(0.036)		-0.054(0.024)	0.123(0.075)	0.608(0.121)		0.125(0.071)
Mgmt.&Mkting	0.249(0.031)	0.363(0.036)		0.259(0.028)	-0.144(0.107)	0.443(0.196)		-0.134(0.101)
Econ,Fin.,Acct	-0.143(0.027)	0.051(0.036)	0.008(0.025)		0.235(0.096)	0.597(0.184)	-0.514(0.127)	

Note: STEM departments are in bold. Agriculture is the excluded department for all cases except only; Engineering is the excluded category in the STEM-only specifications. Standard errors are in parentheses.

TABLE B.5

COMPLETE TABLE – ESTIMATES OF PREFERENCE PARAMETERS

	E(grades)		Fem x E(grade)		Fem x Fem Prof							
All Cohorts x	0.927	(0.006)	0.257	(0.009)	0.141	(0.007)			No. of Classes in Dept. Last Year			
	Req. Class		ln Open Classes		STEM Class		Fem x STEM Class		One		More than One	
Freshmen x	3.006	(0.021)	0.476	(0.004)	0.469	(0.018)	-0.046	(0.025)	—		—	
Sophomores x	1.510	(0.021)	0.388	(0.004)	0.422	(0.016)	-0.145	(0.023)	0.212	(0.014)	0.832	(0.014)
	Major Req. (A)		Major Req. (B)		Upper-Lev. Class							
Upperclassmen x	3.558	(0.008)	2.091	(0.017)	1.861	(0.007)						

	Female		ACT read		ACT Math		HS GPA		Type 3		Type 2	
Regional Studies	-0.091	(0.027)	-0.231	(0.021)	-0.062	(0.024)	-0.057	(0.022)	-1.580	(0.072)	-1.555	(0.032)
Communications	-0.401	(0.022)	-0.247	(0.016)	0.036	(0.018)	-0.090	(0.016)	-0.987	(0.043)	-1.484	(0.022)
Education & Health	0.139	(0.023)	-0.336	(0.018)	0.010	(0.020)	0.075	(0.017)	-1.227	(0.047)	-2.063	(0.026)
Engineering	-0.897	(0.032)	-0.275	(0.020)	0.494	(0.021)	0.048	(0.020)	-0.802	(0.060)	-1.237	(0.028)
Languages	-0.159	(0.024)	-0.116	(0.018)	0.042	(0.020)	-0.154	(0.018)	-1.383	(0.055)	-1.471	(0.026)
English	0.127	(0.029)	0.024	(0.023)	-0.208	(0.025)	0.051	(0.023)	-0.430	(0.073)	-1.580	(0.036)
Biology	0.504	(0.026)	-0.238	(0.019)	0.033	(0.021)	-0.136	(0.019)	-0.688	(0.056)	-1.176	(0.028)
Math	-0.133	(0.026)	-0.154	(0.017)	-0.193	(0.019)	-0.200	(0.017)	-1.265	(0.049)	-1.095	(0.024)
Chem. & Physics	0.069	(0.026)	-0.177	(0.017)	0.143	(0.019)	-0.178	(0.017)	-1.051	(0.050)	-1.131	(0.025)
Psychology	0.162	(0.025)	-0.375	(0.020)	0.047	(0.023)	-0.202	(0.019)	-0.816	(0.055)	-1.066	(0.029)
Social Sciences	-0.350	(0.022)	-0.175	(0.016)	-0.010	(0.018)	-0.182	(0.016)	-0.966	(0.044)	-1.247	(0.021)
Mgmt. & Mkting	-0.049	(0.028)	-0.243	(0.021)	0.138	(0.023)	-0.081	(0.021)	-1.392	(0.075)	-1.551	(0.033)
Econ., Fin., Acct.	-0.378	(0.026)	-0.295	(0.018)	0.131	(0.020)	-0.040	(0.018)	-1.760	(0.061)	-1.477	(0.026)

Note: "Major Req. (A)" refers to whether the course was required for the major; "Major Req. (B)" refers to whether the course was one of two or more required for the major.

B.7. *Share of courses taken in each department by unobserved type*

In Table B.6 we show how types are distributed across departments. The order of the rows is given by the ranking on the ratio of the type 1 (high-ability) share to the type 3 (low-ability) share, implying positive selection into courses listed in the first few rows.

TABLE B.6

SHARE OF COURSES TAKEN IN EACH DEPARTMENT BY UNOBSERVED TYPE

	Type 1 (High Ability)	Type 2 (Medium Ability)	Type 3 (Low Ability)
Econ., Fin., Acct.	0.0944	0.0614	0.0415
Management & Marketing	0.0477	0.0278	0.0230
Regional Studies	0.0383	0.0319	0.0279
Biology	0.0656	0.0540	0.0478
Engineering	0.0643	0.0543	0.0471
Chem. & Physics	0.0985	0.0916	0.0803
Languages	0.0691	0.0622	0.0669
Math	0.1165	0.1291	0.1137
English	0.0237	0.0222	0.0236
Psychology	0.0502	0.0496	0.0530
Social Sciences	0.1092	0.1277	0.1360
Communications	0.1211	0.1547	0.1612
Education & Health	0.0889	0.0573	0.1269
Agriculture	0.0126	0.0763	0.0511

Note: The shares of each type are 64.1%, 29.5%, and 6.4% respectively. STEM departments are in bold. The order is given by the ranking on the ratio of the type 1 (high-ability) share to the type 3 (low-ability) share.

B.8. *Additional regression parameters relating course demand to average grades and workloads*

In Table B.7 we show the full set of professor parameters (see Table X for a subset of the parameters).

APPENDIX C: METHODS APPENDIX – ONLINE

This appendix provides additional details regarding our empirical methods:

TABLE B.7

COMPLETE TABLE – RELATING COURSE DEMAND TO GRADES AND WORKLOADS

	Average Grades				γ			
	OLS		IV		OLS		IV	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Constant	3.418	(0.091)	4.375	(0.097)	0.124	(0.021)	-0.109	(0.022)
Ln Enroll	-0.066	(0.020)	-0.352	(0.023)	0.007	(0.004)	0.077	(0.005)
Grad. Student	-0.057	(0.044)	-0.045	(0.042)	0.006	(0.010)	0.003	(0.010)
Lecturer	-0.016	(0.043)	0.111	(0.041)	0.015	(0.010)	-0.016	(0.009)
Asst. Prof.	-0.136	(0.049)	-0.079	(0.047)	0.040	(0.011)	0.027	(0.011)
Tenured Prof.	-0.091	(0.040)	-0.030	(0.039)	0.018	(0.009)	0.003	(0.009)
Female Prof.	0.102	(0.030)	0.123	(0.029)	0.008	(0.007)	0.003	(0.007)
Female Prof. X STEM	-0.023	(0.062)	-0.065	(0.059)	-0.020	(0.014)	-0.010	(0.013)
Upper-Level Class	0.084	(0.034)	-0.016	(0.032)	0.009	(0.008)	0.033	(0.007)
Upper-Level Class X STEM	0.088	(0.065)	-0.013	(0.062)	0.035	(0.015)	0.060	(0.014)
Regional Studies	0.051	(0.075)	0.028	(0.071)	0.061	(0.017)	0.066	(0.016)
Communications	0.086	(0.062)	0.110	(0.059)	0.020	(0.014)	0.014	(0.013)
Education & Health	0.225	(0.068)	0.275	(0.064)	-0.009	(0.015)	-0.022	(0.015)
Engineering	-0.127	(0.079)	-0.011	(0.075)	0.267	(0.018)	0.238	(0.017)
Language	-0.024	(0.068)	-0.031	(0.065)	0.036	(0.015)	0.037	(0.015)
English	-0.123	(0.082)	-0.171	(0.078)	0.053	(0.019)	0.065	(0.018)
Biology	-0.276	(0.105)	0.109	(0.102)	0.146	(0.024)	0.052	(0.023)
Math	-0.396	(0.073)	-0.209	(0.070)	0.120	(0.016)	0.074	(0.016)
Chem. & Physics	-0.287	(0.085)	-0.037	(0.082)	0.071	(0.019)	0.011	(0.019)
Psychology	-0.145	(0.098)	0.069	(0.094)	0.095	(0.022)	0.043	(0.021)
Social Science	-0.180	(0.064)	-0.111	(0.061)	0.020	(0.015)	0.003	(0.014)
Mgmt. & Mkting	0.213	(0.087)	0.339	(0.083)	-0.022	(0.020)	-0.052	(0.019)
Econ., Fin., Acct.	-0.278	(0.091)	-0.018	(0.088)	0.070	(0.021)	0.007	(0.020)

Note: The analysis is at the course level. The estimates are from 951 courses where $\gamma_j > 0$ and the course capacity constraint does not bind.

1. our modified EM algorithm for recovering the parameters of the grade process and conditional probabilities of a student being each unobserved type,
2. the fixed point algorithm we use when estimating the structural utility parameters of the students,

3. and our method of solving for counterfactual choice probabilities in the presence of capacity constraints.

C.1. *Modified EM algorithm*

We first describe our estimation procedure in the presence of unobserved heterogeneity. 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 that the error is distributed $N(0, \sigma_\eta)$. In theory, one could use the structural choice likelihood in Equation (23) to capture the likelihood of making observed course choices; however, maximizing Equation (23) at every iteration of the EM algorithm is computationally infeasible. Instead, we construct an alternative course choice likelihood function based on a flexible analog of the structural model. For the reduced-form choice problem, we abstract from the bundling of courses, treating each course choice as its own decision problem. To facilitate computation, at points, we break down the problem into the probability of taking a course from department k and then the probability choosing the specific course j:

$$p_{ijk} = p_{ik} p_{ij|k}$$

We specify the reduced-form payoff of taking class j as:

$$v_{ij} = (\phi_1^* + w_i \phi_2^*) g_{ij}(\gamma_j^N, \theta_{j(k)}^N, X_i) + \delta_{0j}^* + w_i \delta_{1j}^* + Z_{1i} \delta_{2k(j)}^* + Z_{2ij} \delta_3^* + \epsilon_{ij}^* \tag{35}$$

where $g_{ij}(\cdot)$ represents the expected grade of student i in course j and ϵ_{ij}^* is assumed to follow a nested logit structure with nesting at the department level characterized by ν. The full set of choice parameters is then $\varphi = \{\phi^*, \delta^*, \nu\}$. Note that although we will not be interpreting the estimates of φ, the structure of utility in Equation (35) is very similar to the structure in Equation (9).[30] This ensures that the conditional type probabilities from this specification are appropriate for classifying students for the estimation of Equation (9).

[30]The structure of utility in Equation (35) differs from the structure in Equation (9) in three ways: First, Equation (35) does not subtract γ_j from expected grades. Second, Equation (35) assumes nested logit preference shocks, while Equation (9) assumes independent Type 1 extreme value errors. Finally, Equation (35) assumes that contemporaneous choices are independent, while Equation (9) models students choosing bundles of courses simultaneously.

Let φ represent the parameters of this flexible choice process. The integrated log likelihood is then:

$$\sum_i \ln\left(\sum_{s=1}^{S} \pi_s \mathcal{L}_{igs}(\theta,\gamma)\,\mathcal{L}_{ics}(\varphi)\right) \tag{36}$$

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

We iterate on the following steps until convergence, where the mth step follows:

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

$$q_{is}^{(m)} = \frac{\pi_s^{(m)}\mathcal{L}_{igs}\left(\theta^{(m-1)},\gamma^{(m-1)}\right)\mathcal{L}_{ics}\left(\varphi^{(m-1)}\right)}{\sum_{s'}\pi_{s'}^{(m)}\mathcal{L}_{igs'}\left(\theta^{(m-1)},\gamma^{(m-1)}\right)\mathcal{L}_{ics'}\left(\varphi^{(m-1)}\right)} \tag{37}$$

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 $\{\theta^{(m)},\gamma^{(m)},\varphi^{(m)}\}$ by maximizing:

$$\sum_i\sum_s q_{is}^{(m)}\left(\ln\left[\mathcal{L}_{igs}(\theta,\gamma)\right] + \ln\left[\mathcal{L}_{ics}(\varphi)\right]\right) \tag{38}$$

To facilitate computation, the maximization step (step 3) is conducted in stages. Denote as $f(g_{ij},\gamma_j^N,\theta_{k(j)}^N)$ the likelihood of observing g_{ij} given the parameters γ_j^N and $\theta_{k(j)}^N$. Denote as $\varphi(!A)$ φ absent the Ath component. Finally, denote as d_{ij} an indicator for whether i chose course j, as d_{ijk} an indicator for whether i chose course j in department k, and as d_{ik} an indicator for whether i's choice was in department k. Maximization then proceeds as follows:

1. For each department $k\in K$, taking φ as given, choose γ_j^N and θ_k^N to maximize:

$$\sum_i\sum_{j\in k} d_{ijk}\left(\ln[f(g_{ij},\gamma_j^N,\theta_k^N)] + \ln[p_{ij|k}(\gamma_j^N,\theta_k^N,\varphi)]\right) \tag{39}$$

2. Taking γ_j^N, θ_k^N, and $\varphi(!\phi^*, !\delta_3^*)$ as given, choose ϕ^* and δ_3^* to maximize:

$$\sum_i \sum_j d_{ij} \ln \left[p_{ijk} \left(\gamma_j^N, \theta_k^N, \varphi(!\phi^*, !\delta_3^*), \phi^*, \delta_3^* \right) \right] \tag{40}$$

3. For each department $k \in K$, taking γ_j^N, θ_k^N, and $\varphi(!\delta_0^*)$ as given, choose δ_{0j}^* (relative to one course in each department) to maximize:

$$\sum_i \sum_{j \in k} d_{ijk} \ln \left[p_{ij|k} \left(\gamma_j^N, \theta_k^N, \varphi(!\delta_{0j}^*), \delta_{0j}^* \right) \right] \tag{41}$$

4. Taking γ_j^N, θ_k^N, and $\varphi(!\delta_1^*, !\delta_2^*, !\nu)$ as given, choose $\delta_{1j(k)}^*$, $\delta_{2j(k)}^*$, and ν to maximize:[31]

$$\sum_i \sum_k d_{ik} \ln \left[p_{ik} \left(\gamma_j^N, \theta_k^N, \varphi(!\delta_1^*, !\delta_2^*, !\nu), \delta_1^*, \delta_2^*, \nu \right) \right] \tag{42}$$

The advantage of this sequential strategy is that it limits the number of parameters being estimated at each stage and limits the number of times that the 1,003 choice probabilities are calculated for each individual. Further, when the 1,003 choice probabilities are calculated within the maximization routine at step 2, the number of parameters over which we are maximizing is limited.

Once the algorithm has converged, we have consistent estimates of $\{\theta, \gamma, \varphi\}$ and the conditional probabilities of a student being of 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 of the study process in (20). Finally, we use the estimates of q_{is} as weights in estimating the structural choice parameters using (23).

C.2. *Fixed-point algorithm*

We now describe our fixed-point algorithm used in each calculation of the student choice likelihood. Let $\tilde{\Theta} = \{\delta_{1k(j)}, \delta_{2k(j)}, \delta_3, \phi_0, \phi_1\}$ represent choice parameters other than δ_{0j}, let S_j^d represent the share of students choosing course j in the data, and let $S_j\left(\delta_{0j}, \tilde{\Theta}\right)$ represent the predicted share of students choosing course j as a function of δ_{0j} and other choice parameters. Given a new guess of $\tilde{\Theta}$, we use the δ_{0j}s from the previous guess $\delta_{0j}^0(\tilde{\Theta})$

[31] At this step, we also recover the δ_{0j}^*s for the normalized courses in each department from step 3.

and calculate $S_j\left(\delta_{0j}^0, \tilde{\Theta}\right)$. The mth iteration of the fixed-point problem updates δ_{0j}^m using:

$$\delta_{0j}^m = \delta_{0j}^{m-1} + \ln\left[S_j^d\right] - \ln\left[S_j\left(\delta^{m-1}, \tilde{\Theta}\right)\right] \tag{43}$$

Given the δ_{0j}^m, we update $S_j\left(\delta_{0j}^m, \tilde{\Theta}\right)$. These steps are repeated until the predicted and actual enrollment shares are arbitrarily close.

C.3. *Counterfactuals in the presence of capacity constraints*

Embedded within each counterfactual are each student's conditional choice probabilities. To ensure that capacity constraints are not exceeded, we work backward based on the registration ordering given by the timestamps (see Section B.3). We proceed in the following manner for each student n, where n refers to the ordering based on the student's timestamp:

1. Calculate the choice probabilities for student n over the courses where the student has met the prerequisites and where the course is not already filled.
2. If adding the choice probabilities to the probabilities of the previous $n-1$ students does not cause any of the classes to exceed the course capacity, proceed to the next student.
3. If one or more of the courses exceeds capacity in step 2, identify the class where adding n's probability causes the capacity constraint to be exceeded by the greatest amount. Label this excess capacity c and the choice probability p. Note that $c < p$, as the course previously had open space.
4. Assign the probability that the course identified in step 3 is in n's choice set as $1 - \frac{c}{p}$. Take the choice probabilities when this course is in the choice set, multiply them by $1 - \frac{c}{p}$, and add them to the number of enrollees in each course. This ensures that the identified course will be exactly filled. Repeat step 1 for student n, taking into account that all probabilities from the new choice set will be multiplied by $\frac{c}{p}$ and where the identified course is no longer in n's choice set.

The algorithm ensures that any capacity-constrained courses are exactly filled, with filled courses no longer available to individuals with later timestamps.

APPENDIX D: EXTENDING THE SUPPLY-SIDE MODEL TO INCLUDE PROFESSOR EFFORT – ONLINE

In this section, we extend our model of professor choices to include efforts exerted to affect enrollment directly. In the section, we

1. show how effort affects the course payoffs,
2. show how professor effort is measured,
3. show how the extension affects the modeling and estimation of the professor's objective function,
4. present estimates of the parameters of the professor's objective function and the equilibrium counterfactuals with the extended model, and

We extend our model to allow professors to directly influence demand for courses by exerting effort, in addition to setting grading parameters. We decompose the course fixed effect in the student's utility function, δ_{0j}, into intrinsic demand, δ_{0j}^*, and the effort of the professor, τ_j:

$$\delta_{0j} = \rho\tau_j + \delta_{0j}^* \tag{44}$$

where ρ measures how professor effort translates into course utility. Two major complications arise in extending the model. First, clean measures of professor effort are difficult to obtain from administrative data. We use student responses from the evaluation data and purge potentially contaminating endogenous effects to arrive at a viable measure. Second, we do not have a way to recover ρ. As a result, we estimate the model under different assumed values of ρ.

D.1. *Measuring Professor Effort*

Outside of a time-use survey or rigidly prescribed schedules (for example, unionized manufacturing jobs), it is often difficult to gather data on worker effort. For professors, whose time could have multiple uses (for example, data analysis or writing an article/book could yield benefits for both research and teaching), even direct measures of inputs become problematic. Instead, we use information about students' receptivity to the professor's teaching to capture a measure of the professor's effort, τ. Of the twenty questions in the evaluations, we focus on three with students answering on a five-point Likert scale:

TABLE D.1

CORRELATION AMONG CLASS EVALUATION AND GRADES THAT STUDENTS EXPECT TO RECEIVE

	Expected Grade	Q09	Q13	Q19	(Q09+Q13+Q19)/3
Expected Grade	1.0000				
Q09	0.2132	1.0000			
Q13	0.2392	0.7364	1.0000		
Q19	0.2192	0.5878	0.7387	1.0000	
(Q09+Q13+Q19)/3	0.2518	0.8575	0.9291	0.8818	1.0000

Note: Expected Grades are grades that students expect to receive (as indicated on class evaluations). Questions receive responses on the evaluation on a 5-point Likert scale and are worded as follows: Did the instructor (1) present the material effectively – Q09, (2) stimulate interest in the subject – Q13, and (3) stimulate you to read further beyond the class – Q19?

- Q09: Did the professor present class materials effectively?
- Q13: Did the professor stimulate your interest in the subject?
- Q19: Did the professor stimulate you to read further in the subject beyond the class?

We average these three measures to create a student i's perception of professor effort in course j, $\tau_{ij}^{(1)}$.

There are at least two issues with using this average as a measure of effort. First, as shown in Table D.1, professors who give high grades may receive better evaluations because of the high grades rather than because of the effort exerted by the professor.[32]

We are able to purge the effort measure of grade effects because the evaluation data contain the expected grade of each student filling out the evaluation. Using evaluation data across multiple semesters (fall 2011 to spring 2013), we regress $\tau_{ij}^{(1)}$ on a course fixed effect and dummy variables for each expected grade. The course fixed effect, $\tau_j^{(2)}$, gives us a measure of effort purged of the effect of offering high grades. The results from this regression are given in the top half of Appendix Table D.2 and show that higher expected grades are associated with better evaluations.

The second issue is that, conditional on the same amount of effort, some instructors may be better in the classroom than others. Since we are interested in discretionary effort rather than fixed instructor ability, we purge our effort measure of instructor effects, using multi-

[32]See Insler et al. (2021), Nelson and Lynch (1984), and Zangenehzadeh (1988), who also find this positive relationship.

TABLE D.2

PROFESSOR EFFORT RESIDUALIZATION & REGRESSION OF EFFORT MEASURE ON LOG ENROLLMENT

	Coef.	Std. Err.
Expected Grade:		
A	0.9519	(0.0358)
B	0.7709	(0.0358)
C	0.5336	(0.0361)
D	0.2933	(0.0370)
log(class size)	-0.0901	(0.0082)

Note: The dependent variable in the top half is the average response to questions 9, 11, and 13 from the evaluation data. Regressors include class times semester fixed effects. The dependent variable in the bottom half is the average response to the three evaluation question minus the grade effects estimated in the top half of the table. Regressors include professor and semester fixed effects. Sample size for the top panel is 150,303 and 4,075 for the bottom panel. Both use evaluation data from Fall 2011 to Spring 2013.

ple semesters of the evaluation data. To do so, we collapse the multi-semester data to the class-year-semester level and regress $\tau_j^{(2)}$ on an instructor fixed effect (taking advantage of the panel nature of the data) and log enrollment. The regression results in the bottom half of Appendix Table D.2 show that the coefficient on log enrollment is large and negative, implying that perceived quality of the class is lower when enrollment is high given the same instructor. We then subtract the instructor fixed effect but leave in the effect of log enrollment: effort should be correlated with log enrollment if it is responding to characteristics of the class. We then standardize this variable to have mean zero and standard deviation one. It is this standardized variable that we use for τ_j.

For estimation of the professor model with professor effort, we impose additional restrictions on the sample. Here, we need professors to have at least two measures of effort across fall 2011 to spring 2013, in addition to having one of those measures for our semester of analysis, fall 2012. This reduces our sample to 748 courses.

D.2. *Model Extension and Estimation*

Professors choose their effort level, τ_j, in addition to grading policy parameters β_j and γ_j. The professor has an ideal effort level e_{3j}, which depends on his or her observed and unobserved characteristics. Then, our equilibrium objects, expected grades, probability of student i enrolling in a class, and log enrollments, are now defined as $\overline{G}_j(\beta,\gamma,\tau)$, $P_{ij}(\beta,\gamma,\tau)$, and $\ln[E_j(\beta,\gamma,\tau)]$, respectively. The professor's objective function now has an extra term to maximize:

$$V_j(\beta,\gamma,\tau) = -\left(\ln[E_j(\beta,\gamma,\tau)] - e_{0j}\right)^2 - \lambda_1\left(\overline{G}_j(\beta,\gamma,\tau) - e_{1j}\right)^2 \\ -\lambda_2(\gamma_j - e_{2j})^2 - \lambda_3(\tau_j - e_{3j})^2 \tag{45}$$

where $e_{3j} = W_{3j}\Psi_3 + \varepsilon_{3j}$. Solving for ideal effort proceeds similarly to the procedure in the main model. There is an extra first-order condition:

$$0 = -\left(\ln[E_j(\beta,\gamma,\tau)] - e_{0j}\right)\frac{\partial \ln E_j}{\partial \tau_j} - \lambda_1\left(\overline{G}_j(\beta,\gamma,\tau) - e_{1j}\right)\frac{\partial \overline{G}_j}{\partial \tau_j} - \lambda_3(\tau_j - e_{3j}) \tag{46}$$

In recovering Ψ_0, Ψ_2, and λ_2, we create our instruments with β^0, γ^0, and τ^0.

To estimate λ_3 and Ψ_3, we take Ψ_0 as given and eliminate λ_1 using using Equations (29) and (46) to solve for τ_j:

$$\tau_j = (1/\lambda_3)C_j\left(\ln[E_j] - \Psi_0\right) + \Psi_3 + \varepsilon_{3j} \tag{47}$$

where C_j is given by:

$$C_j = \left[\frac{\partial \ln[E_j]}{\partial \beta_j}\frac{\partial \overline{G}_j}{\partial \tau_j}\Bigg/\frac{\partial \overline{G}_j}{\partial \beta_j}\right] - \frac{\partial \ln[E_j]}{\partial \tau_j} \tag{48}$$

We then instrument for $C_j\left(\ln[E_j] - \Psi_0\right)$ by evaluating C_j and $\ln[E_j]$ at the common grading and effort policies, β^0, γ^0, and τ^0. Recovery of Ψ_1 and λ_1 proceeds as before.

D.3. *Professor preference estimates and equilibrium counterfactuals under different values of ρ*

We estimate the professor preference parameters in this extended model $\rho = 0.05$ and $\rho = 0.2$. Table D.3 shows the estimates of professor preferences (shown in Table XI) at alternate values of ρ. Table D.4 shows the general equilibrium counterfactual results (shown in Table XIV) at alternative values of ρ. Allowing for professor endogenous effort mutes the effects of the grading policy but does so only slightly. For example, at $\rho = 0.2$ average class for STEM courses is 99.5, or about one student less than when professors could not adjust their effort.

TABLE D.3

ESTIMATES OF PROFESSOR PREFERENCES AT ALTERNATIVE ρ VALUES

	$\rho = 0.2$		$\rho = 0.05$		$\rho = 0.2$		$\rho = 0.05$	
	Ideal grade				Ideal workload			
λ	2.933	(0.448)	2.898	(0.434)	48.167	(5.025)	46.874	(4.675)
Constant	2.627	(0.120)	2.619	(0.119)	0.275	(0.038)	0.277	(0.038)
Upper-Level Class	0.436	(0.065)	0.440	(0.065)	-0.074	(0.038)	-0.075	(0.038)
Upper-Level X STEM	-0.050	(0.069)	-0.052	(0.069)	0.073	(0.021)	0.074	(0.021)
Grad. Student	-0.009	(0.050)	-0.008	(0.051)	0.008	(0.014)	0.008	(0.014)
Lecturer	0.112	(0.052)	0.114	(0.053)	-0.012	(0.013)	-0.013	(0.013)
Asst. Prof.	-0.106	(0.054)	-0.105	(0.054)	0.035	(0.015)	0.035	(0.015)
Tenured Prof.	-0.070	(0.047)	-0.069	(0.047)	0.007	(0.012)	0.007	(0.012)
Female Prof.	0.102	(0.032)	0.102	(0.033)	0.002	(0.009)	0.002	(0.009)
Female Prof. X STEM	-0.056	(0.065)	-0.056	(0.065)	-0.002	(0.018)	-0.002	(0.018)
Regional Studies	-0.015	(0.074)	-0.015	(0.074)	0.050	(0.021)	0.050	(0.021)
Communications	0.086	(0.069)	0.087	(0.069)	0.004	(0.019)	0.004	(0.019)
Education & Health	0.218	(0.068)	0.218	(0.069)	-0.011	(0.019)	-0.011	(0.019)
Engineering	-0.043	(0.082)	-0.041	(0.082)	0.206	(0.024)	0.205	(0.024)
Language	-0.057	(0.066)	-0.057	(0.066)	0.042	(0.018)	0.042	(0.018)
English	-0.223	(0.082)	-0.224	(0.082)	0.094	(0.024)	0.094	(0.024)
Biology	0.016	(0.129)	0.022	(0.129)	0.035	(0.031)	0.033	(0.030)
Math	-0.331	(0.081)	-0.328	(0.081)	0.094	(0.021)	0.092	(0.021)
Chem. & Physics	-0.159	(0.106)	-0.156	(0.106)	0.024	(0.026)	0.022	(0.026)
Psychology	-0.023	(0.101)	-0.020	(0.101)	0.055	(0.026)	0.053	(0.026)
Social Science	-0.138	(0.064)	-0.137	(0.064)	0.023	(0.018)	0.023	(0.018)
Mgmt. & Mkting	0.308	(0.088)	0.310	(0.088)	-0.042	(0.024)	-0.043	(0.024)
Econ., Fin., Acct.	-0.075	(0.101)	-0.071	(0.101)	0.008	(0.026)	0.006	(0.026)
	Ideal log enrl				Ideal prof. effort			
λ	1.000		1.000		0.528	(0.073)	0.252	(0.064)
Constant	5.196	(0.662)	5.173	(0.639)	-0.347	(0.061)	-0.179	(0.059)
Upper-Level Class	-1.385	(0.528)	-1.372	(0.511)				

Note: Ideal enrollment λ is normalized to equal 1. The base for rank is "Instructor," who are adjunct instructors contracted by the course/semester. "Lecturers" are offered longer-term contracts and are salaried.

TABLE D.4

COUNTERFACTUAL SCENARIOS IN GENERAL EQUILIBRIUM AT ALTERNATIVE ρ VALUES

		Class Size		STEM Enrollment Share		
	ρ	STEM	Non-STEM	Overall	Female	Male
Baseline		82.6	45.0	41.8%	34.6%	49.5%
Grade Around 3$^{\diamond}$	0	100.6	37.9	50.9%	45.1%	57.2%
	0.05	100.5	37.9	50.9%	44.9%	57.3%
	0.2	99.5	38.3	50.4%	44.4%	56.8%

Note: $\diamond$: "Grade Around 3" adjusts the mean grade in all courses to a B, affecting both men and women. Professors change their grading strategies based on student responses to changes in preferences and abilities for the general equilibrium analysis.