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

Thomas Ahn	Peter Arcidiacono	Amy Hopson	James Thomas
University of Kentucky	Duke University	Duke University	Duke University
	NBER & IZA		

October 21, 2016

Abstract

We estimate an equilibrium model of grading policies where professors set both an intercept and a returns to studying and ability. Professors value enrollment, learning, and student study time and set their policies taking into the account the policies of the other professors. Students respond to grading policies in their selection of courses and how much to study conditional on enrolling. Men and women are allowed to have different preferences over course types, the benefits associated with higher grades, and the cost of exerting more effort. Two decompositions are performed. First, we separate out how much of the differences in grading policies across fields is driven by differences in demand for courses in those fields and how much is due to differences in professor preferences across fields. Second, we separate out differences in female/male course taking across fields is driven by i) differences in cognitive skills, ii) differences in the valuation of grades, iii) differences in the cost of studying, and iv) differences in field preferences. We then use the structural parameters to evaluate restrictions on grading policies. Restrictions on grading policies that equalize grade distributions across classes result in higher (lower) grades in science (nonscience) fields but more (less) work being required. As women are willing to study more than men, this restriction on grading policies results in more women pursuing the sciences and more men pursuing the nonsciences.

*Preliminary and incomplete. Please do not quote without permission.

1 Introduction

Even after accounting for selection, substantial earnings differences exist across majors. Majors in engineering and the sciences, as well as economics and business, pay substantially more than other fields.[1] Further, earnings disparities across majors have increased substantially over time (Altonji et al. (2014) and Gemici & Wiswall (2014)). Despite their value in the marketplace, STEM (Science, Technology, Engineering, and Mathematics) fields are perceived to be undersubscribed. A report by the President's Council of Advisors on Science and Technology (2012) suggests substantial needs to increase the number of STEM majors. Florida has proposed freezing tuition for STEM majors (Alvarez (2012)), and the state of New York is offering free tuition for high performing students who enroll in public institutions as STEM majors, conditional on working in the state for at least five years (Chapman (2014)).

But many more students enroll in college expecting to major in a STEM field than actually finish in a STEM field (Arcidiacono (2004), Arcidiacono et al. (forthcoming), Stinebrickner & Stinebrickner (2014)). This is not just due to students dropping out: many students switch from STEM to non-STEM fields, particularly in comparison to those who switch from non-STEM to STEM fields. Further, it is predictable who will switch. Those who have relatively weak academic preparation (e.g. SAT scores or HS grades) are much more likely to leave STEM fields. While relatively high levels of academic preparation are associated with persisting in STEM majors, there is little evidence that high levels of academic preparation are more rewarded in the labor market for STEM majors than for non-STEM majors. Women are also more likely to switch: Arcidiacono et al. (2012) show with data from Duke University, that differences in academic preparation can account for the large differences in switching behavior across races but is unable to explain the substantial gender gap.

A potential channel for influencing the number and composition of STEM majors are grading policies. Should grading policies prove to be an important predictor of major choice, they may serve as a relatively cheap way of increasing STEM majors. While other means such as increasing pre-college academic preparation or the share of underrepresented groups in STEM fields may also be effective, these methods are also very costly with the benefits coming much later. Shifting the way teachers teach the sciences and introducing more laboratory-based curricula are both expensive.

[1] See Altonji et al. (2012) and Altonji et al. (forthcoming) for reviews.

Altering training, hiring, and promotion in academia, government agencies, and firms is also costly, as are adjusting long-standing cultural attitudes in the home, school, and workplace.

There is evidence that grades affect sorting into majors. The same majors that pay well also give (on average) significantly lower grades (Sabot & Wakeman-Linn (1991), Johnson (2003)) and are associated with more study time (Brint et al. (2012), Stinebrickner & Stinebrickner (2014)). Lower grades and higher study times deter enrollment. Sabot & Wakeman-Linn (1991) show that the absolute level of grades was a far more important indicator of taking further courses in the subject than their ranking within the class. Butcher et al. (2014) showed that Wellesley's policy of capping the fraction of A's given resulted in shifts towards science classes and science majors. There is also evidence that students enter unaware of the extent of cross-department differences in grading standards. Stinebrickner & Stinebrickner (2014) show that the over-optimism regarding performance at Berea College is primarily driven by students over-predicting their performance in the sciences. As students take more classes, students generally revise their expected performance in the sciences downward. This holds true even for students who persist in the sciences who ought to have received relatively positive grade realizations.

With students responding to grading practices through their choice of courses, departments may set their grading policies in order to deter or encourage enrollment. Those with low enrollments may find it difficult to increase or maintain their faculty size. Hence, incentives exist to raise grades in order to encourage enrollment in these departments. On the other hand, departments that are flush with students may have incentives to lower grades to keep their enrollments to a more manageable size. Within any given department, individual professors may also seek to influence enrollment up or down for his or her class to minimize teaching effort or maximize student learning.[2]

Differences in grading policies may have differing effects for males and females. In principle female students should be particularly interested in STEM fields. Women report studying substantially more than men (Stinebrickner & Stinebrickner (2014), Arcidiacono et al. (2012)), and they should be undeterred by the higher study requirements of these classes. Yet, females are substantially less likely to graduate with a STEM major than males.[3] Ideas for why this might happen

[2]This issue is becoming even more salient as more universities move toward a fiscal model where departmental budgets are more directly determined by enrollment size or credits generated.

[3]The gender gap is not uniform across STEM fields. Indeed, in some STEM fields, such as biosciences, women receiving BA's actually outnumber men.

have been numerous, including role model effects (Rask & Bailey (2002), Hoffmann & Oreopoulos (2009), Carrell et al. (2010)) and future labor market considerations (Gemici & Wiswall (2014), Bronson (2014)) among many others. In addition to these channels, women may study more in part because they value the benefits of studying–higher grades–more than their male counterparts (Rask & Bailey (2002), Rask & Tiefenthaler (2008)). Good grades may yield direct psychic benefit, or they may impact time to graduation or ability to qualify for grants and scholarships. For example, if female students are more risk averse or pessimistic about attrition probability compared to their male counterparts, grades may hold more value. Again, the advantage to focusing on grading policies is that it may be relatively cheap to do so compared to alternative programs.

We propose to estimate an equilibrium model of student course enrollment and effort decisions as well as professor decisions regarding grading standards. How professors set grades affects enrollment and how much students study, though differentially for men and women. The professor objective function includes enrollments, so part of how professors set grades is determined by course demand. With the estimates of the equilibrium model, we will be able to evaluate how differences in grading practices across fields affect, partly as a result of demand, the share of courses taken in different fields. Further, we can see whether cross-departmental differences in professor preferences over enrollment either exacerbate or mitigate the differences in grading across fields.

2 Data

Estimating such a model requires rich data on student course taking, study hours, and grades. We use a detailed student enrollment data set from the University of Kentucky (UK). UK, the state's flagship public post-secondary institution, has a current undergraduate enrollment of approximately 21,000. The school was ranked 119 out of approximately 200 'National Universities' by U.S. News & World Report (U.S. News & World Report 2013). This places UK in the middle of the distribution of large post-secondary institutions, and the student body serves as a good cross-section of college students nationwide.

The data set contains student demographic and course enrollment information. Each semester, the entire student body's course selections and grades are recorded by the Registrar's Office. This data set is particularly valuable because every student outcome in every class is captured, allowing us to estimate a rich model of student and professor interactions. Furthermore, we can analyze course selection and performance explicitly modeling the student choosing his or her semester course-load (which may be contain one to five courses) from the entire course catalog. We also collect information on course pre-requisites. Restricting the student's choice set using this data aids computation as well as creating a more true-to-life representation of the decisions a student faces when selecting his or her courses. For this study, we focus on student enrollment observations from one semester, Fall 2012.[4]

In addition, we have access to class evaluation surveys completed by students at the end of the semester. We note that coverage is not complete, as some departments chose not to make evaluation data available. Linking the class evaluation data to the enrollment data is complex, as rules for identifying the course (or sections within the course) and instructor (or sub-instructor - frequently a graduate student - teaching under the supervision of a head-instructor) are defined independently by the department. As students do not identify themselves in the evaluation forms, we aggregate the data up to the class level. We are able to match 76 percent of classes successfully. We then restrict the data to classes with at least a 70 percent response rate and drop classes with small numbers of respondents to prevent possible identification. Most critically for our research, students are asked about the number of hours per week they spend on studying for this particular course. This information makes the identification of structural parameters on study effort possible, allowing us to analyze how differences in average study time relates to course and student characteristics across courses.[5]

Our Fall 2012 sample yields 89,582 student/class observations. There are 19,527 unique undergraduates, implying that on average, each student enrolls in (but not necessarily completes) four to five courses.[6] Table 1 provides demographic summary statistics, separated by gender. Overall,

[4]We have the capability to link student data across multiple semesters, spanning Fall 2008 to Spring 2013. This yields approximately 1.4 million student/class observations. We restrict ourselves to one semester due to computational constraints.

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

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

women and men look similar when entering college. Women have slightly higher high school grades and slightly lower standardized ACT scores.[7] Women also have higher grades while in college. Sharp differences show up in major selection. While women comprise a slight majority at UK overall, the ratio between men and women in STEM majors is approximately 1.6. In contrast to students from more selective institutions (seen in many other studies of higher education outcomes), over 30 percent of students at UK are part-time students, taking less than 12 credits during the semester.

Table 2 summarizes class-level characteristics separated by STEM-status of the course. STEM classes are substantially larger and give significantly lower grades compared to non-STEM courses. As implied by Table 1, female students are the minority in STEM classes. This is despite the fact that they perform better, on average, than their male counterparts in these courses. On average, each STEM course requires one more hour of study time per week (or 30 percent more time) than a non-STEM course. The study time difference actually understates the true gap across STEM and non-STEM courses. On average, students with higher academic ability will select more often into STEM courses, so each hour spent studying should yield more learning. Yet, STEM classes average much lower grades. A student attempting to generate an equivalent grade across a STEM and non-STEM course will have to invest significantly more than an extra 30 percent in study time in the STEM course.

Table 3 presents simple OLS results showing the relationship between individual and class characteristics with grades and study hours after controlling for a large number of academic background measures.[8] The grades regression sample is at the student/class level, and the study hours per week regression sample is at the class level. The first column gives the results for grades. The patterns are consistent with those in Table 2, STEM classes give lower grades and females have higher grades. Classes that have a higher fraction of female students also give higher grades. This is consistent with there not being a grade curve that is common across STEM or non-STEM departments else

[7]SAT scores are converted to equivalent ACT scores, and the math and verbal sections are averaged.

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

Table 1: Descriptive Statistics by Gender

	Men	Women
High school GPA	3.13	3.34
	(1.20)	(1.16)
ACT Score	25.2	24.4
	(4.42)	(4.18)
Fall 2012 GPA	3.02	3.24
	(0.713)	(0.665)
Fall 2012 Credits	11.7	12.0
	(4.29)	(4.22)
STEM Major	38.0%	23.8%

Note: Fall 2012 University of Kentucky undergraduate students, 9,729 men, 9,798 women. Standard deviations in parentheses.

the higher grades females receive would translate into lower grades for everyone else. Class size has a negative effect on grades. The coefficient on class size confounds two effects that work in opposite directions. On the one hand, students prefer higher grades so higher enrollments should be associated with higher grades. On the other, courses that have high intrinsic demand may have lower grades since these courses do not need to have high grades to attract students.

The second column on Table 3 shows regressions of study hours on the average characteristics of the class. STEM classes are associated with an extra half hour of study, slightly less than what is seen in the descriptive statistics. This suggests that STEM classes are attracting students who are willing to study more, with the grading policies in the STEM classes further spurring on these students to commit more time to study. Classes that have more women also study more, consistent with the previous literature (DiPrete and Buchmann 2013). But perhaps the most interesting coefficient is that on average grades. Courses that give higher grades have less study time, suggesting grades should be interpreted as relative, not absolute, measures of accomplishment, as well as suggesting grade inflation may have negative consequences for learning.

Table 2: Descriptive Statistics by Course Type

	STEM	Non-STEM
Class Size	78.1	46.3
	(101.1)	(64.0)
Average Grade	3.03	3.31
	(0.50)	(0.46)
Average Grade | Female	3.11	3.40
	(0.59)	(0.46)
Study Hours	3.61	2.70
	(1.68)	(1.12)
Percent Female	37.0%	55.9%

Note: Fall 2012 University of Kentucky courses with enrollments of 16 or more students, 379 STEM courses, 1,164 non-STEM courses. For study hours, 293 STEM courses and 793 non-STEM courses. Standard deviations in parentheses.

3 Model

Individual i chooses n courses from the set $[1, \ldots, J]$. Let $d_{ij} = 1$ if j is one of the n courses chosen by student i and zero otherwise. The payoff associated with a bundle of courses is given by the sum of the payoffs for each of the individual courses where the payoffs do not depend on the other courses in the bundle. We specify the payoff for a particular course j as depending on student i's preference for the course, δ_{ij}, the amount of study effort the individual chooses to exert in the course, s_{ij}, and the expected grade conditional on study effort, $\mathbb{E}\left[g_{ij} \,|\, s_{ij}\right]$:

$$U_{ij} = \phi_i \mathbb{E}\left[g_{ij} \,|\, s_{ij}\right] - \psi_i s_{ij} + \delta_{ij} \tag{1}$$

Students then solve the following maximization problem when choosing their optimal course bundle:

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

The grade student i receives in course j, g_{ij}, depends on the academic preparation of student

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

Dependent Var.	Grade	Study hours per week
STEM Class	-0.325	0.520
	(0.009)	(0.148)
Female	0.140	
	(0.008)	
Percent Female	0.395	0.547
	(0.203)	(0.191)
Average Grade		-0.635
		(0.089)
ln(Class Size)	-0.116	-0.396
	(0.004)	(0.048)
Observations	72,449	1,085

Note: Additional controls for grades regression include, minority status, freshman, STEM major, pell grant, in-state student, ACT score, HS gpa, percent minority, percent freshman. Additional controls in study hours regression include percent freshmen, percent STEM major, percent pell grant, percent in-state, average ACT score, average HS gpa, percent minority.

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

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

Grading policies by the professors are then choices over an intercept, β_j, and a return to academic preparation and effort, γ_j.[9] Gains from study effort enters in as a log to capture the diminishing returns to studying. Along with the linear study effort cost defined in the utility function, this ensures an interior solution for the optimal amount of study time.

Students are assumed to know the professors' grading policies.[10] Substituting in for expected grades in (1) yields:

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

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

$$\begin{aligned} 0 &= \frac{\phi_i \gamma_j}{s_{ij}} - \psi_i \\ s^{\star}_{ij} &= \frac{\phi_i \gamma_j}{\psi_i} \end{aligned} \tag{5}$$

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

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

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

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

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

[9] For example, if there is a university-wide (or department-level) mandated/recommended grade distribution, we will be able to capture such a policy, as β_j and γ_j will have lower variance.

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

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

The key equations for estimation are then given by:

(i) the solution to the students maximization problem where (6) is substituted into (2),

(ii) the grade production process given in (7), and

(iii) the optimal study effort given in (5).

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

4 Estimation

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

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

In the second stage, we relate the optimal study effort given in Eq. (5) to the student evaluation data. The evaluations are collected for each class, and students report how many hours they spent studying in that class. This helps us recover some of the study effort parameters, as well as unravel some of the normalizations on the γs required in the first stage. We are able to relate all γs across departments, but still cannot only identify the γ's relative to one course.

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

4.1 Parameterizations

To estimate the model, we need to place some structure on course preferences, δ_{ij}, the value of grades, ψ_i, and the cost of effort, ϕ_i. Further, we must relate academic preparation, A_{ij}, to what we see in the data. Denote $w_i = 1$ if individual i is female and zero otherwise. Denote X_i as a row vector of explanatory variables such as ACT scores, high school grades, race, etc.[11] Denote Z_i a a row vector of explanatory variables that affect preferences for particular departments or levels of courses within departments. Hence Z_i includes gender as well as year in school, allowing women to have a preference for classes in particular departments and the attraction of upper-division versus lower-division classes to vary by department and year in school. Preference shocks for courses are represented by ϵ_{ij}. Finally, we partition courses into K departments, $K < J$, where $k(j)$ gives the department for the jth course. We then parameterize the model as follows:

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

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

$$\psi_i = \exp\left(\psi_0 + w_i\psi_1 + X_i\psi_2\right) \tag{10}$$

$$\phi_i = \phi_0 + w_i\phi_1 \tag{11}$$

There is no intercept in A_{ij} as it can not be identified separately from the β_j's. Note that the same variables enter into academic preparation, preferences, and effort costs, only with different coefficients. Preferences for courses allow for both course fixed effects as well as students with particular characteristics preferring courses in particular departments, $\delta_{1k(j)}$. Note also that the effort costs are exponential in the explanatory variables. This ensures that effort costs are positive. Finally, preferences for grades are only allowed to vary by gender. In principle, we could allow them to vary with X_i as well, but this would substantially complicate the model.

Having separate estimates by gender across all the relevant parameters will help uncover some of the driving forces behind the gender gap in STEM. For example, if female intrinsic demand for courses in STEM departments is relatively low ($\delta_{1k(j)}$ negative) while preferences for grades and cost of effort are relatively equal across males and females (ψ_1 and ϕ_1 close to zero), then changing grading policies will have no effect on the gender gap in STEM. In this case, it would require figuring out why females are not interested in STEM fields, and policies would have to be geared more

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

towards early education about opportunities in STEM for females, or changing cultural attitudes towards females in STEM. On the other hand, if females have significantly different preferences over grades and study effort than males, then altering grading policies could affect the gender distributions within classes and departments. For example, if females have higher preferences for grades (ϕ_1 positive) and lower cost of effort (ψ_1 negative) than males, then increasing γ_j and correspondingly changing β to keep enrollments in STEM courses the same would result in an increase in the fraction of females in STEM.

4.2 Estimation without Unobserved Heterogeneity

4.2.1 Grade parameters

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

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

where:

$$\begin{aligned}
\theta_{0j} &= \beta_j + \gamma_j (\ln(\phi_0) + \ln(\gamma_j) - \psi_0) && (13)\\
\theta_{1k(j)} &= \alpha_{1k(j)} + \ln(\phi_0 + \phi_1) - \ln(\phi_0) - \psi_1 && (14)\\
\theta_{2k(j)} &= \alpha_{2k(j)} - \psi_2 && (15)
\end{aligned}$$

We estimate the reduced form parameters $\{\theta_{0j}, \theta_1, \theta_2\}$ as well as the structural slopes, the γ_j's, using nonlinear least squares. A normalization must be made for every department as scaling up the θ's by some factor and scaling down the γ's by the same factor would be observationally equivalent. We set one γ_j equal to one for each department.[12] Denote C_k as the normalization for department k. We then estimate γ_j^N where $\gamma_N = \gamma_j / C_{k(j)}$. Similarly, we estimate $\theta_{1k(j)}^N$ and $\theta_{2k(j)}^N$ where $\theta_{1k(j)}^N = \theta_{1k(j)} C_{k(j)}$ and $\theta_{2k(j)}^N = \theta_{2k(j)} C_{k(j)}$.

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

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

4.2.2 Study parameters

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

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

$$\ln(h_{ij}) = \mu \ln(s^*_{ij}) + \zeta_{ij} \tag{16}$$

Substituting in for s^*_{ij} yields:

$$\begin{aligned} \ln(h_{ij}) &= \mu\left(\ln(\mu_1) + \ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i)\right) + \zeta_{ij} && (17) \\ &= \kappa_0 + w_i\kappa_1 - X_i\kappa_2 + \ln(\gamma_j) + \zeta_{ij} && (18) \end{aligned}$$

where:

$$\begin{aligned} \kappa_0 &= \mu(\ln(\phi_0) - \psi_0) && (19) \\ \kappa_1 &= \mu(\ln(\phi_0 + \phi_1) - \ln(\phi_0) - \psi_1) && (20) \\ \kappa_2 &= \mu\psi_2 && (21) \end{aligned}$$

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

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

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

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

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

$$\frac{\sum_i (l_i = l) d_{ij} \ln(h_{ij})}{\sum_i (l_i = l) d_{ij}} = \tilde{\kappa}_0 + w_{jl}\kappa_1 - X_{jl}\psi_2 + \kappa_{3k(j)} + \mu_2 \ln(\hat{\gamma}^N) + \zeta_{jl} \tag{23}$$

where w_{jl} and X_{jl} are the averages of these characteristics for those of year level l enrolled in course j. We correct for potential bias due to measurement error in γ_j using instrumental variables. We use the share of freshmen, sophomores, and juniors in each class as instruments for $\log(\gamma_j)$, and estimate using two stage least squares.

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

4.2.3 Utility parameters

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

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

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

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

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

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

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

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

More subtle is the identification on the returns to the normalized course. If separate course fixed effects were estimated for both men and women then C_1 would not be identified as it would be subsumed into the course fixed effects. But by allowing females preferences to vary at the department rather than the course level,[13] the extent to which sorting happens beyond the effect through grades themselves identifies C_1.

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

Simulated maximum likelihood

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

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

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

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

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

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

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

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

Recovering the remaining structural parameters

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

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

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

The remaining structural parameters of the grade equation, equation (12), are the course intercepts, β_j, and the returns to observed abilities $\alpha_{1k(j)}$ and $\alpha_{2k(j)}$. These can be recovered using:

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

4.3 Estimation with Unobserved Heterogeneity

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

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

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

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

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

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

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

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

2. Update $\pi_s^{(m)}$ using $\left(\sum_{i=1}^{N} q_{is}^{(m)}\right)/N$.

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

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

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

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

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

4.4 Implications from the Demand-Side Estimation

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

(i) differences in preferences (δ_{ij}),

(ii) differences in value of grades (ϕ_i),

(iii) differences in study costs (ψ_{ij}).

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

5 Results

5.1 Preference estimates

Table 4 presents the preference parameters with the exception of the study costs, the class-specific intercepts, and the coefficients on year in school cross department cross level of the course. Recall that the parameter on expected grades is identified from variation in how abilities are rewarded in different classes. Both men and women value grades, with the estimates suggesting that women value grades around 18% more than men. The estimate of female preferences for female professors is positive, with the estimate suggesting that women be indifferent between a class that had a female professor and one that had a male professor who gave grades that were a little under 0.3 points higher. This coefficient is likely biased upward due to the aggregation of departments. To the extent that female professors are more likely to be in departments that females have a preference for and there is variation within our aggregated groups, we may be picking up within-group preferences for departments.

The second set of rows of Table 4 shows female preferences (relative to male preferences) for different departments. The omitted category is Agriculture. The largest difference in preferences is between Engineering and English: 1.48 points, which translates into 3 grade points. Engineering, however, is an outlier with all the other gaps smaller that one point (or 2 grade points).

5.2 Study effort estimates

Estimates of the study effort parameters are presented in Table 5.[14] Lower study costs result in women studying a little over 7% more than men conditional on taking the same class. However, women also study more because they value grades more, with this effect at over 18%,[15] again conditional on taking the same class resulting in an overall effect of over 25%.

All of the measures of academic preparation (ACT scores and high school grades) have the expected sign but are small in magnitude. While blacks and first generation students have higher

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

[15]This number comes from the difference in the log of the preferences for grades: $\ln(\phi_0+\phi_1)-\ln(\phi_0) = \ln(.4916) - \ln(.4157)$.

Table 4: Estimates of Preference Parameters

Preference for:	Coeff.	Std. Error
Expected grades (ϕ)	0.4157	(0.0199)
Female $\times$ expected grades	0.0759	(0.0157)
Female $\times$ female professor	0.1455	(0.0186)
C_1 (normalizing constant)	0.9581	(0.2593)
Female preferences for Departments		
Engineering	-1.0615	(0.0737)
Econ., Fin., Acct.	-0.5091	(0.0593)
Social Sciences	-0.2862	(0.0545)
Communication	-0.1528	(0.0537)
Chemistry & Physics	-0.1482	(0.0599)
Languages	-0.1033	(0.0582)
Mathematics	-0.0072	(0.0688)
Mgmt. & Mkting.	0.1153	(0.0662)
Regional Studies	0.2216	(0.0698)
Biology	0.2546	(0.0638)
Education & Health	0.3287	(0.0581)
Psychology	0.3758	(0.0659)
English	0.4167	(0.0769)

study costs, Hispanics and miscellaneous minorities have lower study costs. Those who are the second unobserved type have higher study costs but, as we will see in subsequent tables, are more able.[16] Finally,

The second set of columns shows how the returns to study effort vary across classes, taking the median γ class for each course grouping. The heterogeneity is quite large. A ten percent increase in study effort would translate into over a third of a grade point increase in mathematics but would translate into less than a tenth of a grade point in agriculture, management and marketing, and

[16] The population probability of being the second unobserved type is 0.213. The information on grades and course-taking does a good job of sorting individuals into types. See Appendix Figure X for a histogram of the conditional type probabilities.

education.

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

	Study Effort Coeff. ($-\psi$)	Std. Error	Department	Median γ Coeff.
Female	-0.0737	-0.0799	*Mathematics*	3.5863
ACT read	-0.0022	-0.0097	*Engineering*	3.1147
ACT math	-0.0182	-0.0112	*Biology*	2.0092
HS GPA	-0.0041	-0.0872	English	1.7984
Black	0.2342	-0.1724	*Chemistry & Physics*	1.7766
Hispanic	-0.3174	-0.2519	Psychology	1.6432
Other Min.	-0.2687	-0.2609	Econ., Fin., Acct.	1.4914
First Generation	0.1068	-0.1149	Regional Studies	1.4854
Unobs. Type	0.1979	-0.0842	Communication	1.3657
	Effort elasticity		Languages	1.3352
ln(γ)	0.5445	(0.2055)	Social Sciences	1.2367
			Agriculture	0.9215
			Mgmt. & Mkting.	0.7984
			Education & Health	0.7696

5.3 Grade estimates

The estimated α's, the department-specific ability weights, are given in Table 6. These are calculated by taking the reduced-form θ's, undoing the normalization on the γ's, and subtracting off the part of the reduced form that θ's that reflect the study time (taken from ψ). The departments are sorted such that those with the lowest female estimate are listed first. Note that in all departments the female estimate is negative. This occurs because females study substantially more than males yet receive only slightly higher grades. Given that sorting into universities takes place on both cognitive and non-cognitive skills and that women have a comparative advantage in non-cognitive skills, males at Kentucky have higher cognitive skills than their female counterpart even though in the population cognitive skills are similar between men and women.

Negative estimates are also found for Hispanics. While Hispanics have higher grades than African Americans, our estimates of the study costs suggested that they also studied substantially more. Given the very high estimate of Hispanic study time we would have expected Hispanics to perform even better in the classroom than they actually did if their baseline abilities were similar to African Americans.

With the estimates of the grading equation, we can reported expected grades for an average student. We do this for freshmen, separately by gender, both unconditionally and conditional on taking courses in that department in the semester we study. Results are presented in Table 7. Three patterns stand out. First, there is positive selection into STEM courses: generally those who take STEM classes are expected to perform better than the average student. This is the not the case for many departments. Indeed, the second pattern is that negative selection is more likely to occur in departments with higher grades. Finally, women are disproportionately represented in departments that give higher grades for the average student. Of the seven departments that give the highest grades for the average student (female or male), all have a larger fraction female than the overall population. In contrast, of the seven departments that give the lowest grades, only three have a fraction female that is larger than the overall population.

5.4 Drivers of the STEM gap

Given the estimates of the student's choices over classes and effort and the estimates of the grading process, we now turn to examining sources of the male-female gap in choice of STEM classes. Table 8 shows share of STEM classes taken for males and females as well as how that share changes for women as we change different characteristics. The baseline share of STEM classes for men and women is 0.400 and 0.284, respectively. The first counterfactual changes female preferences for grades to be the same as male preferences for grades. This increases the share of STEM courses for women by almost one percentage point, closing the gender gap by almost eight percent.

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

Counterfactuals (4) and (5) look at differences in tastes. Counterfactual (5) turns off taste

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

	Chem & Physics	Econ. & Fin.	*Bio*	*Math*	English	Eng	Ag	Mgmt Mkting	Soc Sci	Lang Studies	Psych	Region	Comm	Ed & Health
Female	-0.297	-0.267	-0.240	-0.225	-0.225	-0.224	-0.209	-0.198	-0.195	-0.190	-0.179	-0.150	-0.145	-0.075
ACT read	0.000	0.005	0.005	-0.003	0.007	0.000	0.011	0.009	0.013	0.003	0.010	0.000	0.004	0.007
ACT math	0.011	0.011	-0.002	0.000	-0.007	-0.008	0.005	-0.010	-0.006	-0.003	-0.004	-0.001	-0.014	0.016
HS GPA	0.226	0.212	0.187	0.097	0.198	0.064	0.472	0.446	0.316	0.226	0.260	0.192	0.231	0.443
Black	0.115	0.094	0.091	0.196	0.217	0.124	0.034	-0.122	0.106	0.126	0.075	0.171	0.085	-0.121
Hispanic	-0.326	-0.276	-0.323	-0.286	-0.299	-0.306	-0.453	-0.285	-0.293	-0.421	-0.363	-0.180	-0.335	-0.387
Misc. Min.	-0.305	-0.261	-0.383	-0.304	-0.334	-0.349	-0.334	-0.214	-0.455	-0.285	-0.363	-0.350	-0.341	-0.534
1st Gen	0.038	0.057	0.033	0.073	0.010	0.102	0.017	0.040	-0.023	0.041	0.057	0.095	0.041	0.033
Type	0.913	0.982	0.897	0.575	1.221	0.630	0.939	1.989	1.418	1.199	1.018	1.143	1.186	1.724

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

	EGPA Females Unconditional	EGPA Females Conditional	EGPA Males Unconditional	EGPA Males Conditional	Share Female
Education & Health	3.71	3.67	3.54	3.16	0.70
Agriculture	3.49	3.28	3.39	2.96	0.56
Communication	3.42	3.39	3.16	3.17	0.56
Mgmt. & Mkting.	3.30	3.41	3.14	3.27	0.52
Languages	3.26	3.28	3.09	3.12	0.55
Regional Studies	3.20	3.31	2.97	3.11	0.66
Social Sciences	3.10	3.11	2.94	2.90	0.51
English	3.07	3.12	2.93	3.00	0.65
Psychology	2.98	2.98	2.75	2.73	0.67
Engineering	2.96	3.01	2.82	3.03	0.18
Econ., Fin., Acct.	2.78	2.93	2.77	2.91	0.37
Mathematics	2.72	2.78	2.59	2.71	0.47
Biology	2.69	2.82	2.59	2.78	0.60
Chem & Physics	2.47	2.62	2.51	2.74	0.47
Overall					0.51

differences for departments, which increases the share of women to 0.3, closing the STEM gap by 13 percent. These taste differences may be a mixture of pre-college experiences and the culture of different departments. Hence anything the university can do to close the STEM gap on this end is likely bounded above by this number. Counterfactual (6) turns off female preferences for female professors. One way of closing the gender gap in STEM would be to hire more female professors. However, even representation across fields would only close the gap by a little over three percent.

Finally, we examine how changing expected grades across departments affects the gender gap. Namely, we equalize mean grades across courses by increasing (or decreasing) the course-specific intercepts. However, there is still heterogeneity in grades due to the relative difference in γ's and α's, the former being especially important as it dictates the returns to studying. This counter-

Table 8: Decomposing the Gender STEM Gap

	PE Female		PE Male	
		Increase		Increase
	Share	Over Base	Share	Over Base
Baseline	28.40%		40.02%	
Turn off grade prefs (1)	29.34%	0.94%		
Turn off gender ability (2)	32.17%	3.78%		
Turn off observed ability (3)	28.89%	0.50%		
Turn off tastes (4)	29.94%	1.54%	29.2%	
Turn off female prof pref (5)	28.79%	0.40%		
Equalize expected grades for average student (6)	33.40%	5.00%	44.00%	3.98%

factual raises the share of STEM courses taken by females to about a third, higher than any of the counterfactuals. The reason the effects are larger here on the gender ratio than in the first counterfactual is that the returns to studying are much higher in STEM courses and women are willing to study more than men, due both to valuing grades more and having lower study costs. Note that the effect of the STEM gap is mitigated because men also shift towards STEM under this policy.

While the patterns here suggest a potentially cheap way of closing the gender gap is to equalize average grades across fields, professors are likely to respond to restrictions on grading policies. However, the response may further reduce the gender gap. The reason is that, if STEM courses are forced to give higher grades, they are likely to assign more work to deter entry. More work translates into higher γ's which make STEM courses relatively more attractive to women. The reverse holds for departments that are forced to lower their grades: in order to attract more students, they must lower workloads, implying lower values of γ which makes these courses relatively less attractive to women.

6 Equilibrium Grading Policies

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

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

In this section, we develop a model which describes how professors choose grading policies and propose a method for estimating the professor preference parameters of this model. The model assumes professors care about both the composition and outcomes of students who take their class and set policy parameters β_j and γ_j to influence these factors. In this framework, one source of differences in grading policies is differences in professor preferences for composition and outcomes of students. For example, professors with especially strong preferences for small classes will grade harshly to deter more students from choosing their course.

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

Because grading policies in all courses affect the choices of students, the composition of students in each course depends on the grading policies of all professors. This general equilibrium feature means that each professor's optimal grading policy depends on the grading policies of all other professors. We assume professors do not collude when choosing grading policies implying policies are set in a non-cooperative game between professors.

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

6.1 The Professor's Problem

We assume professors choose grading policy parameters β_j and γ_j to maximize an objective function which depends on both the composition and outcomes of students who take their class. This objective function needs to reconcile the following patterns in the data:

- the correlation between β_j and γ_j is extremely negative at -0.96,
- δ_{0j} is negatively correlated with β_j (-0.42) and positively correlated with γ_j (0.42), and
- STEM classes have even stronger negative (positive) correlations with β_j (γ_j) at -0.61 (0.71).

To capture these features of the data, we set up the professor's objective function to depend on the amount of learning in the class. However, there is diminishing marginal utility in classroom learning. In contrast, the cost to having more students in the class increases at a linear rate. Learning for student i in class j is given by the ability in the class plus study time: $A_{ij} + \ln(\phi_i) + \ln(\psi_i)$. Note that learning translates directly into grades by taking this term, multiplying it by γ_j and adding β_j. We also assume that there are quadratic costs associated with assigning more work and that these costs are heterogeneous across professors. Finally, professors have direct preferences over grades, not wanting their grades to deviate from certain (possibly departmental) norms. We specify these preferences over the expected performance of the average student.

Denote $\tilde{A}_{ij}$ and $\overline{A}_j$ as expected learning net of $\ln(\gamma)$ for student i and the average student respectively:

$$\tilde{A}_{ij} = A_{ij} + \ln(\phi_i) - \ln(\psi_i), \qquad \overline{A}_j = \frac{\sum_i^N A_{ij} + \ln(\phi_i) - \ln(\psi_i)}{N}$$

Then the objective function the professors maximize is:

$$\begin{aligned}
V_j(\beta,\gamma) &= \lambda_{1j}\ln\left[\sum_i P_{ij}(\beta,\gamma)\left(\tilde{A}_{ij}+\ln(\gamma_j)\right)\right] \\
&\quad -\lambda_{2j}\left[\sum_i P_{ij}(\beta,\gamma)\right]-(\lambda_3+e_{2j})\gamma-\lambda_4\gamma^2 \\
&\quad -\left[\beta_j+\gamma_j\left(\overline{A}_j+\ln(\gamma_j)\right)-e_{1j}\right]^2 \qquad (32)
\end{aligned}$$

where e_{j1} represents the professor's ideal grade for the average student[17] and e_{j2} allows for heterogeneity in the cost of assigning work.

We parameterize λ_{1j} and λ_{2j} so that they are allowed to vary with whether it is a STEM course and whether it is an upper division course. We are more flexible in how we parameterize e^*_{1j} and e^*_{2j}, allowing these vary flexibly at the department cross upper/lower level as well as having an error term. This yields the following structure for these four terms:

$$\begin{aligned}
\lambda_{1j} &= X_{\lambda j}\lambda_1 \\
\lambda_{2j} &= X_{\lambda j}\lambda_2 \\
e_{1j} &= X_{ej}\Psi_1+e^*_{1j} \\
e_{2j} &= X_{ej}\Psi_2+e^*_{2j}
\end{aligned}$$

where $X_{\lambda j}$ has a constant term and indicators for whether course j is STEM and upper division, where X_{ej} contains indicators for each department as well as interactions for each department with whether the course is upper division, and e^*_{1j} and e^*_{2j} are unobserved preferences that are orthogonal to the department and the level of the course.

Taking the first order condition with respect to β:

$$\begin{aligned}
0 &= \lambda_{1j}\left[\frac{\sum_i \frac{\partial P_{ij}}{\partial \beta_j}\left(\tilde{A}_{ij}+\ln(\gamma_j)\right)}{\sum_i P_{ij}\left(\tilde{A}_{ij}+\ln(\gamma_j)\right)}\right]-\lambda_{2j}\left[\sum_i \frac{\partial P_{ij}}{\partial \beta_j}\right] \\
&\quad -2\left[\beta_j+\gamma_j\left(\overline{A}_j+\ln(\gamma_j)\right)-e^*_{1j}\right] \qquad (33)
\end{aligned}$$

Given our estimates of student demand, grading policies, and study effort, the only unknowns in (33) are λ_{1j}, λ_{2j}, and e^*_{j1}. However, e^*_{1j} is correlated with β_j and γ_j.

[17] Note that this is the average student at the University of Kentucky, not the average student in the professor's class.

To deal with the endogeneity of β_j and γ_j, we instrument for β_j and γ_j using the innate demand for the course, δ_{0j}. The assumption is then that innate course demand once expected grades and effort are conditioned out is uncorrelated with professor grade preferences and workload targets. Following **?** we then substitute in for β_j and γ_j with the predicted values from a regression of $\hat{\beta}_j$ and $\hat{\gamma}_j$ on $\hat{\delta}_{0j}$ and department cross upper/lower level. Our estimating equation is then given by:

$$
\begin{aligned}
2\left[\hat{\beta}_j + \hat{\gamma}_j\left(\overline{A}_j + \ln(\hat{\gamma}_j)\right)\right] \quad = \quad & \lambda_{1j}\left[\frac{\sum_i \frac{\partial P^*_{ij}}{\partial \beta_j}\left(\tilde{A}_{ij} + \ln\left(\gamma^*_j\right)\right)}{\sum_i P^*_{ij}\left(\tilde{A}_{ij} + \ln\left(\gamma^*_j\right)\right)}\right] - \lambda_{2j}\left[\sum_i \frac{\partial P^*_{ij}}{\partial \beta_j}\right] \\
& -2X_{ej}\Psi_1 + e^*_{1j} \qquad (34)
\end{aligned}
$$

where terms with stars (with the exception of e^*_{1j}) contain β_j and/or γ_j are therefore instrumented. We then use (34) to obtain estimates of λ_1, λ_2, and Ψ_1 and can back out estimates of e^*_{1j}.

The remaining structural parameters are λ_3, λ_4, Ψ_2 and e^*_{2j}. We obtain these using the first order condition of (32) with respect to γ_j:

$$
\begin{aligned}
0 \quad = \quad & \lambda_{1j}\left[\frac{\sum_i \frac{\partial P_{ij}}{\partial \gamma_j}\left(\tilde{A}_{ij} + \ln(\gamma_j)\right)}{\sum_i P_{ij}\left(\tilde{A}_{ij} + \ln(\gamma_j)\right)} + \frac{\sum_i P_{ij}}{\sum_i P_{ij}\left(\tilde{A}_{ij} + \ln(\gamma_j)\right)\gamma_j}\right] \\
& -2\lambda_{2j}\left[\sum_i \frac{\partial P_{ij}}{\partial \gamma_{ij}}\right] - 2\left[\beta_j + \gamma_j\left(\overline{A}_j + \ln(\gamma_j)\right) - e_{1j}\right]\left[\overline{A}_j + \ln(\gamma_j) + 1\right] \\
& -(\lambda_{3j} + e_{2j}) - 2\gamma_j\lambda_4
\end{aligned}
$$

Note that we have estimates of all the terms on the first two lines. Call the sum of the first three terms D. Our estimating equation is then:

$$
D = \lambda_3 + 2\gamma_j\lambda_4 + X_{ej}\Psi_2 + e^*_{2j}
$$

We then instrument for γ_j to get λ_4 using the same first stage as in our original estimating equation.

6.2 Calculating equilibrium counterfactuals

Conditional on estimates of the β's and γ's, it is relatively simple to recover estimates of the professors' preference parameters. However, the reverse is not true. Given λ's and a counterfactual environment, it is computationally very demanding to find the 2,100-plus counterfactual β's and γ's.

To circumvent the computational demands, we approximate grading policies based on a set of covariates. Namely, we project the β's and $\ln(\gamma)$'s on a set of covariates including a flexible function of the professor parameters, baseline preferences for the course (δ_0), department fixed effects that are allowed vary by upper and lower level classes, and whether the professor was female.[18] These covariates explain over 90% of the variation in for both the β's and the γ's. We then use this set of covariates–plus the residual of this regression–to estimate counterfactual grading policies as functions of these covariates. Note that since the residual is included the professor first order conditions will be exactly satisfied in the baseline scenario.

Given this set of covariates, we search for coefficients that minimize the sum of squared errors associated with the first order conditions, minimizing over 70 parameters rather than over 2100. Although this is an approximation, we know exactly how well the approximation is performing and can additional covariates if the fit is poor.[19] Further, given the counterfactual grading policies, we can reestimate the professor parameters in the counterfactual environment. If the estimated professor preferences in the counterfactual are close to the values from the baseline scenario then we can be at least somewhat assured that we have a reasonable approximation to the counterfactual grading policies.

7 Conclusion

The lack of graduates in STEM majors–particularly among under-represented groups–has been of some policy concern. We show that there is a potentially cheap way to change the number and composition of STEM majors. Namely, grading policies have a substantial effect on sorting into STEM classes. We show that a substantial portion of the gender STEM gap can be removed by having STEM classes give grades that are on average similar to those in non-STEM classes.

These grading policies, however, are in part choices by professors. Hence administrative policies designed to change how professors grade will elicit responses by professors on other dimensions such as workload. These responses by professors may result in an even further closing of the STEM gender gap. Namely, if classes across departments are forced to give similar grades on average, then STEM (non-STEM) classes will employ alternative means to deter (encourage) enrollment in their courses by changing workloads. This will result in STEM classes assigning even more work

[18]The flexible function of the professor parameters includes the level of both parameters, the log of both parameters, and, in the case of β, the preference for studying taken to the $-.5$ power.

[19]Note we can see how much the fit improves relative to the fit provided by the baseline grading policies.

and non-STEM classes assigning less work. Since women are willing to study more, the increased STEM workload works as less of a deterrent to women taking STEM courses.

References

Altonji, J. G., Arcidiacono, P. & Maurel, A. (forthcoming), *Handbook of Labor Economics*, Vol. 5, Elsevier, chapter 7.

Altonji, J. G., Blom, E. & Meghir, C. (2012), Heterogeneity in human capital investments: High school curriculum, college major, and careers, Technical report, Annual Review of Economics.

Altonji, J. G., Kahn, L. B. & Speer, J. D. (2014), 'Trends in earnings differentials across college majors and the changing task composition of jobs', *The American Economic Review* **104**(5), 387–393.

Alvarez, L. (2012), 'To steer students toward jobs, florida may cut tuition for select majors', *New York Times* **Dec. 9th**.

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

Arcidiacono, P., Aucejo, E. M. & Hotz, V. J. (forthcoming), 'University differences in the graduation of minorities in stem fields: Evidence from california', *American Economic Review* .

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

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

Arcidiacono, P. & Miller, R. A. (2011), 'Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity', *Econometrica* **7**(6), 1823–1868.

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.

Bronson, M. A. (2014), Degrees are forever: Marriage, educational investment, and lifecycle labor decisions of men and women. Georgetown University.

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

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

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

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

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.

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

Pre (2012), *Report to the President: Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics.*

Rask, K. N. & Bailey, E. M. (2002), 'Are faculty role models? Evidence from major choice in an undergraduate institution', *The 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.

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

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

Appendix

Table A.9: STEM Classification by Department

Non-STEM			STEM
Aerospace Studies	Dept of Mkt and Supply Chain	Library & Info Science	Biology
Agr Economics	Dietetics and Nutrition	Linguistics	Chemical & Materials Engineering
Agricultural Biotechnology	Early Child, Spec Ed, Rehab	Merchand,Apparel & Textile	Chemistry
Agricultural Education	Education	Mil Sci and Leadership	Civil Engineering
Agriculture General	Education Curriculum & Instr	Modern & Classical Lang	Computer Science
Allied Health Ed & Rsrch	Ed Policy Studies and Evaluation	Nursing	Earth and Environmental Sciences
Animal & Food Sciences	Ed, School and Counseling Psych	Philosophy	Electrical & Computer Engineering
Anthropology	English	Plant Pathology	Engineering
Appalachian Studies	Environmental Studies	Plant and Soil Sciences	Entomology
Arts Administration	Family Sciences	Political Science	Mathematics
Accountancy	Fine Arts - Music	Psychology	Mechanical Engineering
Economics	Fine Arts - Theatre Arts	Public Health	Mining Engineering
Biosystems & Agr Engineering	Forestry	STEM Education	Physics And Astronomy
Business and Economics	Gender and Women's Studies	Schl Of Journalism & Telecomm	School of Architecture
Communication	Geography	Schl of Art and Visual Studies	Statistics
Communication Disorders	Health Sciences Education	Schl of Human Env Sci	
Communication & Info Studies	Hispanic Studies	Schl of Interior Design	
Community & Leadership Dev	History	Social Work	
Dept of Management	Kinesiology - Health Promotion	Sociology	
Dept of Gerontology	Landscape Architecture	Sustainable Agriculture	
Dept of Fin & Quant Methods	Latin American Studies		

Table A.10: Aggregation of Departments

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

Table A.11: Students with and without ACT scores

	Non-missing		Missing		
Variable	Mean	Std Dev	Mean	Std. Dev	p-value
Female	.50	.50	.49	.50	.353
Minority	.20	.40	.21	.47	.112
STEM Major	.31	.46	.29	.45	.033
GPA	2.99	.74	2.88	.80	.000
Observations	17,664		2,540		