# 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

February 24, 2023

**Abstract**

Substantial earnings differences exist across occupations, with STEM occupations paying especially well. We show that stricter grading policies in STEM courses reduce STEM enrollment, especially for women. To show this, we estimate a model of student demand for courses and optimal effort choices of students given professor grading policies. Professor grading policies are treated as equilibrium objects that in part depend on student demand for courses. Restrictions on grading policies that equalize average grades across classes 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 that students take. Human capital in science, technology, engineering, and mathematics (STEM) is 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 thus centers on promoting STEM skills to increase their supply.[2] Of particular concern is the lack of

*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 Bureau of Labor Statistics or the Federal Trade Commission. We thank Erin Denison and Hema Shah for helpful assistance.

[1]There is also evidence that the heterogeneity across fields of study is increasing over time. See Gemici & Wiswall (2014).

[2]Various state legislatures have introduced bills to waive/reduce tuition or forgive student loans for STEM students who remain in-state to teach (Hinz (2019), Latek (2019), Chapman (2014)). At the national level, in 2019, a bill to expand STEM education and especially to increase girls' participation in computer science was signed into law

female representation in STEM occupations; women represent 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 classes in STEM than women. STEM classes at UK are characterized by enrollments that are almost twice as high, grades that are 0.33 points lower, and study times that are 44% higher than their non-STEM counterparts.[3] Furthermore, given existing evidence that women value grades more than men, lower STEM grades may disproportionately deter women from choosing STEM classes (Rask & Bailey (2002), Rask & Tiefenthaler (2008)).

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 students by changing their course bundles and study times.

We estimate our model of course choices using UK transcript data. Our model of course choices allows us to be quite flexible over student preferences for courses in particular departments while still being able to separately identify 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 assumption 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

(Building Blocks of STEM Act (S. 737)) (Stevens 2021), and the GI Bill was expanded to support veterans seeking STEM degrees (Veteran STEM Scholarship Improvement Act) (Gross 2019). Moreover, in 2021, legislation to speed green card issuance for international STEM students was introduced (Keep STEM Talent Act, STEM Jobs Act (H.R. 6429)) (Graeml 2019).

[3] Statistics from other institutions show similar patterns. See Freeman (1999), Bagues et al. (2008), and Rask (2010) on grading differences and Brint et al. (2012) and Stinebrickner & 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.

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 grades but lower ACT math scores than their male counterparts. Higher grades and lower ACT math scores are associated with relatively higher grades in non-STEM courses and 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 larger for women, shrinking the STEM gender gap. Further restricting the slopes of the grading policies to be the same across all courses would decrease the gender gap even more due to women having a comparative advantage in non-STEM courses.[5]

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, which may mediate 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 preferences over grades, enrollments, and

[5] Here the slopes are set to the median across all courses.

workloads that vary across departments. Classes that have 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 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 consider professor responses, requiring the same average grade across courses would increase female STEM enrollment by 30.3% and male STEM enrollment by 15.6%.[6] Since our simulations hold major choices fixed for juniors and seniors and major choice has substantial effects on course choices for these students, our estimates are best interpreted as short-run predictions. In the long run, when major choices can vary, the effects may be substantially larger. Thus, a standardized curve grading policy could be a low cost and effective way to increase overall STEM participation and decrease the gender gap in STEM.

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. Models of the effects of grades on wages typically focus on overall grade point average, as this is what traditionally appears on résumés (Piopiunk et al. (2020)). Indeed, that grade distributions are not fully known serves as an impetus for theoretical models of grade inflation resulting from competition between universities (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 supported by findings in the sociology and psychology literature (Rosenberg et al. (1995), Crocker et al. (2003), and Seymour & Hewitt (1997)).

Our paper relates to a long line of literature examining how college students make higher education decisions. One strand of this literature focuses on the effects of grades on these choices. This literature generally finds that low earned grades reduce persistence in STEM (Astorne-Figari & Speer (2018, 2019), McEwan et al. (2021)) and that grades affect educational choices for female students more than for male students (Ost (2010), Owen (2010), Rask & Bailey (2002), Rask &

[6]The corresponding numbers when ignoring professor responses are 34.7% and 17.8%.

Tiefenthaler (2008), Zafar (2013)).[7] 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, because Wellesley is a women's college, the authors cannot estimate differential effects by gender. In this context, one of our contributions is to show 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.

Methodologically, we also contribute to the literature on educational choices by estimating a structural model of course choices. Many of the existing structural models of within-college decisions focus on major choices, with some important limitations.[8] First, majors represent bundles of courses, grading policies, and instructors, making it difficult to separately identify preferences for individual bundle elements. By analyzing course choices, we observe variation in grading policies and instructor characteristics between very similar courses, enhancing our ability to separately identify preferences for various attributes. Second, there is substantial variation in human capital accumulation conditional on major.[9] By focusing on course choices, we obtain a more complete measure of the skills that students acquire.

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 (2021) 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. This builds upon descriptive evidence on the heterogeneity of grading policies over time and across departments (Babcock (2010), Johnson (2003), Sabot & Wakeman-Linn (1991)), policy experiments to reduce grading differences (Butcher et al. (2014), Bar et al. (2009)), and theoretical work on how to set 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

[7]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.

[8]A notable exception is Thomas (2021), who also models course choices; however, this paper assumes that courses within broad academic fields are identical.

[9]In our data, the interquartile range for the share of courses completed in STEM for seniors is 50% to 74% for STEM majors and 13% to 33% for non-STEM majors.

times than non-STEM courses, suggesting that student demand may be a factor in instructors' grading policies. These grading policies in turn have implications for cross-gender differences in course choices. We show that women receive higher grades than men in both STEM and non-STEM courses and study more than their male counterparts. However, women enroll in substantially fewer STEM courses and sort into classes where the average grades are higher. Taken together, these facts motivate an equilibrium model where students choose courses and study times in response to the grading policies of professors, with differential responses to these policies by male and female students.

Our analysis focuses 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 tenure status, enrollment caps, prerequisites and whether the course is a prerequisite for other courses, and whether the course fulfills any university or major requirements. The third is end-of-semester class evaluation surveys completed by students on their course quality perceptions, their expected grades, and the number of hours that they 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.[10]

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 16,109 students and 1,003 courses. 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 1. Women at UK arrive with higher high school grades but lower ACT math scores. For the semester that we analyze, women have significantly higher grades. They also take fewer courses in STEM: While men take almost half their courses in STEM departments, for women, the share is a little over a third. Key to our analysis is understanding this gap.

Part of the gender gap in STEM course enrollment may be due to different expectations and

[10] 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 two data sets. The evaluation data are also incomplete because some courses opt out of participating in the evaluation.

Table 1: Descriptive Statistics by Gender

| | Men | Women | p-value ($H_0$: Men = Women) |
|---|---|---|---|
| *Precollege outcomes* | | | |
| High school GPA | 3.49 (0.472) | 3.62 (0.401) | 0.00 |
| ACT Reading Score | 26.1 (5.13) | 26.0 (4.84) | 0.37 |
| ACT Math Score | 25.7 (4.65) | 23.9 (4.23) | 0.00 |
| *Fall 2012 outcomes* | | | |
| GPA | 2.86 (0.938) | 3.12 (0.848) | 0.00 |
| Share Courses in STEM | 49.5% | 34.6% | 0.00 |

Note: Fall 2012 University of Kentucky undergraduate students; 7,904 men and 8,286 women. SAT scores are converted to equivalent ACT scores. Share Courses in STEM is the share of STEM courses in total courses taken. Standard deviations in parentheses.

Table 2: Descriptive Statistics by Course Type

| | STEM | Non-STEM |
|---|---|---|
| Class Size | 80.2 (99.3) | 41.4 (46.0) |
| Average Grade | 2.94 (0.45) | 3.27 (0.42) |
| Average Grade \| Female | 3.00 (0.56) | 3.37 (0.43) |
| Study Hours | 3.46 (1.52) | 2.40 (0.81) |
| Percent Female | 37.0% | 58.3% |
| Percent Fem. Prof. | 27.0% | 46.4% |
| Percent Upper Level | 40.6% | 44.3% |

Note: Fall 2012 University of Kentucky courses with enrollments of 15 or more students; 341 STEM courses and 743 non-STEM courses (for study hours, 136 STEM courses and 398 non-STEM courses). Standard deviations in parentheses. Tests of equality in means across STEM/non-STEM yield p-values below 0.01 for all characteristics.

environments in STEM courses. Table 2 summarizes course characteristics by STEM status. STEM courses are almost double the size of non-STEM courses and have grades that are more than 0.3 points lower. Students spend an extra hour per week (40% more time) studying in STEM courses. While women earn higher grades in STEM courses than men, the gap is smaller than that for non-STEM courses.

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 2. Table 3 presents the results for grades across all classes and students in the first column; the other columns present the results for different subsets of courses (elective, upper level, and STEM/non-STEM) and students (juniors and seniors). Odd columns control for whether the class is in STEM; even columns control for our fourteen category effects.

Across all the columns of Table 3, the following patterns emerge. First, 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 reflects supply considerations, suggesting that classes with low demand inflate grading to attract students or that classes with high demand give lower grades to deter students. Second, women outperform men across the board, and this is especially true in non-STEM classes. Third, classes with a higher fraction of female students are associated with higher grades. This is consistent with the lack of a grade curve; otherwise, the higher grades that females receive would translate into lower grades for others. It also suggests that women may sort into classes that give higher grades. Finally, even when we control for academic background and course enrollment, we observe that STEM courses give lower grades.

There are, however, key differences across the subgroups. First, the negative association between class size and grades is especially strong in upper-division courses. Given that enrollments are lower in these courses, this may reflect classes with low demand inflating their grades to attract more students. Second, the association between college grades and academic background (as measured by test scores and high school grades) is heterogeneous across courses. The relationship between academic background and grades is weaker in upper-division courses, non-STEM courses, and courses taken by upperclassmen, all of which are associated with courses that would typically be smaller in size. This again suggests that grading policies may be sensitive to course demand. Finally, what academic factors matter is different for STEM and non-STEM courses, with ACT math scores showing a much stronger relationship with STEM grades than with non-STEM grades.

The descriptive regressions for study hours are more limited since we can use variation only at the cohort–class level.[11] We show in Table 4 the results of regressions of class–cohort study hours on the average characteristics of the class. The share of the class–cohort that is female is positively correlated with higher study times, and the coefficient is largest when course fixed effects are included. This holds true both overall and when we focus only on elective courses. That women study more than men is consistent with survey evidence from Arcidiacono et al. (2012) and Stinebrickner & Stinebrickner (2012).[12] STEM classes show higher study times, though not in elective courses where enrollments are lower.

Perhaps most interesting is the coefficient on average grade in the course.[13] Courses with higher

[11]Nonetheless, there is substantial variation in gender both across and within classes at the cohort level. The standard deviation of share female is 0.285. When we remove the course fixed effect and look at classes with at least two cohorts, the standard deviation is 0.177. The variation across cohorts within a class is driven by the small numbers of students in particular cohorts for particular classes and by differences in when prerequisites for classes are met.

[12]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.

[13]Note that this is calculated at the class level, not the cohort–class level.

Table 3: Regressions of Grades on Student and/or Class Characteristics

| | All Classes | | Upper Level | | Elective | Juniors & | STEM | Non-STEM |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | Classes (5) | Seniors (6) | Classes (7) | Classes (8) |
| Female | 0.102 | 0.100 | 0.103 | 0.099 | 0.118 | 0.067 | 0.043 | 0.152 |
| | (0.009) | (0.009) | (0.016) | (0.016) | (0.012) | (0.013) | (0.014) | (0.012) |
| Percent Female | 0.422 | 0.461 | 0.452 | 0.379 | 0.535 | 0.423 | 1.013 | 0.178 |
| | (0.024) | (0.028) | (0.039) | (0.047) | (0.040) | (0.038) | (0.048) | (0.034) |
| ACT Reading* | 0.038 | 0.049 | 0.046 | 0.068 | 0.058 | 0.040 | 0.061 | 0.035 |
| | (0.005) | (0.005) | (0.008) | (0.008) | (0.007) | (0.007) | (0.007) | (0.006) |
| ACT Math* | 0.134 | 0.131 | 0.089 | 0.080 | 0.144 | 0.123 | 0.201 | 0.084 |
| | (0.006) | (0.006) | (0.010) | (0.010) | (0.008) | (0.008) | (0.009) | (0.007) |
| High School GPA* | 0.265 | 0.266 | 0.182 | 0.185 | 0.314 | 0.214 | 0.280 | 0.254 |
| | (0.005) | (0.005) | (0.009) | (0.009) | (0.007) | (0.007) | (0.008) | (0.006) |
| ln(Class Size) | -0.031 | -0.037 | -0.048 | -0.106 | -0.029 | -0.059 | -0.047 | -0.027 |
| | (0.005) | (0.005) | (0.008) | (0.009) | (0.007) | (0.007) | (0.008) | (0.006) |
| STEM Class | -0.444 | | -0.308 | | | | | |
| | (0.010) | | (0.018) | | | | | |
| Dept. FE | No | Yes | No | Yes | Yes | Yes | Yes | Yes |
| Observations | 55,701 | 55,701 | 15,458 | 15,458 | 29,638 | 26,380 | 26,076 | 29,625 |

Note: $*$ indicates that the variable is z-scored. Additional controls include indicators for upper-level classes (those with course numbers 300+) and minority and first-generation college students. Regressions (2) and (4)–(8) split STEM/non-STEM into 14 department categories.

Table 4: Regressions of Study Time on Average Student and Class Characteristics

| | All Classes | | | Elective Classes | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Female | 0.175 | 0.245 | 0.286 | 0.109 | 0.159 | 0.272 |
| | (0.071) | (0.070) | (0.087) | (0.098) | (0.097) | (0.123) |
| ACT Reading* | 0.049 | 0.038 | -0.035 | 0.182 | 0.173 | 0.070 |
| | (0.041) | (0.041) | (0.049) | (0.059) | (0.059) | (0.083) |
| ACT Math* | 0.065 | 0.044 | -0.020 | 0.016 | 0.014 | -0.020 |
| | (0.048) | (0.048) | (0.055) | (0.063) | (0.065) | (0.079) |
| High School GPA* | -0.064 | -0.070 | -0.093 | -0.056 | -0.097 | -0.138 |
| | (0.046) | (0.045) | (0.047) | (0.064) | (0.062) | (0.069) |
| Average Grade | -0.308 | -0.265 | | -0.276 | -0.215 | |
| | (0.044) | (0.046) | | (0.070) | (0.071) | |
| STEM Class | 0.362 | | | -0.058 | | |
| | (0.050) | | | (0.084) | | |
| ln(Class Size) | -0.124 | -0.083 | | -0.146 | -0.158 | |
| | (0.027) | (0.028) | | (0.049) | (0.051) | |
| Dept. FE | No | Yes | No | No | Yes | No |
| Class FE | No | No | Yes | No | No | Yes |
| Observations | 866 | 866 | 541 | 346 | 346 | 206 |

Note: $*$ indicates that the variable is z-scored. Observations are at the class–cohort level. Additional controls include an indicator for upper-level classes (those with class numbers 300+), % minority, and % first-generation. Regressions (2) and (5) split STEM/non-STEM into 14 department categories.

grades are associated with *less* study time even after we condition on category fixed effects. This suggests that grades are relative measures of accomplishment and, in conjunction with assigned workload, may be a tool used by professors to influence demand for their courses.

## 3 Demand-Side Model

The descriptive results in Section 2 revealed significant differences in grading and study times across departments. Motivated by these descriptive patterns, we next develop a model of how students make course choices and study student effort decisions. These decisions are made in part as a response to professor choices on 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 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 a 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 nongrade preferences for particular courses. How each of these equations is estimated and which identification assumptions we rely on for the estimation are described in Section 4.

### 3.1 Choice set

Student $i$ chooses $n_i$ courses from a subset of all courses $\mathcal{J}_i \subset [1, \ldots, J]$, 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 modifications 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 Appendix B.1 for a description of the supplemental data that we collected and of how we utilized them to form the choice sets.

### 3.2 Course payoffs

We specify the payoff for a particular course $j$ as dependent on student $i$'s nongrade preference for the course, $\delta_{ij}$, 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.[14] The individual's realized utility from choosing course $j$ and exerting $s_{ij}$ units of effort is given by:

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

We parameterize $\phi_i$ as dependent on the student's gender. Denoting a female student $i$ as $w_i$, $\phi_i = \phi_0 + \phi_1 w_i$. The costs of studying in course $j$, $\psi_{ij}$, are specified to depend on $w_i$ and a set of characteristics $X_i$, defined in Table 5. The costs of studying also depend on a shock $\zeta_{ij}$ that is revealed after the student has chosen his or her courses, implying that students form expectations

[14]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 which case students would be choosing among all possible bundles of courses. 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.

over the realizations of $\zeta_{ij}$ when making their course choices. In particular, we specify $\psi_{ij}$ 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 $\zeta_{ij}$ is log-normally distributed.

Preferences for courses net of grades and study costs, $\delta_{ij}$, 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 $j$th 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. Embedded in $Z_{2ij}$ is the recognition that certain course combinations are required for particular majors. At UK, students must choose a major after their sophomore year. For juniors and seniors, we take their major choice as given; one of the components of $Z_{2ij}$ is then whether the course fulfills a requirement in the chosen major. For freshmen and sophomores, we approximate how courses affects future majors by controlling for the number of courses for which $j$ is a prerequisite. For sophomores, we also control for whether the student has taken courses in a particular department in previous semesters. We also include interactions between cohort and courses that fulfill particular university requirements in $Z_{2ij}$. The full listing of what is included in $Z_{1i}$ and $Z_{2ij}$ is shown in Table 5.

We then parameterize $\delta_{ij}$ as:

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

where $\delta_{0j}$ are course fixed effects and $\epsilon_{ij}$ is $i$'s unobserved (to the econometrician) preference for $j$ that 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 $\phi_i$, $\psi_{ij}$ and $\delta_{ij}$ into (1) yields:

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

Estimating department preferences, preferences for grades, and study costs separately by student

Table 5: List of Controls

| | |
|---|---|
| $w_i$ | gender: affects department-specific academic preparation and preferences as well as study costs and grade preferences |
| | *Covariates for academic preparation and cost of study effort* |
| $X_i$ | ACT reading, ACT math, high school grades, minority, first generation, unobserved type |
| | *Covariates for preferences that vary by department* |
| $Z_{1i}$ | ACT reading, ACT math, high school grades, unobserved type |
| | *Covariates for preferences that vary by class match* |
| $Z_{2ij}$ | female × female professor; freshman and sophomore × STEM × 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 is a university core requirement in the "statistical inferential reasoning" area, typically taken during the sophomore year; CIS/WRD110 is a university core requirement in the "composition and communications" area, typically taken during the freshman year.

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 low ($\delta_{1k(j)}$ negative) while the preferences for grades and cost of effort are relatively equal across men and women ($\phi_1$ and $\psi_1$ 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 private information that the student has that is not observed by the econometrician. Hence, there may be some students who have low study costs and have preferences for courses in particular departments beyond what is captured in the observed elements of $X_i$ and $Z_{1i}$. 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 5. To account for selection on unobservables, namely, the student's

private information about his or her own abilities, one of the characteristics in $X_i$ is the individual's unobserved type.

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, $\eta_{ij}$. Given study effort $s_{ij}$, $g_{ij}$ is specified as[15]:

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

Professors' grading policies are then choices over an intercept, $\beta_j$, and a return to academic preparation and effort, $\gamma_j$. 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.[16] After we substitute the grading process (5) into the utility function (4), the optimal study effort given a realization of $\zeta_{ij}$ can be found by differentiating $U_{ij}(s_{ij})$ with respect to $s_{ij}$:

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

A higher $\gamma_j$ increases the incentives to study, resulting in more study effort. Those who value grades more (have higher values of $\phi_i$) and have lower study costs (lower values of $\psi_{ij}$) 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:

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

[15] Note that any biases in the grading process against particular groups would manifest at the department level. Thus, we cannot distinguish between department-specific abilities and department-specific biases in the grading process.

[16] Students have a number of formal and informal resources to learn about grading policies, be it through friends or course syllabi. Course evaluations, which show average expected grades for courses, are online and publicly available. 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, $\zeta_{ij}$, but that $\zeta_{ij}$ is unknown at the time of course selection. The shock to grades, $\eta_{ij}$, 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 $\zeta_{ij}$s. Taking expectations over $\zeta_{ij}$ and $\eta_{ij}$ 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 we substitute the best study effort responses from (6) and the corresponding expected grades given in (8) into (1) and take 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:

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

$$\text{subject to:} \sum_{j=1}^{J} d_{ij} = n_i,\ \ d_{ij} \in \{0,1\}\forall j$$

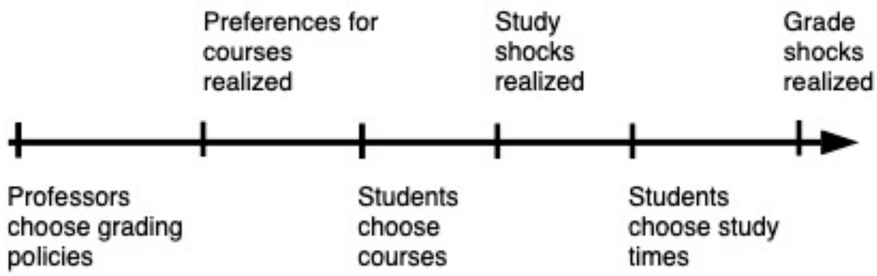
where $n_i$ is taken as given. We then obtain our third estimating equation by solving the maximization problem in (10).

# 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.

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 as outlined in Figure 1. In particular, new information on the costs of studying are revealed after the course choice decisions

Figure 1: Model Timing




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:

$$\begin{aligned} g_{ij} &= \beta_j + \gamma_j \left[ w_i(\alpha_{1k(j)} - \psi_1) + X_i(\alpha_{2k(j)} - \psi_2) + \ln(\phi_0 + \phi_1 w_i) + \ln(\gamma_j) - \psi_0 \right] + \eta_{ij} - \gamma_j \ln(\zeta_{ij}) \\ &= \theta_{0j} + \gamma_j \left( w_i \theta_{1k(j)} + X_i \theta_{2k(j)} \right) + \eta^*_{ij} \end{aligned} \tag{11}$$

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 $\{\theta_{0j}, \theta_1, \theta_2\}$ and the structural slopes, the $\gamma_j$s—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 $\theta$s by some factor and scaling down the $\gamma$s by the

same factor would be observationally equivalent. We set one $\gamma_j$ 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 $\gamma_j^N$, where $\gamma_j^N = \gamma_j / C_{k(j)}$. Similarly, we estimate $\theta_{1k(j)}^N$ and $\theta_{2k(j)}^N$, where $\theta_{1k(j)}^N = \theta_{1k(j)} C_{k(j)}$ and $\theta_{2k(j)}^N = \theta_{2k(j)} C_{k(j)}$.

The variation in the data used to identify $\{\theta_1^N, \theta_2^N\}$ comes from the relationship between student characteristics and grades in each department. The variation in the data used to identify the $\gamma_j^N$s is how these characteristics translate into grades relative to how they do so in the normalized courses. Key to separately obtaining $\gamma_j^N$ from $\{\theta_1^N, \theta_2^N\}$ is the parameterization of course-specific abilities as a weighted index of the characteristics of the students, where the weights vary at the department level (see Equation (5)). Absent this restriction, we would not be able to separate out the parameters that are choices of the professors (the $\gamma_j$s) from course-specific abilities.

## 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:

$$\begin{aligned} \ln(s_{ij}^*) &= \ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i) - \ln(\zeta_{ij}) && (16) \\ &= \kappa_0 + w_i \kappa_1 - X_i \psi_2 + \ln(\gamma_j) - \ln(\zeta_{ij}) && (17) \end{aligned}$$

where

$$\begin{aligned} \kappa_0 &= \ln(\phi_0) - \psi_0 && (18) \\ \kappa_1 &= \ln(\phi_0 + \phi_1) - \ln(\phi_0) - \psi_1 && (19) \end{aligned}$$

Recall that one $\gamma_j$ for every department was normalized in the grade equation. Substituting in $\hat{\gamma}_j^N C_{k(j)}$ for $\gamma_j$ in (17) and rearranging yields:

$$\ln(s_{ij}^*) - \ln(\hat{\gamma}_j^N) = \tilde{\kappa}_0 + w_i \kappa_1 - X_i \psi_2 + \kappa_{2k(j)} - \ln(\zeta_{ij}) \tag{20}$$

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 year in school (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 (20) at the class-cohort level yields our estimating equation:

$$\frac{\sum_i (l_i = l) d_{ij} \ln(s^*_{ij})}{\sum_i (l_i = l) d_{ij}} - \ln(\hat{\gamma}^N_j) = \tilde{\kappa}_0 + w_{jl}\kappa_1 - X_{jl}\psi_2 + \kappa_{2k(j)} - \ln(\zeta_{jl}) \tag{21}$$

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 (21) 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 $\gamma$s, solving for $\gamma$s that are normalized with respect to one course rather than to one course in each department. Namely, let $\gamma^P_j = \gamma^N_j \exp(\kappa_{2k(j)})$. $\hat{\gamma}^P_j$ then provides an estimate of $\gamma_j / C_1$.

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 results from the study cost shocks, $\ln(\zeta_{ij})$, mattering more in classes where $\gamma_j$ is high. In particular, the residual from Equation (11) can be written as:

$$g_{ij} - E(g_{ij}) = -C_1 \gamma^P_j \ln(\zeta_{ij}) + \eta_{ij} \tag{22}$$

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

$$\sigma^2_j = (\gamma^P_j)^2 \kappa_3 + \sigma^2_\eta \tag{23}$$

where $\kappa_3 = C^2_1 \sigma^2_{\ln(\zeta)}$. Regressing $\sigma^2_j$ on $(\gamma^P_j)^2$ then gives us an estimate of $C^2_1$ up to the variance of the study cost shock.

We can recover an estimate of $\sigma^2_{\ln(\zeta)}$ using the residuals of Equation (21) for cohort–course combinations with one student. Let this estimate be given by $\hat{\sigma}^2_{\ln(\zeta)}$. 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 (21) and where the $\ln(\zeta_{ij})$s are drawn from a normal distribution and with mean zero and variance calibrated so that, after censoring into bins, the censored residuals have variance equal to $\hat{\sigma}^2_{\ln(\zeta)}$. This calibrated variance then allows us to recover $C_1$.

### 4.3 Estimation of utility 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 $\gamma$s follow from the previous steps. With these taken as given, the variation in the data that identifies $\phi_0$ and $\phi_1$ 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 $\phi_0$; the difference in this sorting behavior between men and women identifies $\phi_1$.

We use simulated maximum likelihood coupled with a nested fixed-point algorithm to estimate the choice parameters.[17] To illustrate the approach, denote as $K_i$ the set of courses chosen by $i$. Denote as $M_i$ the highest payoff among the unchosen courses:

$$\begin{aligned} M_i = \max_{j \notin K_i} \quad & \big\{(\phi_0 + \phi_1 w_i)\big(\theta_{0j} + \gamma_j(w_i\theta_{1k(j)} + X_i\theta_{2k(j)} - 1\big) \\ & +\delta_{0j} + w_i\delta_{1k(j)} + Z_{1i}\delta_{2k(j)} + Z_{2ij}\delta_3 + \epsilon_{ij}\big\} \end{aligned} \tag{24}$$

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

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

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

Since the $\epsilon_{ij}$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 $r$th draw of the unchosen $\epsilon_{ij}$s and as $R$ the

[17] If individuals were choosing one course, estimation of the parameters in (9) would follow a multinomial logit. Students, however, choose bundles of courses. Even though the structure of the model is such that there are no complementarities from choosing particular combinations of courses, the probability of choosing a particular bundle does not reduce to the probabilities of choosing each of the courses separately.

number of simulation draws, the full log-likelihood function is then given by[18]:

$$\ln \mathcal{L} = \sum_i \ln \left( \left[ \sum_{r=1}^{R} \prod_{j=1}^{J} \left(1 - G(M_{ir} - \overline{U}_{ij})\right)^{d_{ij}} \right] / R \right) \tag{25}$$

While in theory one could estimate all of the choice parameters $\{\delta_{0j}, \delta_{1k(j)}, \delta_{2k(j)}, \delta_3, \phi_0, \phi_1\}$ by solving for the parameter values that maximize Equation (25), the large number of courses makes doing so computationally challenging. To circumvent this issue, 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 $\delta_{0j}$ directly to data on enrollment shares. This approach imposes that the predicted enrollment shares exactly match the observed enrollment shares. The details of the fixed-point algorithm can be found in Appendix C.2.

The remaining structural parameters from the study effort estimation, Equation (21), are the study cost intercept, $\psi_0$, and women's (relative) study costs, $\psi_1$. These can be recovered using:

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

The remaining structural parameters of the grade equation, Equation (11), are the course intercepts, $\beta_j$, and women's ability parameters $\alpha_{1k(j)}$. These can be recovered using:

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

## 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 $\pi_s$ is the unconditional probability of the $s$th value. Types are identified through the correlation of students' grades in each of their chosen courses and the probabilities of choosing different course combinations after conditioning on the observables. In practice, we set $S$ to three.

[18] Our setup is similar in spirit to that in Nevo et al. (2005). The estimator in Nevo et al. (2005) randomly samples rankings of chosen options, computes likelihood contributions conditional on rankings, and averages across 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.

Including unobserved types helps account for selection into courses based in part on student abilities not captured by our observed measures. For example, there may be a type who has especially high grades in STEM courses and who as a result chooses more STEM courses. However, this type may also have a preference for STEM courses on top of any grade effect, and this too is accounted for through the type component of $Z_{1i}$.

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 & Jones (2003) and Arcidiacono & Miller (2011), we can estimate some of the parameters in a first stage via a modified expectations maximization (EM) algorithm. Our full estimation procedure is described in Appendix C.1.

## 4.5 Discussion

Our model of course choices requires a number of assumptions. Here, we discuss a few of these as well as what patterns in the data we would miss if these assumptions were violated. In particular, we consider our assumptions regarding the timing of information, information about the grading processes, course bundling, and the way gender impacts course preferences.

### 4.5.1 Timing of information

Our model relies on study cost shocks and grade shocks that are revealed after course choices are made. In the case of private information about how well a student will perform in a particular class, this will generally work to attenuate our estimates of comparative advantage, provided that comparative advantage and nongrade preferences do not move in opposite directions. For example, suppose that women have a comparative advantage in non-STEM from a grade perspective and also have a preference for non-STEM subjects. Both of these factors mean that those women who take STEM courses will have been more on the margin of doing so. As a result, they will be more selected on this private information than their male counterparts, leading to upward-biased estimates of female ability in STEM courses.

Private information on study costs in particular classes could work in the same direction. Unlike in the case of grades, however, we do not have different study costs across different departments. Hence, to the extent that men have stronger preferences for STEM courses than women and that this in turn is associated with lower study costs in STEM courses, we might expect women to study

relatively less than men in STEM courses.[19] Adding an interaction between female and STEM department in Equation (21) shows a positive (but insignificant) coefficient on this term. Similarly, we find a small insignificant coefficient on this interaction when we add it to our descriptive results where we include course fixed effects in the study hours regression. Both these features point toward private information on study times not having a substantial effect on our model.[20]

#### 4.5.2 Information on instructor grading policies

We have assumed that students know the course-specific intercepts ($\beta_j$s) and slopes ($\gamma_j$s) for each class. The model can incorporate some limited forms of measurement error. For example, if a student were equally overoptimistic in all courses, then the overoptimism would cancel out in the choice problem; it would not affect the utility ranking of the courses. If all students were overly optimistic about grades in particular departments, then this too would not affect our estimates of the value of grades, even if this overoptimism varied between men and women, as these would be captured in the course fixed effects ($\delta_{0j}$) and women's departmental preference effects ($\delta_{1k(j)}$). In the latter case, however, we would mistakenly attribute some of female departmental preference effects to differential optimism.

However, we might expect differential uncertainty over the grading process, as some students would have had more time to learn about the processes or could be better networked. If this were the case, we would expect upperclassmen to sort better into classes that reward their abilities than lowerclassmen. We might also expect women to be more informed about the grading processes of classes that are taken by more women. Given the model estimates, we can simulate course choices and test whether certain groups have higher or lower expected grades from their simulated choices than what would be expected from their actual choices. For example, if upperclassmen were more informed than lowerclassmen, then upperclassmen (lowerclassmen) would have higher (lower) expected grades for their actual choices than for their model-simulated choices. Appendix A.1 spells out how we test for differential sorting and provides evidence that our model is able to match the sorting patterns seen in the data.

[19]Because we find that women value grades more than men and have lower study costs, they study more than men in both STEM and non-STEM courses. Here, we are referring to the difference in difference.

[20]We can further test whether preferences for courses affect study costs by using additional information from the evaluation data on why students took the course. Students could report that they took the course as an elective, as a university requirement, or as a major requirement. Regressing the individual responses of log study hours, a course fixed effect, and indicators for each reason shows little difference in study times across reasons for taking the course. There is no difference between those who took the class to fulfill a university requirement and those who took the class to fulfill a major requirement; those who took the class as an elective reported slightly (less than 4%) lower study hours.

### 4.5.3 Course bundling

Course choices from the model can also be simulated to evaluate whether we are missing key features of the data regarding how students bundle courses. Namely, we can examine whether the model overpredicts the degree to which students take only classes that have high workloads or only classes that have low workloads. Similarly, given that STEM classes are associated with higher workloads, we can examine whether the model misses how much students balance STEM classes with non-STEM classes. We show the results of this exercise in Appendix A.2. Despite not allowing for interactions between courses in the student utility function, our model matches the data (both overall and focusing only on upperclassmen) on the within-student distribution of workloads and share of classes in STEM.[21] While this may be surprising, we also show in Appendix A.2 that study times are quite low at UK, implying that balancing workloads may be of limited importance to these students.[22]

### 4.5.4 Mechanisms for the gender gap

Although our model includes a variety of mechanisms that could explain the gender gap in STEM, there may be features of key STEM courses that deter women but are not captured by the model. In Appendix A.3, we again use the model to see how well it predicts the share of women in various subsets of classes such as those required for a STEM major and those with an enrollment of at least one hundred students. The simulated and actual shares of women are quite close in all cases.

Finally, it may also be the case that women perform better in classes taught by women, a feature not allowed for in our model. We estimate a version of our model where we allow for an interaction between female and female professor in our grade production process. 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.

[21] That the model performs equally well in predicting the distribution of STEM classes for upper- and lowerclassmen despite the inclusion of extensive controls for major for upperclassmen suggests that whatever hidden information there is with regard to intended major for lowerclassmen does not substantially affect our model.

[22] The low levels of studying are consistent with the sharp declines in studying over time observed in the United States. Babcock & Marks (2011) show that reported study time outside of class in 2003 was between 13 and 14 hours per week, which is approximately half of what was reported in 1961.

## 5 Demand-Side Estimates

### 5.1 Preference estimates

Table 6 presents a subset of the preference parameters. Recall that the parameter on expected grades is identified from within-department variation in how abilities are rewarded in different classes. 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. Female students prefer classes with female professors, with the estimate suggesting that women would have to be compensated by approximately one-tenth of a grade point to enroll in the same class taught by a male professor.[23]

In the first column of the second panel of Table 6, we show women's preferences (relative to men's) for different departments, with the omitted category being Agriculture. The 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 0.47 grade points.

These departmental preferences of women emerge after we account for grade considerations and sorting on other ability factors.[24] 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 (see Table 9). 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 6 how nongrade 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.[25] 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

[23] The coefficient may be biased upward due to aggregation of departments. To the extent that 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 (Carrell et al. (2010), Rask & Bailey (2002), Hoffmann & Oreopoulos (2009)).

[24] See Jacob et al. (2013), Jiang (2019), Kaganovich et al. (2021), Wiswall & Zafar (2015), and Zafar (2013) for other examples in the literature exploring nongrade preferences for departments or majors.

[25] 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.

classes.

Table 6: Estimates of Preference Parameters

| Preference for: | Coef. | Std. Error | | |
|---|---|---|---|---|
| Expected grades ($\phi$) | 0.927 | (0.006) | | |
| Female x expected grade | 0.257 | (0.009) | | |
| Female x female professor | 0.141 | (0.007) | | |
| Departments | Female | ACT read* | ACT math* | HS GPA* |
| **Biology** | 0.505 | -0.238 | 0.033 | -0.136 |
| Psychology | 0.161 | -0.375 | 0.047 | -0.202 |
| Education & Health | 0.140 | -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 | -0.153 | -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 | -0.294 | 0.131 | -0.040 |
| Communications | -0.400 | -0.247 | 0.036 | -0.090 |
| **Engineering** | -0.898 | -0.275 | 0.493 | 0.048 |

Note: The specification includes study costs, class-specific intercepts, parameters on university and major requirement courses, unobserved types, a measure of the "option value" of the course (log of the new courses opened up by taking this course), and coefficient on upper-/lowerclassmen x course level and upper-/lowerclassmen x STEM x gender. See Appendix Table B.4 for complete results. $*$ indicates that the variable is z-scored. *Female* is women's nongrade preference for departments relative to men's. Department preferences are relative to agriculture. STEM departments are in bold.

## 5.2 Study effort estimates

Estimates of the study cost parameters are presented in Table 7. Women have 5.4% lower study costs than men, though the estimate is insignificant. Conditional on having the same observed characteristics and taking the same class, women study 30% more than men (a result that is statistically significant), but our estimates of $\phi_0$ and $\phi_1$ 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 $\gamma$ 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 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.

Table 7: Estimates of Study Costs and Departmental Returns to Studying

| | Study Effort Coef. ($\psi$) | Std. Error | Department | Median $\gamma$ Coef. |
|---|---|---|---|---|
| Female | -0.054 | (0.007) | **Engineering** | 0.441 |
| ACT Reading Score* | -0.057 | (0.050) | **Biology** | 0.308 |
| ACT Math Score* | 0.065 | (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 $\gamma$.

## 5.3 Grade estimates

The estimated department-specific ability weights, the $\alpha$s, are given in Table 8. These are calculated by taking the reduced-form $\theta$s, undoing the normalization on the $\gamma$s, and subtracting the part of the component that reflects study time (taken from $\psi$). 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 2: Women have higher grades than men in both STEM and non-STEM 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 (noneconomics) Social Sciences, Psychology, English, and Languages.

For all STEM categories but one (Biology), a standard deviation increase in the ACT math score has a larger impact on grades than a standard deviation increase in the high school GPA. This

is in contrast to the reduced-form results presented in Table 3, which suggested that high school grades were more important, even in STEM. What Table 3 misses is the differential returns to studying and ability across courses (heterogeneity in $\gamma_j$) and their correlation with course-specific mean grades (heterogeneity in $\beta_j$). The classes with higher returns to studying and course-specific ability (higher $\gamma_j$) are also classes that tend to have lower average grades (lower $\beta_j$).[26] Students with relatively low ACT math scores disproportionately sort into STEM classes where $\gamma_j$ is low. Since these classes also tend to have higher average grades, this sorting works to attenuate the coefficient on ACT math scores in Table 3.

Table 8: Estimates of Department-Specific Ability Weights ($\alpha$)

| | Female | ACT read* | ACT math* | HS GPA* |
|---|---|---|---|---|
| Education & Health | 0.522 | 0.164 | 0.407 | 0.750 |
| Communications | 0.372 | 0.167 | 0.167 | 0.960 |
| Regional Studies | 0.361 | 0.071 | 0.660 | 1.020 |
| Agriculture | 0.171 | 0.223 | 0.613 | 1.121 |
| Psychology | -0.055 | 0.415 | 0.517 | 0.940 |
| English | -0.072 | 0.296 | 0.453 | 1.006 |
| Languages | -0.140 | 0.311 | 0.530 | 0.929 |
| Social Sciences | -0.199 | 0.480 | 0.408 | 1.082 |
| **Math** | -0.244 | -0.074 | 1.594 | 0.984 |
| Mgmt. & Mkting | -0.308 | 0.166 | 0.440 | 0.985 |
| **Biology** | -0.439 | 0.166 | 0.648 | 0.827 |
| **Engineering** | -0.441 | -0.019 | 0.632 | 0.362 |
| **Econ., Fin., Acct.** | -0.547 | 0.146 | 0.980 | 0.917 |
| **Chem. & Physics** | -0.708 | 0.042 | 1.286 | 1.166 |

Note: $*$ indicates that the variable is z-scored. STEM departments are in bold. Departments are sorted by women's $\alpha$. See Table B.5 for the corresponding standard errors.

With the estimates of the grade production process, we can calculate the expected grades for students in a representative class for each department. We create a representative class by taking an enrollment-weighted average of the $\beta$s and $\gamma$s for each department. We calculate the expected grades in these representative classes separately by gender for two groups of students. The estimate for the first group is unconditional: We look at all students and weight by the total number of classes that each student takes. The second estimate conditions on the student taking classes in the department, where the weights are now given by the number of classes taken in the particular department.

The results are presented in Table 9 and are sorted based on women's unconditional grades, which are shown in the first column. Four patterns stand out. First, there is positive selection into

[26]Recall from Table 4 that students in classes with higher average grades report fewer study hours.

STEM courses: generally those who take STEM classes perform better than the average student. This is the not the case for all departments. Indeed, the second pattern is that negative selection is more likely to occur in departments with higher grades. Third, women are disproportionately represented in departments that give higher grades to the average student. Of the five departments that give the lowest grades (all of which fall under the STEM umbrella), women are underrepresented relative to the overall population in all but one, Biology. Finally, and consistent with Table 8, women have a comparative advantage in non-STEM courses. In all non-STEM categories, the unconditional expected grades for women are higher than those of men (and the difference is statistically different from zero), in part because women study more. For STEM categories, the unconditional expected grades are similar across genders. This occurs despite women studying more and is especially the case in the STEM classes where the returns to studying are highest.

Table 9: Expected GPA for Average Classes by Department, Unconditional and Conditional on the Student Taking Courses in That Department

| | EGPA Females Unconditional | EGPA Females Conditional | EGPA Males Unconditional | EGPA Males Conditional | Share Female |
|---|---|---|---|---|---|
| Education & Health | 3.49* | 3.47 | 3.31 | 3.23 | 0.693 |
| Communications | 3.44* | 3.38 | 3.22 | 3.13 | 0.558 |
| Agriculture | 3.34* | 3.21 | 3.20 | 2.86 | 0.572 |
| Mgmt. & Mkting | 3.20* | 3.36 | 3.12 | 3.28 | 0.500 |
| Languages | 3.20* | 3.24 | 3.10 | 3.09 | 0.549 |
| Regional Studies | 3.18* | 3.28 | 2.96 | 3.05 | 0.662 |
| Social Sciences | 3.09* | 3.07 | 2.98 | 2.86 | 0.505 |
| English | 3.08* | 3.09 | 2.93 | 2.97 | 0.652 |
| Psychology | 2.96* | 2.95 | 2.81 | 2.71 | 0.669 |
| **Econ., Fin., Acct.** | 2.63 | 2.88 | 2.66 | 2.86 | 0.377 |
| **Biology** | 2.56* | 2.74 | 2.54 | 2.72 | 0.599 |
| **Math** | 2.53 | 2.57 | 2.54 | 2.65 | 0.465 |
| **Engineering** | 2.48 | 2.83 | 2.54 | 2.93 | 0.183 |
| **Chem. & Physics** | 2.30* | 2.54 | 2.39 | 2.68 | 0.473 |
| Overall | | | | | 0.518 |

Note: $*$ denotes statistically significant differences between women's and men's unconditional EGPA at the 5% level. The comparisons across the unconditional and conditional EGPAs for each gender separately are different for every category at the 5% level. "Share Female" is % of students enrolled in courses offered in the category who are female. "Unconditional" represents the avg. grade outcome under the assumption that the entire student population enrolls in the course and weighted by the number of courses the student takes. STEM categories are in bold. Categories are sorted by women's unconditional grades. See Table B.6 for the corresponding standard errors.

Each part of the demand model is also influenced by the unobserved types.[27] These types are largely distinguished by class performance. Type 1s have average grades of 3.4, type 2s of 2.6, and

[27] Recall that the types are identified from the unexplained correlation between a student's grades in different courses and his or her course choices.

type 3s of 1.3.[28] While the model allows for the possibility of one type being stronger in some departments and another type being stronger in others, we instead find that type 1s have higher abilities in all departments, followed by type 2s. We show in Online Appendix Table B.7 the share of classes that each type takes in each department. Type 1s disproportionately select into classes in STEM departments—especially Biology, Economics, and Engineering—implying positive selection on unobservables into these classes. Ignoring this selection would lead to upward-biased estimates of the grading intercepts for courses in these departments.

### 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 our attention on freshmen and sophomores because junior and seniors have already chosen their majors.[29] In all simulations, we change the parameters or characteristics for women to match the parameters or characteristics for men.[30]

We show in Table 10 the share of classes taken in STEM for women and men and the changes in these shares for women under different scenarios.[31] 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 10 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.

The predicted outcomes when women's preferences for grades are changed to be the same as men's ($\phi_1$ 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 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

[28] The shares of types 1, 2, and 3 are 62.6%, 29.7%, and 7.7%, respectively.

[29] 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.

[30] 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.

[31] Our counterfactual simulations hold the utilities of courses with $\gamma < 0.01$ fixed. See Appendix C.3 for how the counterfactual choice probabilities are calculated in the presence of capacity constraints.

Table 10: STEM Enrollment for Freshmen and Sophomores in Counterfactual Scenarios (Partial Eq)

| | | 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 | 43.4% | | 9.9 |
| (5) | Shift unobs. abil. in grades | 45.2% | | 8.1 |
| (6) | Equalize unobs. pref. for depts. | 41.3% | | 12.0 |
| (7) | Female professor effect turned off | 41.3% | | 12.0 |
| | Grade Around a B | | | |
| (8) | Same grades in all courses:◇ $\gamma_j = 0$ | 62.3% | 65.5% | 3.2 |
| (9) | Same $\gamma$ in all courses: $\gamma_j = 0.251$ | 60.7% | 67.9% | 7.2 |
| (10) | $\gamma_j$'s fixed at estimated values | 55.7% | 63.9% | 8.2 |

Note: †: Women's preference and ability parameters are adjusted to be identical to men's preferences and abilities. ◇: "Same grades in all courses" sets grades and $\gamma$s to be the same in all courses, which is equivalent to turning off the effects of grades in course utility. Counterfactuals are partial equilibrium, as the grading policies of professors are held fixed.

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 nongrade department preferences. Because men are stronger on the math ACT and this makes STEM classes more attractive both through grades and through the nongrade department preferences, equalizing observed abilities reduces the gender gap by 2.4 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 reduce the gender gap, but only by a small amount (0.3 percentage points). Overall, we find that the nongrade 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 10 examines the role that differential grading across fields plays in

driving the STEM gap. Each row fixes the average grades in each course to a B under three partial equilibrium settings. First, we set the $\gamma_j$s to zero. This is equivalent to removing grades from the utility function altogether; sorting on abilities still matters, but only through preferences. The results in the third-to-last row show substantial increases for both men and women in the share of their courses in STEM. That both groups see substantial increases is reflective of the much lower grades and higher effort demanded in STEM courses. However, the results are especially large for women, whose share of courses in STEM increases by more than 50% and reducing the gender gap to 3.2 percentage points. This occurs because women value grades more than men and because this effect is amplified further because of their comparative advantage in non-STEM courses.

The second scenario in the second-to-last row again fixes the average grades in each course to a B but now sets the $\gamma_j$s to the median value across all courses (0.251). The $\beta_j$s are then adjusted to equalize average grades across courses. Relative to that in the case where the $\gamma_j$s are set to zero, the STEM share for women (men) falls (rises). This is again reflective of women's comparative advantage in non-STEM subjects. The resulting gender gap is 7.2 percentage points. The third scenario (last row) fixes the $\gamma_j$s at their estimated values. Since STEM departments have higher $\gamma$s, this leads to lower STEM shares for both men and women, though the drop is larger for women. The estimated gender gap in this scenario is 8.2 percentage points, still substantially lower than the initial gap of 12.3 percentage points.

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 $\gamma$s). Consider regressions of average courses grades, $\overline{G}_j$, and workloads, $\gamma_j$, on log enrollment, $\ln(E_j)$, and course characteristics, $W_j$:

$$\overline{G}_j = W_j\vartheta_{G1} + \ln(E_j)\vartheta_{G2} + \varepsilon_{Gj} \tag{26}$$

$$\overline{\gamma}_j = W_j\vartheta_{\gamma 1} + \ln(E_j)\vartheta_{\gamma 2} + \varepsilon_{\gamma j} \tag{27}$$

$W_j$ includes faculty characteristics (e.g., adjunct/tenure track, professor gender), indicators for each department, an indicator for upper-division courses, and interactions between the upper-division and STEM indicators. $\varepsilon_{kj}$ is the unobserved professor preference for outcome $k$ in course $j$. Log enrollment is endogenous, and the 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 $\beta_j$ and $\gamma_j$ 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 ($\delta_j$) will have higher enrollments. The exogeneity assumption requires that instructor leniency be uncorrelated with innate course demand (as captured by $\delta_j$ 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 ($\delta_j$).

We estimate three versions of Equations (26) and (27): (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 $\gamma_j > 0$.[32] We show in Table 11 the coefficients on log enrollment in each regression and

[32]Recall that twenty-five courses had $\gamma_j$s that hit the zero constraint.

how removing the effects of log enrollment affects the average gap between STEM and non-STEM courses in the outcome.[33] Averaging across courses shows that STEM classes have grades that are 0.34 points lower than the grades of their non-STEM counterparts (row 3, column 1). Accounting for the endogeneity of log enrollment using our instrument (row 3, 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. To put these numbers in perspective, the standard deviation of log enrollment is 0.742, implying that a one-standard-deviation increase is associated with a -0.261-point decrease in the average course grade. The standard deviation of average course grades is 0.444, implying that a one-standard-deviation increase in log enrollment is associated with a decrease of 0.588 standard deviations in average course grades.

We show in the second set of columns in Table 11 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 $\gamma_j$ is 0.115, implying that the average value of $\gamma$ for STEM courses is over one standard deviation larger than that of non-STEM courses (row 3, column 4). Accounting for the endogeneity of log enrollment using our instrument (row 3, column 6) reduces this gap to 0.089 points.

Table 11: Relating Course Demand to Grades and Workloads

| | Average Grades | | | $\gamma$ | | |
|---|---|---|---|---|---|---|
| | Baseline | OLS | IV | Baseline | OLS | IV |
| Ln Enroll | | -0.066 | -0.352 | | 0.0072 | 0.0768 |
| | | (0.020) | (0.023) | | (0.0045) | (0.0052) |
| STEM Gap | 0.341 | 0.307 | 0.164 | -0.128 | -0.124 | -0.089 |

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 Table B.8 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 (26) and (27). We assume that professors choose grading policy parameters $\beta_j$ and $\gamma_j$ 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 ($\gamma$).[34] In partic-

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

[34]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

ular, we specify the professor's objective function to penalize deviations from ideal log enrollment, $e_{0j}$, and the professor's ideal average grade in the class, $e_{1j}$, and ideal workload, $e_{2j}$. These ideals depend on observed and unobserved characteristics of the professor.[35]

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 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 ($\beta$ and $\gamma$). The dependence on $\beta$ and $\gamma$ comes through the composition of the students who take the course. 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:

$$\overline{G}_j(\beta,\gamma) = \beta_j + \gamma_j\left[\frac{\sum_i^N P_{ij}(\beta,\gamma)\left[A_{ij}+\ln(\phi_i)-\ln(\psi_i)\right]}{\sum_i^N P_{ij}(\beta,\gamma)} + \ln(\gamma_j)\right] \tag{28}$$

$$\ln\left[E_j(\beta,\gamma)\right] = \ln\left[\sum_i^N P_{ij}(\beta,\gamma)\right] \tag{29}$$

Then, the objective function that professor $j$ maximizes is:

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

where the coefficient on ideal log enrollment is normalized to one.[36]

Professors choose grading policies given different innate demand for their courses. Absent the first term, the professor of course $j$ would set $\gamma_j$ to $e_{2j}$. Given $\gamma_j$, he or she would then set $\beta_j$ 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) $e_{0j}$ if grades and effort were set to their ideal levels, professors adjust grades (workloads) downward (upward) to move enrollment closer to

Appendix D for the model with professor effort, the description of how professor effort is measured, and the model results with this channel incorporated.

[35] This way of expressing the objective function is equivalent to a quadratic in each of the individual terms.

[36] 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.

the ideal.

### 6.3 Estimation

We use the first-order conditions of the professor's objective function to form our estimating equations. After we divide by two, these are given by[37]:

$$0 = -\left(\ln\left[E_j(\beta,\gamma)\right] - e_{0j}\right)\frac{\partial \ln E_j}{\partial \beta_j} - \lambda_1\left(\overline{G}_j(\beta,\gamma) - e_{1j}\right)\frac{\partial \overline{G}_j}{\partial \beta_j} \tag{31}$$

$$0 = -\left(\ln\left[E_j(\beta,\gamma)\right] - e_{0j}\right)\frac{\partial \ln E_j}{\partial \gamma_j} - \lambda_1\left(\overline{G}_j(\beta,\gamma) - e_{1j}\right)\frac{\partial \overline{G}_j}{\partial \gamma_j} - \lambda_2(\gamma_j - e_{2j}) \tag{32}$$

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

$$e_{lj} = W_{lj}\Psi_l + \varepsilon_{lj} \tag{33}$$

where $\varepsilon_{lj}$ is unobserved professor-specific tastes for the $l$th outcome. Given that there are two first-order conditions and three $\varepsilon$ terms and that we cannot recover three unobservables from two equations, we normalize $\varepsilon_{0j}$ to zero. When $l$ refers to enrollment, we specify $W_{lj}$ to include a constant term and the upper-division indicator, with the latter allowing instructors to prefer lower enrollments when the class is upper division. 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 $\varepsilon_{1j}$ and $\varepsilon_{2j}$ in part determine the optimal choice of $\beta_j$ and $\gamma_j$. 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 $W_{lj}$. That is, if all professors were forced to set their policy parameters to $\beta^0$ and $\gamma^0$, the corresponding enrollment would be uncorrelated with the unobserved preferences of the instructor.

[37]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 (31) and (32) with the first terms of each set to zero.

#### 6.3.1 Recovering $\Psi_0$, $\Psi_2$, and $\lambda_2$

We estimate the professor parameters in two steps. In the first step, we estimate $\Psi_0$ (preferences for ideal enrollment), $\Psi_2$ (preferences for ideal workload), and $\lambda_2$ (the weight on the ideal workload). Rearranging and differencing the first-order conditions given in Equations (31) and (32) to eliminate the $\lambda_1$ term and solving for $\gamma_j$, we obtain:

$$\gamma_j = (1/\lambda_2) \ln [E_j] A_j - \Psi_0 A_j + W_{2j}\Psi_2 + \varepsilon_{2j} \tag{34}$$

where $A_j$ is given by:

$$A_j = \left[ \frac{\partial \ln [E_j]}{\partial \beta_j} \frac{\partial \overline{G}_j}{\partial \gamma_j} \Big/ \frac{\partial \overline{G}_j}{\partial \beta_j} \right] - \frac{\partial \ln [E_j]}{\partial \gamma_j} \tag{35}$$

Both the first and second terms of Equation (34) are correlated with $\varepsilon_{2j}$, as $\varepsilon_{2j}$ affects enrollment through $\gamma_j$ and the corresponding derivatives of enrollment and grades with respect to $\gamma_j$. 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 $\beta^0$ and $\gamma^0$. In practice, we set $\beta^0$ and $\gamma^0$ to the median values across all courses.[38] The variation across classes is then driven by the innate demand for courses given fixed grading policies.

#### 6.3.2 Recovering $\Psi_1$ and $\lambda_1$

In the second step, we recover estimates of $\Psi_1$ (preferences for ideal grades) and $\lambda_1$ (the weight on ideal grades). Equation (31), the first-order condition with respect to $\beta_j$, can be rewritten as:

$$\overline{G}_j(\beta, \gamma) = -(1/\lambda_1) B_j \left( \ln [E_j] - \Psi_0 \right) + W_{1j}\Psi_1 + \varepsilon_{1j} \tag{36}$$

where $\Psi_0$ 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 $\beta^0$ and $\gamma^0$ in each course.

[38] Using alternative values for $\beta$ and $\gamma$ produces similar results.

### 6.4 Supply-side results

We show in Table 12 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 145. 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 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 in order of decreasing size of the ideal grade of professors in the department. Professors in the Management & Marketing and Education & Health departments have the highest ideal grades, while professors in the English and Mathematics departments have the lowest ideal grades. 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 12 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 (34) and (36). The first terms of each equation show how deviations from ideal enrollments (demand adjustments) impact professor choices.

Table 13 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 between Biology and English (the two extremes) due to demand factors is about 0.45 grade points. English has the lowest ideal

Table 12: Estimates of Professor Preferences

| | Ideal grade | | Ideal workload | | Ideal log enrl | |
|---|---|---|---|---|---|---|
| $\lambda$ (Preference Weight) | 2.969 | (0.375) | 46.060 | (3.987) | 1.000 | — |
| Constant | 2.596 | (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, $\lambda_0$, 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 Table D.3 for a specification that includes professor effort. STEM departments are in bold. The baseline department is Agriculture.

grades among all departments except Math (Table 12) yet offers grades around the median in equilibrium (Table 9) 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 13: 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 (34) and (36).

Using the results in Tables 12 and 13, 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 13), ii) level of course offerings (Table 12 rows 3–4), iii) rank of the instructor (Table 12 rows 5–8), iv) female professor representation (Table 12 rows 9–10), and v) departmental effects (Table 12 rows 11–23). The results are presented in Table 14.

We show in the first column of Table 14 how demand factors vary across STEM and non-STEM courses. These are calculated by weighting the department demand adjustments in Table 13 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 12 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 7%. 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 12).

Table 14: 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 | -0.0200 | -0.1130 | -0.2831 |
| Non-STEM grade | 0.0434 | 0.0131 | 0.0031 | 0.0080 | 0.0453 | 0.1134 |
| Diff. | 0.1519 | 0.0457 | 0.0110 | 0.0280 | 0.1582 | 0.3966 |
| Shares | 38.30% | 11.52% | 2.77% | 7.07% | 39.90% | |
| STEM workload | 0.0222 | 0.0211 | -0.0003 | -0.0021 | 0.0472 | 0.0911 |
| Non-STEM workload | -0.0089 | -0.0085 | 0.0001 | 0.0008 | -0.0189 | -0.0365 |
| Diff. | -0.0310 | -0.0296 | 0.0004 | 0.0030 | -0.0661 | -0.1275 |
| Shares | 24.34% | 23.22% | -0.33% | -2.32% | 51.83% | |

Note: The decompositions of STEM/non-STEM differences in grades and workloads ($\gamma$) are calculated by means of the $\Psi$ estimates of department-category intercepts, instructor rank, upper-level class, and female professor from Table 12 and the demand-side adjustments calculated in Table 13.

### 6.5 General equilibrium counterfactuals

Given the estimates of professor preferences, we can now examine how equilibrium grading practices would change in counterfactual scenarios. We focus on three sets of counterfactuals, the first two of which examine the sources of the grading differences across fields and their implications for STEM enrollment. First, we examine the effects of equating intrinsic demand for STEM and non-STEM courses. Second, we examine the effects of removing observed differences in preferences between STEM and non-STEM professors. Finally, we consider a counterfactual that could actually be implemented, examining the effects of mandating an average grade of a B in all courses. This last counterfactual mirrors the partial equilibrium case considered in the last row of Table 10 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.

The counterfactual results are presented in Table 15. The first row shows the data. The second row equates intrinsic demand. There are a number of demand-side mechanisms, including department preferences and previously passed courses, that may lead to heterogeneous demand for STEM and non-STEM courses. To equalize intrinsic demand, we first set grades to be the same in all courses. We then solve for a value $\Delta$ such that decreasing the utility of all STEM classes by $\Delta$ equates average class sizes in STEM and non-STEM courses. $\Delta$ thus quantifies the extent to which demand for STEM courses is greater than demand for non-STEM courses in a scenario where grades are irrelevant. Then, to simulate a counterfactual scenario where intrinsic demand is equal in STEM and non-STEM courses, we subtract $\Delta$ from the utility of all STEM classes and then solve for the equilibrium grading policies and the resulting course enrollments.[39] Given the same intrinsic demand, STEM courses would have lower enrollments because of professor preferences for lower grades and higher workloads. The gap in grades across STEM and non-STEM courses shrinks from 0.45 grade points in the baseline to 0.17 grade points when intrinsic demand is equalized. Hence, much of the differences in grades across STEM and non-STEM courses is due to substantially higher demand for STEM courses.

The next three rows of Table 15 equate observed professor preferences by averaging out differences due to the rank and gender composition of the instructors and the departmental effects. In particular, we modify the instructor preferences for ideal grades and/or ideal workloads given in Equation (33) as follows:

$$e_{lj}^{*} = \frac{\sum_{j'} W_{lj'}\Psi_l}{J} + \varepsilon_{lj} \tag{37}$$

This removes all systematic differences in ideal grades and/or workloads due to instructor characteristics or departments; however, it retains the idiosyncratic preference terms $\varepsilon_{lj}$. Given the new professor preferences, we then solve for the equilibrium grading and effort policies. Equalizing professor preferences across both ideal grades and workloads (row 3) reduces the average differences in grades across STEM and non-STEM to 0.24 points; the remaining difference is due to demand. Decreasing grading differences leads to increases in STEM enrollments, especially for women. Women

[39] We solve for the optimal grading and effort policies in this counterfactual scenario via a fixed-point algorithm. Given guesses of the grading and effort choices, we solve for student choice probabilities, taking into account capacity constraints as in C.3. We then update the grading and effort policies using the first-order conditions of the professor's maximization problem.

increase the share of classes that they take in STEM by 5.6 percentage points; the corresponding increase for men is 5.1 percentage points. We show in rows 4 and 5 that while both equalizing grades and equalizing workloads increase STEM participation, equalizing ideal grades has a larger effect, especially for women. Indeed, equalizing ideal workloads actually increases the STEM gap.

We show in the last three rows of Table 15 the partial and general equilibrium results of imposing that each class's average grade be a 3.0. Row 6 contains results equivalent to those in the last row of Table 10 but now including juniors and seniors. Here, the $\gamma_j$s are fixed at their estimated values. The share of STEM classes increases for men and women by 8.7 and 12 percentage points, respectively. We then show in row 7 what happens when professors are able to partially undo the effects of the policy by changing their workloads ($\gamma_j$s). 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 8 contains results equivalent to those in row 7 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 1.1 percentage points relative to that under a curve that affects all courses.

Table 15: Counterfactuals with Endogenous Professor Responses

| | | Class Size | | STEM Enrollment Share | | | Weighted Avg. Grade | | |
|---|---|---|---|---|---|---|---|---|---|
| | | STEM | Non-STEM | Overall | Female | Male | Overall | STEM | Non-STEM |
| (1) | Baseline | 82.6 | 45.0 | 41.8% | 34.6% | 49.5% | 3.016 | 2.756 | 3.207 |
| (2) | Equate Demand | 44.6 | 59.8 | 22.6% | 17.7% | 27.8% | 3.104 | 2.976 | 3.142 |
| (3) | Equate Prof Pref | 93.2 | 40.8 | 47.2% | 40.2% | 54.6% | 3.110 | 2.986 | 3.224 |
| (4) | Equate Prof Grade Pref | 89.1 | 42.4 | 45.1% | 38.4% | 52.3% | 3.109 | 2.997 | 3.202 |
| (5) | Equate Prof $\gamma$ Pref | 86.5 | 43.4 | 43.8% | 36.2% | 52.0% | 3.016 | 2.780 | 3.205 |
| | Grade Around $3^{\diamond}$ | | | | | | | | |
| (6) | Partial Eq.* | 103.1 | 36.9 | 52.2% | 46.6% | 58.3% | 3.000 | 3.000 | 3.000 |
| (7) | General Eq. | 100.6 | 37.9 | 50.9% | 45.1% | 57.2% | 3.000 | 3.000 | 3.000 |
| (8) | General Eq. Lower Div | 98.4 | 38.8 | 49.8% | 43.8% | 56.3% | 3.046 | 2.996 | 3.096 |

Note: "Weighted" means weighting by class enrollment. $\diamond$: "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 general equilibrium analysis. $*$: Partial equilibrium results corresponding to last row in Table 10. While in Table 10 we show the STEM shares for freshmen and sophomores, the results here are for all students. See Table D.4 for the changes to $\gamma$ and counterfactuals evaluated when professors directly influence enrollment.

# 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 factors that may deter enrollment. Using administrative data form the University of Kentucky, we estimate a model of course choices to understand what influences STEM enrollment and how those influences differentially affect men and women. While we show that a variety of factors influence how students choose courses, we find that differences in grading policies play an important role in suppressing STEM demand and that this is particularly true for female students.

One issue with policies aimed at reducing grading differences is that instructors may respond to these policies by changing other aspects of their courses. To capture these responses and to understand the source of grading differences more generally, our analysis treats grading policies 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 overall STEM participation by 9.1 percentage points (a 22% increase) and female STEM participation by 10.5 percentage points (a 30% 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, J. G., Kahn, L. B. & Speer, J. D. (2016), 'Cashier or consultant? entry labor market conditions, field of study, and career success', *Journal of Labor Economics* **34**(S1), 361–401.

Arcidiacono, P., Aucejo, E. M. & Spenner, K. (2012), 'What happens after enrollment? an analysis of the time path of racial differences in GPA and major choice', *IZA Journal of Labor Economics* **1**(1), 1–24.

Arcidiacono, P. & Jones, J. B. (2003), 'Finite mixture distributions, sequential likelihood and the EM algorithm', *Econometrica* **71**(3), 933–946.

Arcidiacono, P. & Miller, R. A. (2011), 'Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity', *Econometrica* **79**(6), 1823–1868.

Astorne-Figari, C. & Speer, J. D. (2018), 'Drop out, switch majors, or persist? The contrasting gender gaps', *Economics Letters* pp. 82–85.

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

Babcock, P. (2010), 'Real costs of nominal grade inflation? new evidence from student course evaluations', *Economic Inquiry* **48**(4), 983–996.

Babcock, P. & Marks, M. (2011), 'The falling time cost of college: Evidence from half a century of time use data', *Review of Economics and Statistics* **93**(2), 468–478.

Bagues, M., Labini, M. S. & Zinovyeva, N. (2008), 'Differential grading standards and university funding: Evidence from Italy', *CESifo Economics Studies* **52**(2), 149–176.

Bar, T., Kadiyali, V. & Zussman, A. (2009), 'Grade information and grade inflation: The Cornell experiment', *Journal of Economic Perspectives* **23**(3), 93–108.

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

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

Butcher, K. F., McEwan, P. J. & Weerapana, A. (2014), 'The effects of an anti-grade-inflation policy at Wellesley College', *Journal of Economic Perspectives* **8**(3), 189–204.

Carrell, S. E., Page, M. E. & West, J. E. (2010), 'Sex and science: How professor gender perpetuates the gender gap', *Quarterly Journal of Economics* **125**(3), 1101–1144.

Chan, W., Hao, L. & Suen, W. (2007), 'A signaling theory of grade inflation', *International Economic Review* **48**(3), 1065–1090.

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

Crocker, J., Luhtanen, R. K., Cooper, M. L. & Bouvrette, A. (2003), 'Contingencies of self-worth in college students: Theory and measurement', *Journal of Personality and Social Psychology* **85**(6), 894–908.

Epple, D., Romano, R. & Sieg, H. (2006), 'Admission, tuition, and financial aid policies in the market for higher education', *Econometrica* **74**(4), 885–928.

Ferreyra, M., Garriga, C. & Diaz, A. S. (2021), 'raising college access and completion: How much can free college help?', Working paper.

Freeman, D. G. (1999), 'Grade divergence as a market outcome', *Journal of Economic Education* **30**(4), 344–351.

Fu, C. (2014), 'Equilibrium tuition, applications, admissions, and enrollment in the college market', *Journal of Political Economy* **122**(2), 225–281.

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

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

Graeml, Y. (2019), 'Senators introduce bill seeking to help international stem students work in us', *The Daily Californian* .
**URL:** *www.dailycal.org/2019/06/12/senators-introduce-bill-seeking-to-help-international-stem-*

*students-to-work-in-us/*

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

Gross, N. (2019), 'Your gi bill benefits: Everything you need to know', *Military Times* .
**URL:** *www.militarytimes.com/education-transition/2019/07/20/gi-bill-benefits-guide/*

Hinz, G. (2019), 'STEM students would get free tuition under Il. bill', *Crain's Chicago Business* .
**URL:** *www.chicagobusiness.com/greg-hinz-politics/stem-students-would-get-free-tuition-under-illinois-bill*

Hoffmann, F. & Oreopoulos, P. (2009), 'A professor like me the influence of instructor gender on college achievement', *Journal of Human Resources* **44**(2), 479–494.

Insler, M., McQuoid, A. F., Rahman, A. & Smith, K. (2021), Fear and loathing in the classroom: Why does teacher quality matter?

Jacob, B., McCall, B. & Stange, K. M. (2013), College as country club: Do colleges cater to students' preferences for consumption?, Technical report, National Bureau of Economic Research.

Jiang, X. (2019), Women in stem: Ability, preference, and value.

Johnson, V. E. (2003), *Grade inflation: A crisis in college education*, Springer.

Kaganovich, M., Taylor, M. & Xiao, R. (2021), Gender differences in persistence in a field of study.

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

Latek, T. (2019), 'Lawmakers drawing up proposals for 2020 general assembly', *Kentucky Today* .
**URL:** *kentuckytoday.com/stories/lawmakers-drawing-up-proposals-for-2020-general-assembly*

McEwan, P. J., Rogers, S. & Weerapana, A. (2021), 'Grade sensitivity and the economics major at a women's college', *AEA Papers and Proceedings* **11**, 102–106.

Nelson, J. P. & Lynch, K. A. (1984), 'Grade inflation, real income, simultaneity, and teaching evaluations', *Journal of Economic Education* pp. 21–37.

Nevo, A., Rubinfeld, D. L. & McCabe, M. (2005), 'Academic journal pricing and the demand of libraries', *American Economic Review* **95**(2), 447–452.

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

Owen, A. (2010), 'Grades, gender, and encouragement: A regression discontinuity analysis', *The Journal of Economic Education* **81**, 217–234.

Piopiunk, M., Schwerdt, G., Simon, L. & Woessmann, L. (2020), 'Skills, signals, and employability: An experimental investigation', *European Economic Review* **123**, 1–25.

Rask, K. (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.

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

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

Rosenberg, M., Schooler, C., Schoenbach, C. & Rosenberg, F. (1995), 'Global self-esteem and specific self-esteem: Different concepts, different outcomes', *American Sociological Review* **60**(1), 141–156.

Sabot, R. & Wakeman-Linn, J. (1991), 'Grade inflation and course choice', *Journal of Economic Perspectives* pp. 159–170.

Seymour, E. & Hewitt, N. M. (1997), *Talking about leaving*, Westview Press, Boulder, CO.

Stevens, R. H. (2021), 'How congress can help get more women and girls involved in stem education', *The Hill* .
**URL:** *https://thehill.com/blogs/congress-blog/education/554019-how-congress-can-help-get-*

*more-women-and-girls-involved-in-stem*

Stinebrickner, R. & Stinebrickner, T. (2014), 'A major in science? Initial beliefs and final outcomes for college major and dropout', *Review of Economic Studies* **81**(1), 426–472.

Stinebrickner, R. & Stinebrickner, T. R. (2012), 'Learning about academic ability and the college dropout decision', *Journal of Labor Economics* **30**(4), 707–748.

Thomas, J. (2021), What do course offerings imply about university preferences?

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'.

Wiswall, M. & Zafar, B. (2015), 'Determinants of college major choice: Identification using an information experiment', *Review of Economic Studies* **82**(2), 791–824.

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

Zangenehzadeh, H. (1988), 'Grade inflation: A way out', *Journal of Economic Education* pp. 217–226.

Zubrickas, R. (2015), 'Optimal grading', *International Economic Review* **3**(56), 751–776.

# A Appendix – For Inclusion in Text

In this appendix, discuss some of the 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 draw preferences for each student and record $n$, the number of classes that he or she chose in the data;
2. beginning with the student with the earliest timestamp, we take the $n$ courses with the highest utility (the sum of expected utility plus the unobserved preference); and
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 function for every class ($\gamma_j$ and $\beta_j$). In reality, they may be uncertain about the true parameters: upperclassmen may be able to sort better than lowerclassmen, and women may be more informed about classes where there are more women. Both have implications for what the model predicts vs. what is seen in the data.

If upperclassmen have better information about grading processes, the model would overpredict (underpredict) how well lowerclassmen (upperclassmen) sort into classes that match their abilities. To test this, we simulate course choices using model estimates and calculate expected grades using

the simulated and actual choices. If the expected grades in the simulated courses were on average higher than those in the actual courses chosen by lowerclassmen, this would support upperclassmen being better informed. However, the difference between the actual and simulated expected grade points for freshmen, sophomores, and upperclassmen are all less than 0.01. For freshmen, simulated:actual expected grades are 2.982:2.985. For sophomores, they are 3.018:3.017. For upperclassmen, they are 3.079:3.070.

As a central finding of our model is that women value grades more than men, women may have better information in courses where women account for a greater share of enrollment. We would expect women to do a better job matching their abilities for these classes than what the model predicts. To test this, we calculate average expected grades for women in each class based on their actual choices and simulated choices. We then difference at the class level women's actual and model-simulated grades and regress this measure on the actual share of women in the class. The regression results show a small and negative relationship; 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 received from a course does not depend on the other chosen courses. This rules out students balancing workloads ($\gamma$) 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 $\gamma$ 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 (0.3080) is actually slightly lower than SD in the data (0.3202); the similar numbers for upperclassmen are 0.3070 and 0.3208. 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 0.3284 versus 0.3431 in the data; for upperclassmen, the corresponding predicted SDs are 0.3374 and 0.3552.

These results suggest that bundling may not be a primary enrollment strategy, at least in our data. While this may be surprising, the evaluation data show that average course study time is 2.6 hours per week, with 99% of classes associated with less than 9 hours of study time per week. For the limited number of students who have course bundles where all the classes are in the evaluation

Table A.1: Share of Women in Lower-Division STEM classes, Predicted vs. Actual

| | Lower-Division | | Plus Enroll≥100 | |
|---|---|---|---|---|
| | Actual | Simulated | Actual | Simulated |
| STEM, All | 0.458 | 0.452 | 0.488 | 0.482 |
| STEM minus Biology, All | 0.430 | 0.424 | 0.452 | 0.450 |
| STEM, Lowerclassmen | 0.478 | 0.469 | 0.505 | 0.496 |
| STEM minus Biology, Lowerclassmen | 0.448 | 0.444 | 0.469 | 0.467 |
| STEM, Required, All | 0.450 | 0.446 | 0.484 | 0.481 |
| STEM minus Biology, Required, All | 0.416 | 0.415 | 0.445 | 0.447 |
| STEM, Required, Lowerclassmen | 0.466 | 0.465 | 0.500 | 0.499 |
| STEM minus Biology, Required, Lowerclassmen | 0.430 | 0.437 | 0.460 | 0.466 |

data, the mean study hours are 12.2, suggesting substantial room for substitution between study and leisure.

## A.3 Flexibility of gender preferences

Although our model includes a variety of mechanisms to explain the gender gap in STEM, there may be features not captured by the model of 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. The first two columns of Table A.1 show the female share in actual and simulated lower-division courses, while the third and fourth columns also restrict to courses with 100 or more students. Rows 1 and 2 look at all STEM lower-division courses and all STEM lower-division courses except those in biology. Rows 3 and 4 examine the same courses but only for lowerclassmen. Rows 5 through 8 repeat the analysis from rows 1 through 4 but further restrict to classes required for a STEM major. For all specifications, the simulated female share is within one percentage point of the actual share, suggesting that our model captures well how women and men are distributed across courses in different departments.

# 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. construction of student choice sets,
2. our method of aggregating the departments into our fourteen categories,
3. additional parameters from the motivating regressions on grades and hours studied (Tables 3 and 4)
4. our sample selection procedures for the various structural estimation stages,
5. additional student preference parameters (expanding on Table 6),
6. additional tables with standard errors (Tables 8 and 9),
7. the share of courses 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 11).

## B.1 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 had already completed over the prior seven semesters (fall 2008–spring 2011) unless he or she is observed in the class again in fall 2012.
2. Class prerequisites: We compile lists of prerequisite classes (from the 2012 UK Undergraduate Bulletin) for every course. We use the student's academic history and close the choice set unless all prerequisites are met. If the student takes a class without having completed all requirements, we assume that an exemption was granted by the professor.
3. AP exams: Students have the option to bypass introductory courses in certain subjects (defined in the UK Undergraduate Bulletin) with a score of 3 or above on the corresponding AP exam.

Table B.1: 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 |

(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.

4. Room capacity: We have timestamps for all classes that students registered for. Using data on room capacity for each class, we find the timestamp of when/if the class hits capacity. For each student, we find the timestamp of the first class that he or she registered for. If the student's first timestamp is after the class's timestamp, the class is not in the choice set.

Table B.1 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. Over 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 consideration of AP exams (Restriction 3) or capacity constraints (Restriction 4) are marginal.

## B.2 Aggregation of departments

In Table B.2 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.

Table B.2: Aggregation of Departments

| Category | Departments |
|---|---|
| Agriculture | Agricultural Biotechnology, Agricultural Economics, Agricultural Ed, Agriculture General, Animal & Food Sciences, Biosystems & Agr Engineering, Environmental Studies, Forestry, Landscape Architecture, Plant Pathology, Plant & Soil Sciences, Sustainable Agriculture |
| Regional Studies | Appalachian Studies, Family Sciences, Gender & Women's Studies, Hispanic Studies, Latin American Studies |
| Communications | Arts Admin, Communication, Communication & Info Studies, Fine Arts – Music, Fine Arts – Theatre Arts, Schl of Journalism & Telecomm, Schl of Art & Visual Studies, Schl of Interior Design |
| 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** | Chemical & Materials Engineering, Civil Engineering, Computer Science, Electrical & Computer Engineering, Engineering, Mechanical Engineering, Mining Engineering, Schl of Architecture |
| Languages | Linguistics, Modern & Classical Languages, Philosophy |
| English | English |
| **Biology** | Biology, Entomology |
| **Mathematics** | Mathematics, Statistics |
| **Chem & Physics** | Chemistry, Earth & Environmental Sciences, Physics & Astronomy |
| Psychology | Psychology |
| Social Sciences | Anthropology, Geography, History, Political Science, Schl of Human Environmental Sciences, Sociology |
| Mgmt. & 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.

## B.3 Additional parameters from the motivating regressions

In Table B.3 we show estimates of the department indicator variables from Tables 3 and 4. The grade regression results show that the coefficients are lowest for the 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) 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.

Table B.3: Reduced-Form Regressions of Department Grade and Study Hours Parameters

| | Grades | | | | | | | | | | | | Study Hours | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | All Classes | | Upper Level | | Electives | | Upperclassmen | | STEM | | non-STEM | | All Classes | | Electives | |
| Department | Coef. | Std. Err. | Coef. | Std. Err. | Coef. | Std. Err. | Coef. | Std. Err. | Coef. | Std. Err. | Coef. | Std. Err. | Coef. | Std. Err. | Coef. | Std. Err. |
| Agriculture | | | | | | | | | | | | | -0.263 | (0.137) | -0.544 | (0.280) |
| Regional Studies | 0.014 | (0.031) | 0.099 | (0.054) | 0.005 | (0.046) | 0.063 | (0.041) | | | 0.041 | (0.029) | -0.071 | (0.132) | 0.106 | (0.267) |
| Communications | 0.244 | (0.025) | 0.210 | (0.035) | 0.310 | (0.037) | 0.214 | (0.033) | | | 0.229 | (0.023) | -0.081 | (0.134) | -0.122 | (0.271) |
| Education & Health | 0.341 | (0.026) | 0.421 | (0.034) | 0.228 | (0.044) | 0.395 | (0.033) | | | 0.359 | (0.025) | -0.363 | (0.133) | -0.162 | (0.291) |
| **Engineering** | -0.161 | (0.029) | -0.156 | (0.042) | 0.048 | (0.046) | -0.250 | (0.037) | | | | | 0.504 | (0.133) | 0.268 | (0.303) |
| Language | 0.104 | (0.027) | 0.036 | (0.042) | 0.134 | (0.041) | 0.132 | (0.036) | | | 0.096 | (0.025) | -0.169 | (0.127) | -0.032 | (0.260) |
| English | -0.102 | (0.034) | -0.170 | (0.053) | 0.007 | (0.053) | -0.101 | (0.044) | | | -0.086 | (0.032) | 0.006 | (0.138) | 0.077 | (0.268) |
| **Biology** | -0.408 | (0.028) | -0.213 | (0.043) | -0.349 | (0.042) | -0.427 | (0.038) | -0.387 | (0.031) | | | | | -0.173 | (0.303) |
| **Math** | -0.353 | (0.026) | -0.350 | (0.058) | -0.260 | (0.038) | -0.491 | (0.037) | -0.264 | (0.025) | | | 0.224 | (0.143) | | |
| **Chem. & Physics** | -0.487 | (0.027) | -0.286 | (0.092) | -0.449 | (0.039) | -0.487 | (0.037) | -0.420 | (0.026) | | | 0.101 | (0.145) | -0.377 | (0.292) |
| Psychology | -0.212 | (0.030) | -0.117 | (0.044) | -0.165 | (0.044) | -0.195 | (0.041) | -0.195 | (0.034) | | | 0.047 | (0.181) | 0.237 | (0.304) |
| Social Science | -0.022 | (0.025) | 0.057 | (0.036) | 0.002 | (0.038) | 0.038 | (0.033) | | | -0.049 | (0.024) | -0.140 | (0.128) | 0.064 | (0.261) |
| Mgmt. & Mkting | 0.249 | (0.031) | 0.363 | (0.036) | 0.070 | (0.066) | 0.298 | (0.036) | | | 0.229 | (0.029) | -0.407 | (0.146) | -0.101 | (0.301) |
| **Econ., Fin., Acct.**. | -0.143 | (0.027) | 0.051 | (0.036) | -0.100 | (0.045) | -0.126 | (0.034) | -0.018 | (0.026) | | | -0.028 | (0.135) | 0.053 | (0.296) |

Note: Agriculture is the excluded department category for the Grades regression. STEM departments are in bold. Engineering is the excluded category in the STEM-only specification. Biology (Math) is the excluded department category for the All Classes (Electives Only) Study Hours regression.

### B.4 Sample selection

We now describe our sample selection rules for the various stages of the structural 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 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 $\gamma$ for these courses set to 0.

Estimates of the grade parameters show an additional 7 courses where the estimate of $\gamma$ is less than 0.01 (estimates of $\gamma$ 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 $\gamma$ 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. Our sample is further cut to 2,411 once courses with values of $\gamma$ less than 0.01 are removed.

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 850.

For the professor estimation, we do not use courses where $\gamma$ 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.5 Additional student preference parameters

In Table B.4 we show the full set of student preference parameters (see Table 6 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.

### B.6 Tables with standard errors

Tables 8 and 9 in the text are presented without standard errors for readability. Tables B.5 and B.6 present the complete tables. We generate standard errors for the expected GPA results in Table B.6, using the empirical distribution of the grading policy parameters with model estimates as the mean and the inverse Hessian as the covariance matrix to draw 1,000 instances of the parameters.

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

In Table B.7 we show how the types sort into classes offered in each department. 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.4: 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 | 1.511 | (0.021) | 0.388 | (0.004) | 0.422 | (0.016) | -0.145 | (0.023) | 0.211 | (0.014) | 0.832 | (0.014) |
| | Major Req. (A) | | Major Req. (B) | | Upper-Lev. Class | | | | | | | |
| Upperclassmen x | 3.557 | (0.008) | 2.091 | (0.017) | 1.861 | (0.008) | | | | | | |
| | 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 | -0.400 | (0.022) | -0.247 | (0.016) | 0.036 | (0.018) | -0.090 | (0.016) | -0.987 | (0.043) | -1.484 | (0.021) |
| Education & Health | 0.140 | (0.023) | -0.336 | (0.018) | 0.010 | (0.019) | 0.075 | (0.017) | -1.227 | (0.047) | -2.063 | (0.026) |
| Engineering | -0.898 | (0.032) | -0.275 | (0.020) | 0.493 | (0.020) | 0.048 | (0.020) | -0.801 | (0.061) | -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 | (0.025) |
| English | 0.127 | (0.030) | 0.024 | (0.023) | -0.208 | (0.025) | 0.051 | (0.023) | -0.431 | (0.073) | -1.580 | (0.036) |
| Biology | 0.505 | (0.026) | -0.238 | (0.019) | 0.033 | (0.021) | -0.136 | (0.019) | -0.688 | (0.057) | -1.175 | (0.028) |
| Math | -0.133 | (0.026) | -0.153 | (0.017) | -0.193 | (0.019) | -0.200 | (0.017) | -1.265 | (0.049) | -1.095 | (0.023) |
| 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 | 0.161 | (0.025) | -0.375 | (0.020) | 0.047 | (0.022) | -0.202 | (0.019) | -0.816 | (0.055) | -1.067 | (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 | (0.076) | -1.551 | (0.033) |
| Econ., Fin., Acct. | -0.378 | (0.026) | -0.294 | (0.018) | 0.131 | (0.020) | -0.040 | (0.018) | -1.760 | (0.062) | -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.

Table B.5: Estimates of Department-Specific Ability Weights ($\alpha$) with Standard Errors

| | Female | | ACT read* | | ACT math* | | HS GPA* | |
|---|---|---|---|---|---|---|---|---|
| Education & Health | 0.522 | (0.165) | 0.164 | (0.061) | 0.407 | (0.080) | 0.750 | (0.115) |
| Communications | 0.372 | (0.174) | 0.167 | (0.067) | 0.167 | (0.052) | 0.960 | (0.203) |
| Regional Studies | 0.361 | (0.152) | 0.071 | (0.070) | 0.660 | (0.107) | 1.020 | (0.142) |
| Agriculture | 0.171 | (0.197) | 0.223 | (0.106) | 0.613 | (0.130) | 1.121 | (0.168) |
| Psychology | -0.055 | (0.094) | 0.415 | (0.062) | 0.517 | (0.069) | 0.940 | (0.078) |
| English | -0.072 | (0.160) | 0.296 | (0.094) | 0.453 | (0.111) | 1.006 | (0.133) |
| Languages | -0.140 | (0.101) | 0.311 | (0.085) | 0.530 | (0.101) | 0.929 | (0.159) |
| Social Sciences | -0.199 | (0.087) | 0.480 | (0.086) | 0.408 | (0.074) | 1.082 | (0.141) |
| Math | -0.244 | (0.062) | -0.074 | (0.035) | 1.594 | (0.131) | 0.984 | (0.082) |
| Mgmt. & Mkting | -0.308 | (0.125) | 0.166 | (0.093) | 0.440 | (0.124) | 0.985 | (0.247) |
| Biology | -0.439 | (0.078) | 0.166 | (0.049) | 0.648 | (0.074) | 0.827 | (0.084) |
| Engineering | -0.441 | (0.069) | -0.019 | (0.029) | 0.632 | (0.064) | 0.362 | (0.041) |
| Econ., Fin., Acct. | -0.547 | (0.089) | 0.146 | (0.053) | 0.980 | (0.115) | 0.917 | (0.097) |
| Chem. & Physics | -0.708 | (0.081) | 0.042 | (0.044) | 1.286 | (0.073) | 1.166 | (0.070) |

Note: * indicates that the variable is z-scored.

Table B.6: Expected GPA for Average Classes by Department, Unconditional and Conditional on Taking Courses in That Department, with Standard Errors

| | EGPA Females Unconditional | | EGPA Females Conditional | | EGPA Males Unconditional | | EGPA Males Conditional | |
|---|---|---|---|---|---|---|---|---|
| Education & Health | 3.49 | (0.04) | 3.47 | (0.04) | 3.31 | (0.04) | 3.23 | (0.03) |
| Communications | 3.44 | (0.10) | 3.38 | (0.09) | 3.22 | (0.09) | 3.13 | (0.08) |
| Agriculture | 3.34 | (0.04) | 3.21 | (0.03) | 3.20 | (0.04) | 2.86 | (0.03) |
| Mgmt. & Marketing | 3.20 | (0.12) | 3.36 | (0.13) | 3.12 | (0.12) | 3.28 | (0.12) |
| Languages | 3.20 | (0.05) | 3.24 | (0.05) | 3.10 | (0.05) | 3.09 | (0.05) |
| Regional Studies | 3.18 | (0.04) | 3.28 | (0.04) | 2.96 | (0.04) | 3.05 | (0.04) |
| Social Sciences | 3.09 | (0.03) | 3.07 | (0.03) | 2.98 | (0.03) | 2.86 | (0.03) |
| English | 3.08 | (0.04) | 3.09 | (0.04) | 2.93 | (0.04) | 2.97 | (0.04) |
| Psychology | 2.96 | (0.02) | 2.95 | (0.02) | 2.81 | (0.02) | 2.71 | (0.02) |
| Econ., Fin., Acct. | 2.63 | (0.02) | 2.88 | (0.02) | 2.66 | (0.02) | 2.86 | (0.02) |
| Biology | 2.56 | (0.02) | 2.74 | (0.02) | 2.54 | (0.02) | 2.72 | (0.03) |
| Math | 2.53 | (0.02) | 2.57 | (0.02) | 2.54 | (0.02) | 2.65 | (0.02) |
| Engineering | 2.48 | (0.04) | 2.83 | (0.03) | 2.54 | (0.02) | 2.93 | (0.02) |
| Chem. & Physics | 2.30 | (0.02) | 2.54 | (0.02) | 2.39 | (0.02) | 2.68 | (0.01) |

Note: Standard errors generated from bootstrapping with 1,000 draws.

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

In Table B.8 we show the full set of professor parameters (see Table 11 for a subset of the parameters).

Table B.7: 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 | 0.0613 | 0.0415 |
| Management & Marketing | 0.0477 | 0.0278 | 0.023 |
| Regional Studies | 0.0383 | 0.0319 | 0.0279 |
| **Biology** | 0.0656 | 0.054 | 0.0478 |
| **Engineering** | 0.0642 | 0.0545 | 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 | 0.053 |
| Social Sciences | 0.1092 | 0.1277 | 0.136 |
| Communications | 0.1211 | 0.1546 | 0.1611 |
| Education & Health | 0.0889 | 0.0573 | 0.1269 |
| Agriculture | 0.0126 | 0.0763 | 0.0511 |

Note: The shares of each type are 62.6%, 29.7%, and 7.7% 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.

Table B.8: Complete Table – Relating Course Demand to Grades and Workloads

| | Average Grades | | | | $\gamma$ | | | |
|---|---|---|---|---|---|---|---|---|
| | OLS | | IV | | OLS | | IV | |
| | Coef. | Std. Err. | Coef. | Std. Err. | Coef. | Std. Err. | Coef. | Std. Err. |
| Constant | 3.418 | (0.091) | 4.376 | (0.097) | 0.124 | (0.021) | -0.108 | (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.

# C Methods Appendix – Online

This appendix provides additional details regarding our empirical methods:

1. our modified EM algorithm for recovering the parameters of the grade process and conditional probabilities of a student being each unobserved type,
2. 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 $\eta_{ij}$, 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 (25) to capture the likelihood of making observed course choices; however, maximizing Equation (25) 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:

$$v_{ij} = (\phi_1^* + w_i\phi_2^*) g_{ij}(\gamma_j^N, \theta_{j(k)}^N, X_i) + \delta_{0j}^* + w_i\delta_{1j}^* + Z_{1i}\delta_{2k(j)}^* + Z_{2ij}\delta_3^* + \epsilon_{ij}^* \tag{38}$$

where $g_{ij}(\cdot)$ represents the expected grade of student $i$ in course $j$ and $\epsilon_{ij}^*$ is assumed to follow a nested logit structure with nesting at the department level characterized by $\nu$. The full set of choice parameters is then $\varphi = \{\phi^*, \delta^*, \nu\}$. Note that although we will not be interpreting the estimates of $\varphi$, the structure of utility in Equation (38) is very similar to the structure in Equation (9).[40] This

[40] The structure of utility in Equation (38) differs from the structure in Equation (9) in three ways: First, Equation (38) does not subtract $\gamma_j$ from expected grades. Second, Equation (38) assumes nested logit preference shocks, while

ensures that the conditional type probabilities from this specification are appropriate for classifying students for the estimation of Equation (9).

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

$$\sum_i \ln\left(\sum_{s=1}^{S} \pi_s \mathcal{L}_{igs}\left(\theta,\gamma\right)\mathcal{L}_{ics}\left(\varphi\right)\right) \tag{39}$$

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

We iterate on the following steps until convergence, where the $m$th step follows:

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

$$q_{is}^{(m)} = \frac{\pi_s^{(m)}\mathcal{L}_{igs}\left(\theta^{(m-1)},\gamma^{(m-1)}\right)\mathcal{L}_{ics}\left(\varphi^{(m-1)}\right)}{\sum_{s'}\pi_{s'}^{(m)}\mathcal{L}_{igs'}\left(\theta^{(m-1)},\gamma^{(m-1)}\right)\mathcal{L}_{ics'}\left(\varphi^{(m-1)}\right)} \tag{40}$$

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:

$$\sum_i\sum_s q_{is}^{(m)}\left(\ln\left[\mathcal{L}_{igs}\left(\theta,\gamma\right)\right]+\ln\left[\mathcal{L}_{ics}\left(\varphi\right)\right]\right) \tag{41}$$

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 $\gamma_j^N$ and $\theta_{k(j)}^N$. Denote as $\varphi(!A)$ $\varphi$ absent the $A$th 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 $\varphi$ as given, choose $\gamma_j^N$ and $\theta_k^N$ to maximize:

$$\sum_i\sum_{j\in k} d_{ijk}\left(\ln[f(g_{ij},\gamma_j^N,\theta_k^N)]+\ln[p_{ij|k}(\gamma_j^N,\theta_k^N,\varphi)]\right) \tag{42}$$

Equation (9) assumes independent Type 1 extreme value errors. Finally, Equation (38) assumes that contemporaneous choices are independent, while Equation (9) models students choosing bundles of courses simultaneously.

2. Taking $\gamma_j^N$, $\theta_k^N$, and $\varphi(!\phi^*, !\delta_3^*)$ as given, choose $\phi^*$ and $\delta_3^*$ to maximize:

$$\sum_i \sum_j d_{ij} \ln \left[ p_{ijk} \left( \gamma_j^N, \theta_k^N, \varphi(!\phi^*, !\delta_3^*), \phi^*, \delta_3^* \right) \right] \tag{43}$$

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

$$\sum_i \sum_{j \in k} d_{ijk} \ln \left[ p_{ij|k} \left( \gamma_j^N, \theta_k^N, \varphi(!\delta_{0j}^*), \delta_{0j}^* \right) \right] \tag{44}$$

4. Taking $\gamma_j^N$, $\theta_k^N$, and $\varphi(!\delta_1^*, !\delta_2^*, !\nu)$ as given, choose $\delta_{1j(k)}^*$, $\delta_{2j(k)}^*$, and $\nu$ to maximize:[41]

$$\sum_i \sum_k d_{ik} \ln \left[ p_{ik} \left( \gamma_j^N, \theta_k^N, \varphi(!\delta_1^*, !\delta_2^*, !\nu), \delta_1^*, \delta_2^*, \nu \right) \right] \tag{45}$$

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 (21). Finally, we use the estimates of $q_{is}$ as weights in estimating the structural choice parameters using (25).

### 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 $\delta_{0j}$, 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 $\delta_{0j}$ and other choice parameters. Given a new guess of $\tilde{\Theta}$, we use the $\delta_{0j}$s from the previous guess $\delta_{0j}^0(\tilde{\Theta})$ and calculate $S_j \left( \delta_{0j}^0, \tilde{\Theta} \right)$. The $m$th iteration

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

of the fixed-point problem updates $\delta_{0j}^{m}$ using:

$$\delta_{0j}^{m} = \delta_{0j}^{m-1} + \ln\left[S_j^d\right] - \ln\left[S_j\left(\delta^{m-1}, \tilde{\Theta}\right)\right] \tag{46}$$

Given the $\delta_{0j}^{m}$, 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.1). 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.

# 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
5. show that any bias in our professor effort measure from reverse causality is small.

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, $\delta_{0j}$, into intrinsic demand, $\delta_{0j}^*$, and the effort of the professor, $\tau_j$:

$$\delta_{0j} = \rho\tau_j + \delta_{0j}^* \tag{47}$$

where $\rho$ 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 $\rho$. As a result, we estimate the model under different assumed values of $\rho$.

## 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

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 | 0.1590 | 1.0000 | | | |
| Q13 | 0.1997 | 0.7374 | 1.0000 | | |
| Q19 | 0.1805 | 0.5983 | 0.7317 | 1.0000 | |
| (Q09+Q13+Q19)/3 | 0.2025 | 0.8633 | 0.9256 | 0.8815 | 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?

professor's effort, $\tau$. Of the twenty questions in the evaluations, we focus on three with students answering on a five-point Likert scale:

- 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.[42]

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 multiple semesters of the evaluation data. To do so, we 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

[42]See Insler et al. (2021), Nelson & 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 | 1.1773 | (0.0559) |
| B | 0.9736 | (0.0559) |
| C | 0.7606 | (0.0567) |
| D | 0.4305 | (0.0485) |
| log(class size) | -0.0901 | (0.008) |

Note: The dependent variable in the top half is the average response to three questions on a 5-point Likert scale in the evaluations: Did the instructor (1) present the material effectively, (2) stimulate interest in the subject, and (3) stimulate you to read further beyond the class? 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 first half of the table. Regressors include professor and semester fixed effects.

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.[43] We then standardize this variable to have mean zero and standard deviation one. It is this standardized variable that we use for $\tau_j$.

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, $\tau_j$, in addition to grading policy parameters $\beta_j$ and $\gamma_j$. 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:

$$V_j(\beta, \gamma, \tau) = -\left(\ln[E_j(\beta, \gamma, \tau)] - e_{0j}\right)^2 - \lambda_1\left(\overline{G}_j(\beta, \gamma, \tau) - e_{1j}\right)^2 - \lambda_2(\gamma_j - e_{2j})^2 - \lambda_3(\tau_j - e_{3j})^2 \quad (48)$$

[43] There is potential for bias in this estimated effect of enrollment due to reverse causality. We explore this in detail in Appendix D.4. To summarize, we find that any potential bias would be economically insignificant.

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:

$$0 = -\left(\ln\left[E_j(\beta,\gamma,\tau)\right] - e_{0j}\right)\frac{\partial \ln E_j}{\partial \tau_j} - \lambda_1\left(\overline{G}_j(\beta,\gamma,\tau) - e_{1j}\right)\frac{\partial \overline{G}_j}{\partial \tau_j} - \lambda_3(\tau_j - e_{3j}) \tag{49}$$

In recovering $\Psi_0$, $\Psi_2$, and $\lambda_2$, we create our instruments with $\beta^0$, $\gamma^0$, and $\tau^0$.

To estimate $\lambda_3$ and $\Psi_3$, we take $\Psi_0$ as given and eliminate $\lambda_1$ using using Equations (31) and (49) to solve for $\tau_j$:

$$\tau_j = (1/\lambda_3)C_j\left(\ln\left[E_j\right] - \Psi_0\right) + \Psi_3 + \varepsilon_{3j} \tag{50}$$

where $C_j$ is given by:

$$C_j = \left[\frac{\partial \ln\left[E_j\right]}{\partial \beta_j}\frac{\partial \overline{G}_j}{\partial \tau_j}\bigg/\frac{\partial \overline{G}_j}{\partial \beta_j}\right] - \frac{\partial \ln\left[E_j\right]}{\partial \tau_j} \tag{51}$$

We then instrument for $C_j\left(\ln\left[E_j\right] - \Psi_0\right)$ by evaluating $C_j$ and $\ln\left[E_j\right]$ at the common grading and effort policies, $\beta^0$, $\gamma^0$, and $\tau^0$. Recovery of $\Psi_1$ and $\lambda_1$ proceeds as before.

## D.3 Professor preference estimates and equilibrium counterfactuals under different values of $\rho$

Our model in the paper abstracted from professor effort. This is equivalent to assuming that $\rho = 0$ in our extended model. We also estimate the professor preference parameters under the assumption that $\rho = 0.05$ and $\rho = 0.2$. Table D.3 shows the estimates of professor preferences (shown in Table 12) at alternate values of $\rho$. Table D.4 shows the general equilibrium counterfactual results (shown in Table 15) at alternative values of $\rho$.

## D.4 Size of bias in our professor effort measure

Our measure of professor effort is biased due to reverse causality: professor effort affects enrollment. In this section, we illustrate why this is bias is likely small. We begin by approximating log enrollment for instructor $i$ in course $j$ at time $t$ as:

$$\ln(E_{ijt}) = \delta_{1i} + \delta_{2j} + \delta_{3t} + \rho\tau_{ijt} + \epsilon_{1ijt} \tag{52}$$

In this case, $\rho$ gives the return to effort, $\tau_{ijt}$.

Now express effort, $\tau_{ijt}$ as depending on innate demand—everything in the equation above

Table D.3: Estimates of Professor Preferences at Alternative $\rho$ Values

| | | $\rho = 0.2$ | | $\rho = 0.05$ | | $\rho = 0.2$ | | $\rho = 0.05$ | |
|---|---|---|---|---|---|---|---|---|---|
| | | Ideal grade | | | | Ideal workload | | | |
| | $\lambda$ | 2.933 | (0.448) | 2.899 | (0.434) | 48.160 | (5.024) | 46.868 | (4.674) |
| | Constant | 2.627 | (0.120) | 2.619 | (0.119) | 0.275 | (0.038) | 0.277 | (0.038) |
| Class | Upper-Level Class | 0.435 | (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) |
| Rank | 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) |
| Gender | 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) |
| Dept. | 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) | 0.003 | (0.019) |
| | Education & Hleath | 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) | 0.204 | (0.024) |
| | Language | -0.057 | (0.066) | -0.057 | (0.066) | 0.041 | (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) | 0.021 | (0.129) | 0.035 | (0.031) | 0.032 | (0.030) |
| | **Math** | -0.331 | (0.081) | -0.328 | (0.081) | 0.094 | (0.021) | 0.092 | (0.021) |
| | **Chem. & Physics** | -0.160 | (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 | | | |
| | $\lambda$ | 1.000 | | 1.000 | | 0.528 | (0.073) | 0.251 | (0.064) |
| | Constant | 5.196 | (0.662) | 5.172 | (0.639) | -0.348 | (0.061) | -0.179 | (0.059) |
| Class | Upper-Level Class | -1.384 | (0.528) | -1.372 | (0.511) | | | | |

Note: Ideal enrollment $\lambda$ 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 $\rho$ Values

| | | Class Size | | STEM Enrollment Share | | |
|---|---|---|---|---|---|---|
| | $\rho$ | STEM | Non-STEM | Overall | Female | Male |
| Baseline | | 82.6 | 45.0 | 41.8% | 34.6% | 49.5% |
| Equal Demand | 0 | 44.6 | 59.8 | 22.6% | 17.7% | 27.8% |
| | 0.05 | 45.2 | 59.6 | 22.9% | 18.1% | 28.0% |
| | 0.20 | 45.5 | 59.4 | 23.1% | 18.3% | 28.2% |
| Equal Prof Pref | 0 | 93.2 | 40.8 | 47.2% | 40.2% | 54.6% |
| | 0.05 | 94.6 | 40.2 | 47.9% | 41.1% | 55.2% |
| | 0.2 | 93.7 | 40.6 | 47.4% | 40.7% | 54.7% |
| Equal Prof Grade Pref | 0 | 89.1 | 42.4 | 45.1% | 38.4% | 52.3% |
| | 0.05 | 90.1 | 42.0 | 45.6% | 39.0% | 52.7% |
| | 0.2 | 89.2 | 42.4 | 45.2% | 38.6% | 52.2% |
| Equal Prof $\gamma$ Pref | 0 | 86.5 | 43.4 | 43.8% | 36.2% | 52.0% |
| | 0.05 | 87.4 | 43.1 | 44.2% | 36.7% | 52.4% |
| | 0.2 | 86.7 | 43.4 | 43.9% | 36.4% | 51.9% |
| 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 | 98.5 | 38.7 | 49.9% | 43.9% | 56.3% |

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.

besides the $\tau_{ijt}$ term—plus some instructor and time fixed effects and an error:

$$\tau_{ijt} = \alpha\left(\delta_{1i} + \delta_{2j} + \delta_{3t} + \epsilon_{1ijt}\right) + \delta_{4j} + \delta_{5t} + \epsilon_{2ijt} \tag{53}$$

$\alpha$ then gives how innate demand affects effort. Note that our assumption is that $\alpha < 0$, as higher innate demand implies less effort.

We can then express equation (53) with respect to $\ln(E_{ijt})$:

$$\tau_{ijt} = \alpha \ln(E_{ijt}) - \alpha\rho\tau_{ijt} + \delta_{4j} + \delta_{5t} + \epsilon_{2ijt} \tag{54}$$

Rearranging the terms shows the bias that would arise in our estimate of $\alpha$ due to the dependence of log enrollment on effort:

$$\tau_{ijt} = \frac{\alpha \ln(E_{ijt})}{1+\alpha\rho} + \delta^*_{4j} + \delta^*_{5t} + \epsilon^*_{2ijt} \tag{55}$$

where the starred variables are just scaled up by $\frac{1}{1+\alpha\rho}$.

Equation (55) is what we use to estimate professor effort. The coefficient on log enrollment is less than zero. This tells us two things under the assumption that $\rho$ is greater than zero, that is, that professor effort positively affects enrollment. First, $\alpha$ is indeed negative. Second, $1 > (1+\alpha\rho) > 0$.

Thus, the magnitude of our coefficient on effort is biased upward by the factor $\frac{1}{1+\alpha\rho}$. The closer this number is to one, the smaller the bias.

Our estimated coefficient on log enrollment is -0.09 (see Table D.2). Note that -0.09 is actually biased upward in magnitude. Hence, the actual bias factor is between 1 and $\frac{1}{1-0.09\rho}$. If $\rho$ is 0.05, the bias factor is then less than 1.0045 (where 1 would be no bias). When $\rho$ is 0.2, the bias factor is less than 1.018. In this latter case, the true $\alpha$ would be between -0.088 and -0.09, suggesting that any bias in our effort measure is small.