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		

December 18, 2015

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 Please do not quote without permission.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 enrollment information, spanning Fall 2008 to Spring 2013. Each semester, the entire student body's course selections and grades are recorded by the Registrar's Office. Enrollment data can be linked across semesters to provide a complete panel data of every student's academic trajectories across the ten semesters. This yields approximately

1.4 million student/class observations. 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 in the context of a 'class bundle.' For this study, we focus on student enrollment observations from one semester, Fall 2012.

In addition, we have access to class evaluation surveys completed by students at the end of the semester. We note that coverage is not complete, as some departments chose not to make evaluation data available. Data from classes with a small number of student respondents are deleted, to prevent possible identification. For our Fall 2012 sample, we are able to link 1,086 classes to the enrollment data, which represents a 76 percent successful match rate. We use classes with at least a 70 percent response rate. 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 many hours per week spent studying for this course, expected final grade, and whether the course was a major requirement. Evaluation data cannot be linked to individual students; we use class average data.

Restricting the sample to Fall 2012 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.[4] Table 1 provides demographic summary statistics, separated by gender. Overall, women and men look similar when entering college. Women have slightly higher high school grades and slightly lower standardized ACT scores.[5] 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.

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/effort) than a non-STEM course.

Table 3 presents simple OLS results showing the relationship between individual and class char-

[4] We also observe withdrawal data.

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

acteristics with grades and study hours after controlling for a large number of academic background measures.[6] 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 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 demand for reasons besides grades 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. Classes that have more women also study more, consistent with the previous literature. 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.

3 Model

Individual i chooses n courses from the set $[1, \dots, J]$. Let $d_{ij} = 1$ if j is one of the n courses chosen by student i and zero otherwise. 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, $\mathbb{E}\left[g_{ij} \mid s_{ij}\right]$:

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

[6] We restrict our sample to standard classes with at least 16 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.

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.

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.

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

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

The grade student i receives in course j, g_{ij}, depends on the academic preparation of student 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.[7] 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.[8] 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:

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

[7] 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.

[8] 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.

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:

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

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

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.[9] 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

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

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.

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:

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

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

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

We estimate the reduced form parameters $\{\theta_{0j}, \theta_1, \theta_2\}$ as well as the structural slopes, the γ_j's, using nonlinear least squares. One γ_j must be normalized to separately identify the remaining parameters so we set one $\gamma_j = 1$ for each department.[10] The variation in the data used to identify

[10] 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.

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

4.2.2 Study parameters

We next turn to recovering some of the study effort parameters as well as undoing the normalization made on the γ's. 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 effort translates into hours linearly at rate μ but is reported with multiplicative measurement error ζ_{ij}:

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

Taking logs and substituting in for s^*_{ij} yields:

$$\begin{aligned} \ln(h_{ij}) &= \ln(\mu) + \ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i) + \zeta_{ij} && (17) \\ &= \kappa_0 + w_i\kappa_1 - X_i\psi_2 + \ln(\gamma_j) + \zeta_{ij} && (18) \end{aligned}$$

where:

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

Recall that we had to normalize one γ_j for every department in the grade equation. Denote the normalized values of γ as γ^* which relates to the unnormalized returns through $\gamma^*_j = \gamma_j / C_{k(j)}$.[11] Given our estimate $\hat{\gamma}^*$ from the grade equation and rearranging terms we have:

$$\ln(h_{ij}) - \ln(\hat{\gamma}^*) = \kappa_0 + w_i\kappa_1 - X_i\psi_2 + \kappa_{2k(j)} + \zeta_{ij} \tag{21}$$

where $\kappa_{2k(j)} = \ln(C_{k(j)})$.

[11] Note that one γ_j is still normalized to one, just not one γ_j in each 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:

$$\frac{\sum_i (l_i = l) d_{ij} \ln(h_{ij})}{\sum_i (l_i = l) d_{ij}} - \ln(\hat{\gamma}^*) = \kappa_0 + w_{jl}\kappa_1 - X_{jl}\psi_2 + \kappa_{2k(j)} + \zeta_{jl} \tag{22}$$

where w_{jl} and X_{jl} are the averages of these characteristics for those of year level l enrolled in course j.

Our estimate of the unnormalized returns to study and ability and then be recovered using $\hat{\gamma}_j = \hat{\gamma}^*_j \exp(\kappa_{2k(j)})$. The department-specific weights on each of the observed characteristics (with the exception of gender) can now be recovered given $\hat{\gamma}$ and the estimates of the study costs, ψ_2. Namely, $\hat{\alpha}_{2k(j)} = \hat{\psi}_2 + \hat{\theta}_{2k(j)} \exp(-\kappa_{2k(j)})$.

Disentangling female preparation in each of the departments, $\hat{\alpha}_{1j}$, and female study costs, $\hat{\psi}_1$, first requires estimates of female preferences for grades (described in the next section). We cannot separate out μ, ψ_1, and the normalization on γ for one of the departments. However, the lack of identification does not affect the decompositions we perform or our counterfactual policy changes.

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:

$$\widehat{E[g_{ij}]} = \hat{\theta}_{0j} + \hat{\gamma}^*_j \left(w_i \hat{\theta}_{1k(j)} + X_i \hat{\theta}_{2k(j)} \right) \tag{23}$$

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:

$$U_{ij} = \delta_{0j} + w_i \delta_{1k(j)} + Z_{ij} \delta_{2k(j)} + \left(\widehat{E[g_{ij}]} - \hat{\gamma}_j \right) (\phi_0 + w_i \phi_1) + \epsilon_{ij} \tag{24}$$

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, $\delta_{2k(j)}$, and preferences over grades, ϕ.

We assume that ϵ_{ij} is distributed Type 1 extreme value. If individuals were choosing one course, estimation of the parameters in (24) 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.

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:

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

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:

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

Given $\hat{\phi}_0$ and $\hat{\phi}_1$, we can calculate $\hat{\psi}_1$ using (20) where $\hat{\psi}_1 = \ln(\hat{\phi}_0 + \hat{\phi}_1) - \ln(\hat{\phi}_0) - \hat{\kappa}_1$ from the study equation. We can then recover $\hat{\alpha}_{1k(j)}$ from (14) where $\hat{\alpha}_{1k(j)} = \hat{\theta}_{1k(j)} - \ln(\hat{\phi}_0 + \hat{\phi}_1) + \ln(\hat{\phi}_0) + \hat{\psi}_1$ from the grade equation.

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:

$$\sum_i \ln\left(\sum_{s=1}^{S} \pi_s \mathcal{L}_{igs}\left(\theta, \gamma\right) \mathcal{L}_{ics}\left(\varphi\right)\right) \tag{26}$$

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:

$$q_{is}^{(m)} = \frac{\pi_s^{(m)} \mathcal{L}_{igs}\left(\theta^{(m-1)}, \gamma^{(m-1)}\right) \mathcal{L}_{ics}\left(\varphi^{(m-1)}\right)}{\sum_{s'} \pi_{s'}^{(m)} \mathcal{L}_{igs'}\left(\theta^{(m-1)}, \gamma^{(m-1)}\right) \mathcal{L}_{ics'}\left(\varphi^{(m-1)}\right)} \tag{27}$$

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:

$$\sum_i \sum_s q_{is}^{(m)} \ln \left[\mathcal{L}_{igs}\left(\theta, \gamma\right)\right] \tag{28}$$

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

$$\sum_i \sum_s q_{is}^{(m)} \ln \left[\mathcal{L}_{ics}\left(\varphi\right)\right] \tag{29}$$

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

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 differences in preferences can also be linked to characteristics of the instructor. For example, we can link courses to gender of the instructor and see the extent to which female students prefer female professors by regressing δ_{1j} on indicators for whether the professor was female and departmental fixed effects. We can then use these estimates to forecast how the course choices would change if each department had a larger (or smaller) representation of female professors, holding fixed grading standards.

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 Equilibrium Grading Policies

Examining the effects of grading policies like those at Wellesley where the fraction of A's are capped are difficult to analyze because professors can respond to constraints on the number of A's given by changing the returns to effort as well as the intercept. For example, if certain departments are more generous in handing out A's than what the policy mandates, a way to keep their courses attractive is to require less work. This is equivalent to lowering γ_j. Similarly, if a policy were to mandate that some departments increase their grades, they can deter some of the increases in enrollment by requiring more work by increasing γ_j. In this section we specify the objective function of the professor and, supposing that the professor is restricted to linear grading policies, describe its solution. We then discuss the sorts of counterfactuals that can be conducted once the professor's preferences are recovered.

5.1 The Professor's Problem

Professor payoffs are assumed to be a function of:

(i) the total amount of learning in the course: $a(\beta, \gamma)$,

(ii) total enrollment: $b(\beta, \gamma)$, and

(iii) student study time: $c(\beta, \gamma)$.

Learning for student i is what is rewarded in the grade equation, the term hit by γ_j. Learning for individual i in course j is given by:

$$\begin{aligned} L_{ij}(\gamma_j) &= A_i + \ln(s_{ij}^*) \\ &= A_i + \ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i) \end{aligned} \tag{30}$$

Student study time is included as a way of capturing student complaints as workload increases.[12] Note that this negative component to the professor's utility is *separate and distinct* from the utility gained from students' learning. The professor's objective function is then assumed to follow:

$$\begin{aligned} V_j &= \lambda_{0j} a(\beta,\gamma) - \lambda_{1j} b(\beta,\gamma) - \lambda_{2j} c(\beta,\gamma) \\ &= \lambda_{0j}\left[\sum_i P_{ij}(\beta,\gamma)\left(A_i + \ln(\phi_i) + \ln(\gamma_j) - \ln(\psi_i)\right)\right] \\ &\quad -\lambda_{1j}\left[\sum_i P_{ij}(\beta,\gamma)\right] - \lambda_{2j}\left[\sum_i P_{ij}(\beta,\gamma)\frac{\phi_i\gamma_j}{\psi_i}\right] \end{aligned} \tag{31}$$

The term P_{ij} is the probability of student i choosing course j, defined in the student's problem. Utility functions are only identified up to scale. Hence we normalize λ_{0j} to one.

Professors are assumed to know the preferences of the other professors and all professors simultaneously set their grading standards in a non-cooperative fashion. In a pure strategy equilibrium, the first order conditions of the professors problem must be satisfied at the grading standards observed in the data. The choice of β_j and γ_j satisfy the two first order conditions:

$$\frac{\partial V_j}{\partial \beta_j} = 0 = \frac{\partial a(\beta,\gamma)}{\partial \beta_j} - \lambda_{1j}\frac{\partial b(\beta,\gamma)}{\partial \beta_j} - \lambda_{2j}\frac{\partial c(\beta,\gamma)}{\partial \beta_j} \tag{32}$$

$$\frac{\partial V_j}{\partial \gamma_j} = 0 = \frac{\partial a(\beta,\gamma)}{\partial \gamma_j} - \lambda_{1j}\frac{\partial b(\beta,\gamma)}{\partial \gamma_j} - \lambda_{2j}\frac{\partial c(\beta,\gamma)}{\partial \gamma_j} \tag{33}$$

The solution to this system is then:

$$\begin{bmatrix} \lambda_{1j} \\ \lambda_{2j} \end{bmatrix} = \begin{bmatrix} \frac{\partial b(\beta,\gamma)}{\partial \beta_j} & \frac{\partial c(\beta,\gamma)}{\partial \beta_j} \\ \frac{\partial b(\beta,\gamma)}{\partial \gamma_j} & \frac{\partial c(\beta,\gamma)}{\partial \gamma_j} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial a(\beta,\gamma)}{\partial \beta_j} \\ \frac{\partial a(\beta,\gamma)}{\partial \gamma_j} \end{bmatrix} \tag{34}$$

Note that, by substituting in the estimates of the professor's own grading practices, the estimates of the grading practices of the other professors, and the estimates of student preferences, the derivatives on the right hand side (34) can be calculated, allowing us to recover the professor preferences.

[12] Another way to interpret this term is to regard it as a proxy for professor's own time lost due to increased need to help students learn via extra classes, more preparation, or additional office hours.

5.2 Implications

Once professor preferences are recovered, we can see how characteristics of these vary by department as well as with other characteristics of the professor such as gender or professor rank.[13] Counterfactual policies, however, would require solving for the new equilibrium. Given the professor preferences, we could then see how the solution to the grading standards game would change given different environments. Note that this would entail solving a fixed point problem among all the professors.

Conditional on the feasibility of solving for new equilibria, a number of counterfactuals can be conducted. First, we can see how the equilibrium grading policies would change if all professors had the same preferences and correspondingly how enrollment in STEM courses by males and females would change as well. Second, we can examine what the equilibrium grading policies would look like if preferences for courses, the δ_j's, were the same. This would then give a sense for how much of the differences we see in grading practices are driven by demand versus professor preferences. Finally, we can investigate how placing caps on the mean grade given would affect the equilibrium.

6 Results

6.1 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 a little over 20% 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 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

[13] The data also categorizes teachers into: student, staff, post-doc, faculty, and retired faculty. Rank of faculty is easily discernible from publicly available data.

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: 1.57 points, which translates into over 3 grade points. Engineering, however, is an outlier with all the other gaps smaller that one point (or 2 grade points).

Table 4: Estimates of Preference Parameters

Preference for:	Coeff.	Std. Error
Expected grades (ϕ)	0.4124	(0.0197)
Female × expected grades	0.0800	(0.0131)
Female × female professor	0.1460	(0.0186)
Female preferences for Departments		
Regional Studies	0.1981	(0.0695)
Communication	-0.1714	(0.0535)
Education & Health	0.3266	(0.0582)
Engineering	-1.1963	(0.0664)
Languages	-0.1241	(0.0580)
English	0.3663	(0.0760)
Biology	0.2277	(0.0631)
Mathematics	-0.1212	(0.0598)
Chemistry & Physics	-0.1558	(0.0603)
Psychology	0.3440	(0.0651)
Social Sciences	-0.3067	(0.0542)
Mgmt. & Mkting.	0.1220	(0.0662)
Econ., Fin., Acct.	-0.5278	(0.0589)

6.2 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 19.5%,[15] again conditional on taking the same class resulting in an overall effect of over 26.5%.

The results also suggest lower study costs for higher ACT scores but, surprisingly, higher study costs for higher high school grades. While blacks and first generation students have higher study costs, Hispanics and miscellaneous minorities have lower study costs. Those who are the second unobserved type have lower study costs and, as we will see in subsequent tables, also are more able.[16]

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 almost a quarter of a grade increase in mathematics but would translate into less than a tenth of a grade point in agriculture, management and marketing, and education.

6.3 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

[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 40% 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: $\ln(\phi_0+\phi_1)-\ln(\phi_0)=\ln(.4963)-\ln(.4083)$.

[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.

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

	Study Effort			Median γ
	Coeff. ($-\psi$)	Std. Error	Department	Coeff.
Female	-0.0228	(0.0651)	Agriculture	0.9618
ACT read	-0.0071	(0.0088)	Regional Studies	1.1799
ACT math	0.0297	(0.0095)	Communication	1.1333
HS GPA	-0.0777	(0.0840)	Education & Health	0.8072
Black	0.0152	(0.1437)	Engineering	1.3908
Hispanic	0.2186	(0.2432)	Languages	1.0656
Other Min.	0.3634	(0.3152)	English	1.1132
First Generation	0.0296	(0.1016)	Biology	1.6240
Unobs. Type	0.1091	(0.0554)	Mathematics	2.0694
			Chemistry & Phsyics	1.6304
			Psychology	1.2581
			Social Sciences	0.9741
			Mgmt. & Mkting.	0.9056
			Econ., Fin., Acct.	1.2465

skills, males at Kentucky should have higher cognitive skills even if 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 the actually did if their baseline abilities were similar to African Americans.

STEM classes tend to reward ACT math scores relatively more than other departments, though the effects may not be as large as one might expect. However, it is important to note that STEM classes also have higher γ's: the estimates of the α 's give the returns to ability when γ is one, but the estimates of γ for STEM courses are generally much higher than one.

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, only one has a smaller fraction female than the overall population. In contrast, of the seven departments that give the lowest grades, only two have a fraction female that is larger than the overall population.

6.4 Professor preference estimates

We now turn to the estimates of the professor preference parameters. Recall that professor utility was a function of (i) total learning (the sum of student ability in the class plus the amount of studying), (ii) squared enrollment (so professors may value the first enrolled student more than the second), (iii) a distaste for assigning more work. With the payoff to learning to normalized to one, Table 8 shows the mean and standard deviation of the disutility of squared enrollment and assigning more work by upper and lower classes and STEM/non-STEM.

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

	Chem *Physics*	Econ & Finance	*Bio*	English	Ag	Mgmt Mkting	*Math*	Soc *Sci*	*Eng*	Lang	*Psych*	Reg Studies	Comm	Ed & Health
Female	-0.199	-0.167	-0.135	-0.112	-0.111	-0.106	-0.099	-0.085	-0.084	-0.070	-0.061	-0.021	-0.017	-0.003
ACT read	0.009	0.016	0.016	0.019	0.019	0.017	0.005	0.024	0.011	0.014	0.022	0.009	0.014	0.015
ACT math	0.003	0.004	-0.010	-0.015	-0.008	-0.023	0.002	-0.016	-0.006	-0.010	-0.012	-0.008	-0.024	-0.001
HS GPA	0.329	0.326	0.310	0.349	0.517	0.451	0.256	0.437	0.227	0.371	0.396	0.324	0.368	0.451
Black	-0.146	-0.177	-0.189	-0.038	-0.200	-0.311	-0.084	-0.159	-0.256	-0.152	-0.207	-0.095	-0.200	-0.312
Hispanic	-0.228	-0.171	-0.226	-0.194	-0.343	-0.191	-0.164	-0.191	-0.194	-0.350	-0.273	-0.046	-0.241	-0.277
Other Min.	-0.403	-0.354	-0.502	-0.451	-0.423	-0.318	-0.425	-0.572	-0.539	-0.384	-0.478	-0.466	-0.453	-0.585
First Gen.	-0.105	-0.087	-0.119	-0.159	-0.113	-0.085	-0.090	-0.175	-0.040	-0.113	-0.089	-0.045	-0.111	-0.091
Unobs. Type	0.671	0.794	0.742	1.262	0.574	1.378	0.560	1.261	0.836	1.164	0.881	1.081	1.110	1.165

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	3.52	3.39	3.37	3.10	0.60
Communication	3.35	3.35	3.10	3.09	0.54
Agriculture	3.31	3.22	3.21	3.06	0.74
Languages	3.25	3.22	3.11	3.08	0.54
Mgmt & Mkting	3.10	2.78	2.97	2.72	0.51
Regional Studies	2.99	3.09	2.78	2.94	0.66
Biology	2.97	3.01	2.88	2.85	0.65
English	2.95	2.94	2.83	2.82	0.64
Social Sciences	2.94	2.93	2.79	2.69	0.50
Engineering	2.89	3.04	2.77	2.88	0.16
Mathematics	2.77	2.83	2.66	2.74	0.49
Econ., Fin., Acct.	2.77	2.66	2.77	2.68	0.32
Psychology	2.75	2.69	2.53	2.44	0.69
Chem & Physics	1.99	2.18	2.05	2.28	0.49
Overall					0.52

For lower-level classes, the marginal student is more costly in non-STEM classes than in STEM classes. This actually translates into higher grades in STEM classes as higher grade intercepts increases enrollment with no effect on study time. At the same time, assigning work is less costly in STEM classes (though not significantly so) implying STEM courses will set higher returns to studying, which in turn may result in lower grading intercepts. Large classes are more costly for upper-level STEM courses than lower-level STEM courses, though the pattern is surprisingly reversed for non-STEM courses.

Table 8: Professor Preferences Over Enrollment and Study Time by Course Level and Type

		Disutility of:			
Level	Category	Enrollment[2] (λ_{1j})		Study Time (λ_{1j})	
		Mean	Std. Dev.	Mean	Std. Dev.
Lower	non-STEM	0.060	0.036	1.832	0.784
Level	STEM	0.038*	0.041	1.752	0.489
Upper	non-STEM	0.049†	0.080	1.942	1.016
Level	STEM	0.069*,†	0.036	1.922	1.062

Note: * indicates STEM mean statistically different at 95% level from non-STEM mean at the same course level. † indicates mean upper-level course statistically different at the 95% level from mean lower-level course of the same category (STEM/non-STEM).

7 Counterfactuals

Given the estimates of the student's choices over classes and effort and given the estimates of the grading process, we now turn to examining the sources of the male-female gap in choice of STEM classes. Table 9 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 course for women to 0.293, closing the gender gap by eight percent. Turning off observed ability differences such as differences in ACT scores and high school grades (row 3) and study costs (row 4) have smaller effects on the gap, 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. The next three counterfactuals, which equalize ability differences and costs of studying, all have smaller effects.

Counterfactuals (5) through (7) look at differences in tastes. Counterfactual (5) turns off taste differences for departments, which increases the share of women to 0.31, closing the STEM gap by 22 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.

Table 9: Decomposing the Gender STEM Gap

	STEM (Bio+Math+Chem+Eng)		
	Share	Pct of Gap	Gender Ratio
Baseline Male	0.400		
Baseline Female	0.284		0.710
(1) Turn off differences in grade prefs ($\phi_1 = 0$)	0.293	0.080	0.733
(2) Turn off gender ability differences ($\alpha_1 = 0$)	0.301	0.142	0.751
(3) Turn off observed ability differences ($\overline{X}_f = \overline{X}_m$)	0.289	0.043	0.722
(4) Turn off study effort differences ($\psi_1 = 0$)	0.286	0.013	0.713
(5) Turn off taste differences	0.310	0.220	0.773
(6) Turn off female professor pref	0.288	0.034	0.720
(7) Turn off both female professor pref and taste	0.314	0.324	0.784

Our next set of counterfactuals focus on grading policies. Results are presented in Table 10. We consider two counterfactuals: (i) adjusting grading intercepts for each course such that the expected grade for the average student is the same across courses and (ii) changing preferences of STEM professors to on average match those of humanities professors.

The first grading policy counterfactual equalizes expected grades across courses for the average student by increasing (or decreasing) the course-specific intercepts until expected grades are the same for the average student. However, there is still heterogeneity in grades due to the relative difference in γ's and α's, the former being especially important as it dictates the returns to studying. This counterfactual raises the share of STEM courses taken by females to .334, higher than any of the counterfactuals in Table 9. The gender ratio also tilts significantly towards females but not as much as in counterfactuals (5) and (7) because men too see their probabilities of taking STEM courses increase. 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.

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.

The final counterfactual changes STEM professor preferences by the average difference between STEM and non-STEM professor preferences so that the means are the same. Because professors in lower-level STEM courses have lower disutility from increased class size than their non-STEM counterparts, equalizing preferences exacerbates the gender gap. Lower shares for STEM classes are seen for both men and women but the effects are larger for women.

Table 10: Supply-Side Counterfactuals

	STEM (Bio+Math+Chem+Eng)		
	Male Share	Female Share	Gender Ratio
Baseline	0.400	0.284	0.710
(8) Equalize expected grades for average student	0.440	0.334	0.760
(9) Change STEM prof prefs to non-STEM prof prefs	0.398	0.278	0.700

8 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 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.11: 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.12: 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.13: 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		