

GENDER (STILL) MATTERS IN BUSINESS SCHOOL¹

Aradhna Krishna and A. Yeşim Orhun

Ross School of Business

University of Michigan

December 9, 2020

Abstract

This research documents systematic gender performance differences (GPD) at a top business school using a unique administrative dataset and survey of students. The findings show that women's grades are 11% of a standard deviation lower in quantitative courses than those of men with similar academic aptitude and demographics, and men's grades are 23% of a standard deviation lower in nonquantitative courses than those of comparable women. The authors discuss and test for different reasons for this finding. They show that a female instructor significantly cuts down GPD for quantitative courses by raising the relative grades of female students. In addition, female instructors increase women's interest and performance expectations in these courses and are perceived as role models by their female students. These results provide support for a gender stereotype process for GPD and show that faculty can serve as powerful exemplars to challenge gender stereotypes and increase student achievement. The authors discuss several important implications of these findings for business schools and for society.

¹ The authors are listed alphabetically. We thank Paul Kirsch, Regina Zmich, Heather Bryne, Lillian Chen, Lisa Kozlo and Kevin Gates for their help in providing access to and support with institutional data from the business school. Max Resnick and Bruno Castelo Branco provided excellent research assistance. The authors also thank Timothy McKay and Mark Umbricht for their support, extensive knowledge and help with the administrative dataset.

INTRODUCTION

Women tend to be under-represented in high-paying careers and jobs in management. The overall gender disparity in wages is substantial -- women earn 77 cents for every dollar earned by men -- in management occupations in the U.S.² The paucity of women in some business sectors contributes a large share of the female gender gap in wages (Blau and Kahn 2017) and handicaps economic growth (Hsieh et al., 2019). It can also have a negative impact on corporate structure and culture with compounding downstream consequences for women's success. Discrimination against women in terms of salaries earned and promotions garnered has been amply documented in many fields with lopsided gender representation, notoriously so in the technology firms of Silicon Valley (Cao, 2017); as well as in finance jobs, academia and local government (Kolhatkar, 2017). While lawsuits may take care of specific instances of discrimination, a contributor to the corporate culture that fosters such discrimination – namely, gender imbalance – remains to be addressed.

Why is there gender imbalance in the business fields where it exists, and how can it be changed? Many technology positions in Silicon Valley firms, the current hotbed of highly lopsided gender ratio, represent quantitative business fields and encompass many business disciplines such as marketing, finance, operations. Many other top-paying companies and industries also have substantial levels of gender imbalance and wage gaps, especially at higher levels of management (e.g., financial services, Edelman et al., 2019; consulting, Marriage, 2019). Business schools train the talent that fuel most of the management jobs. Even though business schools in the U.S. boast a 43%–47% of female student representation according to the Department of Education (2018), if female students are not successful in the academic paths that lead to more lucrative career options, the pipeline issue cannot be fixed. Therefore, it is imperative to explore what gendered achievement disparities exist in business education and whether interventions can reduce or expunge these disparities.

² Highlights of Women's Earnings in 2015 U.S. Department of Labor Bureau of Labor Statistics, 2016 (Report 1064).

In this paper, we investigate the differences in academic achievement of male and female students at a top undergraduate business program and examine what drives these differences. Our empirical analyses focus on the grades of 6,312 undergraduate students who represent the 2005–2018 graduating classes, focusing on students' academic performance in the introductory courses of the core curriculum, where coursework is mandatory and students are randomly assigned to different sections of a course. This focus eliminates concerns about student self-selection. Due to the coordinated nature of the core curriculum, all sections of a course in a given term have the same syllabi, materials, and exams, further aiding comparability of performance outcomes. We supplement the administrative data with a survey of current undergraduate business students to assess their expectations, interests, and perceptions across different types of courses.

We document stark gender disparities in academic performance across different types of courses, even after we control for each student's initial academic aptitude measures, current grade point average (GPA), family background, and a rich set of other demographics as captured by their college applications. While the grades of women are systematically lower than those for men in quantitative courses (e.g., 33.7% of a standard deviation lower in finance), they are systematically higher than those for men in others (e.g., 33.5% of a standard deviation higher in organizational behavior). We show a correlation between how quantitative a course is perceived to be by students and the gender performance difference (GPD) levels we document. Specifically, a more quantitatively perceived course is correlated with a positive GPD (men outperform women), and a less quantitatively perceived course is correlated with a negative GPD (women outperform men). We also show that women also are less likely to take jobs in high-paying sectors (i.e. investment banking, technology industry), which leads to a gender wage gap at the time of graduation.

A priori, it is unclear why student gender affect academic performance across quantitative and non-quantitative courses. Why should students of different genders, otherwise similar in their academic aptitudes, family background, and other demographics, perform differently in business school classes? In STEM (science, technology, engineering, and math) programs, women tend to perform worse than men (for a recent review, see Kahn and Ginther 2017). While the commonly cited potential reason of low female

representation does not apply in the case of business education, the education and psychology literatures have discussed two other potential explanations that may apply. The first explanation is that men (women) innately prefer quantitative (nonquantitative) courses (innate preferences hypothesis). The second explanation is based on gender stereotypes, and may be student-based or instructor-based. The student-based explanation is that gender stereotypes held by students can influence their performance in different courses (gender stereotype hypothesis—student-based). The instructor-based explanation is that instructors hold these gender stereotypes and subsequently influence student performance through instructor behaviors, and evaluations (gender stereotype hypothesis—instructor-based).

It is important to examine the respective congruency of these potential explanations with the grade disparities observed in the business education context. If innate preferences drive GPD in business education, then interventions are not likely to be effective and may not improve welfare. However, if GPD is shaped by gender stereotypes, then interventions that challenge those stereotypes might reduce GPD. One intervention that can challenge students' gender stereotypes is an instructor who is a successful counterexample to the stereotype.

We provide evidence for the causal impact of instructor gender on GPD. Overall, we find that women's grades in quantitative courses are 11% of a standard deviation lower than that of men. However, when taught by female instructors, we document that female students' performance improves by 7.7% of a standard deviation in quantitative courses and that this effect is driven by female students with mid to high math aptitudes. In addition, we find that female instructors increase female students' initial interest and performance expectations in quantitative courses and are viewed as role models by female students. This pattern of results provides support for a student-side process explanation of gender stereotypes and demonstrates that instructors can improve students' academic success by providing a counterexample to the stereotype. Furthermore, we show that the data do not corroborate an explanation for GPD based mainly on instructor bias, because the degree of subjectivity in grading does not change the impact of instructor gender on GPD and because students do not perceive a difference in the fairness of instructors based on their genders.

We find, however, that instructor gender does not affect male students' performance, interest, or perception of instructors in any of the courses. The finding that instructor–student gender match in nonquantitative courses, in which male students have an observed handicap, does not improve male student's performance suggests that male students' underperformance may either be driven by innate preferences and not by gender stereotypes, and/or that having a male instructor in nonquantitative classes is not effective in challenging gender stereotypes regarding these courses.

Our paper contributes to the long-standing marketing literature on consumer behavior research that has examined gender and gender-identity differences. This literature not only considered gender differences in responses to advertising and messaging (e.g., Dahl et al., 2009; Fisher and Dube, 2005; Grohmann, 2009; Meyers-Levy and Loken, 2015), but also gender and identity differences in more global preferences (Meyers-Levy and Loken, 2015; Govind, Garg and Mittal, 2020; Gao, Mittal and Zhang, 2020). Our paper also contributes to the education literature, which we discuss in more detail in the next section. In that literature, our paper is closest to Carrell et al. (2010), which documents a positive role of instructor gender in reducing GPD in STEM courses in the Air Force Academy. We contribute to the education literature in three main ways. First, we study GPD within a business school at a large public university, with immediate downstream implications for gender disparities that have garnered much attention in the corporate sector. Second, while most studies of the impact of instructor gender on student achievement have been correlational, as in Carrell et al. (2010), our identification strategy relies on exogenous student-instructor assignment to address identification challenges that arise from students' self-selection into courses or instructors that may have confounded other work. We also improve upon Carrell et al. (2010) by controlling for unequal instructor gender representation across courses, which may otherwise lead to an aggregation bias due to a potential correlation between instructor gender representation and GPD across courses. Such a correlation is plausible because academic success in undergraduate programs in a field is a strong predictor of pursuing graduate work in that field (Sax, 2001). Third, we provide empirical evidence regarding the congruency of different potential mechanisms behind the observed GPD patterns. While these potential mechanisms have been discussed in the literature, causal evidence supportive of them has been scarce. We

provide insight into these mechanisms both by examining the grade outcomes of students across instructors of different genders and by conducting a supplemental survey of the same student population.

Our results have several important implications for business education and the corporate careers it prepares students for. Even though business schools themselves have achieved gender representation close to parity, this paper shows that equal representation has not translated to equal academic achievement. Academic achievement gaps not only create inequity in education, but also may have far-reaching consequences, because they may shape occupational choices (Ost, 2010). We find that female students who have a mid-to-high aptitude for quantitative subjects benefit the most from role models. It is disconcerting that the professional world may be missing out on these women due to the misallocation of talent driven by stereotyped beliefs. Therefore, arresting these gaps in situations where they may arise (such as at business schools) is paramount to create more gender-healthy pipelines for senior management positions.

There is hope. Our results also provide an important avenue for reducing gaps in academic achievement disparities within business schools. We find that female students achieve higher grades in quantitative courses when they are taught by female instructors. As such, business school administrators can help align female students' abilities more accurately with their educational choices by assigning more female instructors to teach quantitative courses.

If better alignment leads to a better allocation of students to careers, this would create a shift with several significant benefits. A reduction of GPD in quantitative courses may increase the representation of talented women in careers that quantitative courses prepare students for (e.g., finance, consulting, consumer technology), which are typically more lucrative. In addition, women's success in quantitative fields can help recruiters hire and retain a more diverse workforce—a goal that many top-paying companies have underscored as being paramount. For both these reasons, business schools also have much to gain from reducing GPD.

Our results indicate that faculty teaching assignments and hiring practices in business schools have important downstream consequences for students and employers. The direct implications of female faculty

representation in quantitative fields with regard to disparities in women's academic achievements in business education also speaks to the importance of reducing gender gaps in universities' hiring practices. With regard to marketing faculty, as the discipline of marketing becomes more data-driven and analytical, our results suggest the need to hire more female faculty to bolster female students' success in these fields.

In what follows, we provide a review of related literature, describe our data, and provide evidence for GPD in business education. We then discuss potential drivers of GPD and test for their implications in our data. We conclude by offering recommendations based on our findings.

LITERATURE REVIEW

Given the equity implications and downstream consequences, a large body of prior research has studied GPD in academic achievement. Prior GPD research has focused primarily on pre-university education (elementary, middle, and high school) around the world. The focus on GPD at the university level is more recent, and only a few studies have been conducted. We first discuss these GPD studies and then discuss research on instructor gender effects on GPD. Importantly, GPD findings in the extant literature, and as well as instructor effects on GPD, are mixed, highlighting the necessity and importance of examining these questions in a particular educational context.

Prior research at the elementary and middle school levels documents evidence, across countries, both for boys outperforming girls and for girls outperforming boys. For example, girls have been shown to do better in reading, writing, and math in elementary and middle school in England (Machin and McNally, 2005). However, in other results, boys outperform girls in math; for example, Fryer and Levitt (2010) find no mean differences between boys and girls in terms of their math performance upon entry to kindergarten in the United States but find that girls lose more than two-tenths of a standard deviation relative to boys over the first six years of school.

GPD findings for high school students are also inconsistent across countries. Using data on individual student performance across 41 countries, including the United States, Machin and Pekkarinen (2008)

demonstrate that among 15-year-olds who take the same standardized tests, female students outperform male students in reading, but male students outperform female students in math. In the contexts of Korean and Chinese high schools, however, Lim and Meer (2017) and Xu and Li (2018) find that girls generally outperform boys, including in math.

Researchers focusing on GPD beyond high school have focused a lot on STEM disciplines, where most courses are quantitative, such as introductory physics programs (Kost, Pollock and Finkelstein 2009; Lorenzo, Crouch and Mazur 2009; Miyake et al. 2010). Koester, Grom, and McKay (2016) use data from a large midwestern university's 116 introductory courses in STEM, social sciences, and humanities. They document a large GPD in STEM courses and in economics in favor of men. They do not find GPD favoring women over men in any courses, including social sciences and humanities. Carrell, Page, and West (2010) examine the context of the Air Force Academy, where women account for only 17% of the student body. They document that female students do worse than their male peers in STEM courses but do not find evidence of GPD in English and history courses.

An important question is whether GPD can be attenuated. Studies have discussed several possibilities (for a recent review, see Ceci et al. 2014). One factor that has been considered is instructor gender. However, only a few studies have been able to address this question causally because of the difficulty of finding a context in which students do not self-select into courses and/or instructors.

Starting again with primary and secondary education, we find that the effect of instructor gender on GPD depends on the context studied. While girls outperform boys in general in Korean and Chinese high schools, having female instructors has a positive effect on female student academic performance (Gong, Lu, and Song 2018; Lim and Meer 2017; Xu and Li 2018). However, in U.S. primary schools, Antecol, Eren, and Ozbeklik (2015) document a negative effect of having a female instructor on female students in math classes. While these studies find a positive or negative gender-match effect for females, other research at the primary and secondary education finds no significant effect of a gender match on outcomes (e.g., Puhani, 2018; Winters et al. 2013).

At the university level, results for gender-match effects have also been mixed, with Hoffman and

Oreopoulos (2009) finding that male students performed worse with a female instructor in the University of Toronto's Arts and Science program, Carrell et al. (2010) finding positive effects only for female instructors on female performance in the Air Force Academy, and Griffith (2014) finding gender-match effects for both male and female instructors at a small, selective liberal arts college in the northeastern United States.

To summarize, both GPD effects and gender-match effects in the extant research are mixed and vary depending on context. Business schools are unique in having both quantitative and nonquantitative courses, with neither course type dominating the curriculum (unlike STEM and liberal arts curricula) and a gender ratio close to equal (unlike the Air Force Academy). Furthermore, business schools attract people who want to prepare for the corporate world. All these factors can result in a very different GPD in business education. This paper contributes to the GPD literature by documenting GPD and the impact of instructor gender in the context of business education at a top public university. An important contribution to the literature stems from the novel evidence we present for the mechanism behind GPD.

DATA

We rely on four sources of data. We obtained the first data set from the business school library of a large public university in the U.S. Midwest; this data set contains undergraduate business administration (UBA) program bulletins listing core courses and their timing requirements. The second data set combines three administrative databases obtained from the same university: (anonymized) student grades in all enrolled classes, student background characteristics obtained from their university application, and instructor demographics. The third data set also comes from the business school library and includes most of the syllabi of the fixed core courses between fall 2006 and winter 2017. The fourth data set includes survey responses of 102 junior UBA students currently enrolled in the same business school; this survey provides us with student perceptions.

Using the administrative data sets, we follow the academic performance of 6,312 UBA students who represent the graduating classes of 2005–2018. Because the school admitted two cohorts as it transitioned

from a two-year program to a three-year program in 2006, the data comprise 15 cohorts of students. The data span all introductory fixed core classes (discussed next) taken between fall 2003 and winter 2017. The survey then provides a lens to examine possible reasons for performance differences. We describe each data set in turn.

Program Bulletins: The Structure of the Core Curriculum

The UBA core (i.e., compulsory) curriculum consists of introductory courses in accounting, business law, business economics, business communications, finance, marketing, operations, organizational behavior, statistics, and strategy. The Business School Registrar publishes a bulletin specifying the timing of these core courses. Some of these courses are “floating core,” meaning that students can choose when to take these and thus can select their professor. In contrast, “fixed core” courses must be taken at a specific time in the program (e.g., fall semester of sophomore year).³

Our analyses focus on students’ academic performance in the fixed-core curriculum of the UBA program because the fixed-core program is mandatory and structurally rules out student self-selection into courses and instructors. The exogenous assignment is ensured by the registrar’s office, which divides each cohort into five or six sections and randomizes students into these sections conditional on the female student proportion being the same across sections. Because instructors are assigned to sections and because students remain with their section mates in the fixed-core program, they are not able to choose the instructor teaching the class. We provide empirical evidence for the registrar’s success in ensuring random assignment in Appendix A.

Administrative Data

The administrative data combine three databases: students’ grades in university classes, students’ background characteristics obtained from their university application, and instructor demographics.

Student grades. As in other studies of GPD, we measure academic performance as students’ grades in

³ The set of floating versus fixed courses has changed several times due to restructuring of the core curriculum. The Web Appendix provides further institutional details that guide our sample construction.

each course.⁴ The grades are determined on an A+, A, A-, B+, B, ..., C-, D, F scale, where an A+ is worth 4.4 grade points, an A is worth 4 grade points, an A- is worth 3.7 grade points, a B+ is worth 3.4 grade points, and so forth. We standardize course grades to have a mean of zero and a variance of one within each course, semester, and year. This standardization allows for a direct comparison across courses and with previous studies of GPD. We also calculate the GPA in other courses (GPAO), which is the cumulative GPA for a student calculated across all semesters, including the current semester, excluding only the course being analyzed. Previous literature has found this variable to be helpful in accounting for potential confounding variables that influence student achievement, reducing both systematic and random sources of error (Huberth et al., 2015; Koester, Grom, and McKay, 2016).

Individual student background characteristics at the time of college application. Approximately 37% of the students are female, 65% are white, 25% are Asian, 3% are black, and 3% are Hispanic. For 65% of the students, English is the primary language spoken at home. The households students come from vary in education and income levels. For example, while 10% of students are first-generation college students, 23% of them have at least one parent with a PhD (in 17% of cases, parental education level is unreported and coded as such). Similarly, while 11% of students come from households with less than \$50,000 income per year, 38% come from households with more than \$150,000 in annual income (in 19% of cases, parental income is unreported and coded as such).

Table 1 breaks down the summary statistics of these characteristics as well as the GPAO variable by student gender and provides a test of differences across genders. We find significant differences across many demographic variables across female and male students. For example, male students are slightly more likely to be white and from a family with at least one parent who has a PhD, whereas female students are more likely to be first-generation college students.

⁴ Previous research has also used grades to quantify GPD. Koester, Grom, and McKay, (2016, p. 2) state that grades are “straight-forward measures... they are natural: they reflect ways in which students themselves might assess performance.”

Table 1: Student Characteristics

	Female Students (N=2349)		Male Students (N=3963)		Male – Female Difference (Total = 6312)	
	Mean	S.D	Mean	S.D	Mean	p-value
<i>Demographics</i>						
Asian	0.29	0.45	0.22	0.42	-0.06***	(0.00)
Black	0.05	0.21	0.03	0.16	-0.02***	(0.00)
Hispanic	0.03	0.18	0.04	0.19	0.00	(0.49)
White	0.60	0.49	0.67	0.47	0.07***	(0.00)
HH Income Unknown	0.20	0.40	0.18	0.38	-0.03**	(0.01)
HH Income 19K-75K	0.10	0.29	0.11	0.32	0.02*	(0.03)
HH Income 75K-150K	0.33	0.47	0.33	0.47	-0.00	(0.74)
HH Income 150K+	0.37	0.48	0.38	0.49	0.01	(0.27)
Prnt. Max. Educ., Unknown	0.18	0.38	0.17	0.37	-0.01	(0.32)
Prnt. Max. Educ., No College	0.11	0.31	0.09	0.29	-0.02*	(0.02)
Prnt. Max. Educ., College	0.22	0.41	0.22	0.42	0.01	(0.49)
Prnt. Max. Educ., Masters	0.29	0.45	0.27	0.44	-0.02	(0.13)
Prnt. Max. Educ., Ph.D.	0.21	0.41	0.25	0.43	0.04***	(0.00)
English is Primary Language	0.66	0.47	0.65	0.48	-0.01	(0.37)
<i>Academic Aptitude Measures</i>						
GPAO	3.49	0.41	3.54	0.41	0.04***	(0.00)
High-school GPA	3.85	0.17	3.80	0.20	-0.05***	(0.00)
Took Calculus in High-school	0.69	0.46	0.70	0.46	0.01	(0.27)
ACT English Test Score	30.82	3.66	30.39	3.65	-0.43***	(0.00)
ACT Math Test Score	30.34	3.28	31.40	3.18	1.06***	(0.00)
ACT Reading Test Score	30.11	4.00	30.08	3.85	-0.03	(0.81)
ACT Science Test Score	28.44	3.72	29.79	3.68	1.35***	(0.00)
ACT Total Score	30.08	2.85	30.58	2.77	0.50***	(0.00)
SAT Math Test Score	698.68	62.08	716.10	59.38	17.42***	(0.00)
SAT Verbal Test Score	642.60	74.26	646.38	70.03	3.78	(0.15)
SAT Total Score	1341.28	109.54	1362.48	105.41	21.20***	(0.00)
AP Calculus Test Score	4.03	1.16	4.18	1.07	0.15**	(0.01)
AP English Lit. Test Score	3.80	0.82	3.56	0.89	-0.24***	(0.00)
AP English Test Score	4.01	0.82	4.03	0.83	0.02	(0.65)
AP MacroEcon Test Score	3.95	0.99	4.15	0.91	0.21***	(0.00)
AP MicroEcon Test Score	4.01	0.85	4.24	0.82	0.23***	(0.00)
AP Statistics Test Score	4.07	0.88	4.27	0.80	0.20***	(0.00)

Notes: Gender was recorded in the administrative data as a binary variable.

The data also provide several measures of academic performance before students joined the university (e.g., SAT component scores, advanced placement [AP] subject test scores, high school GPA). The application process allows for SAT and/or ACT test scores. We have SAT exam scores for 53% of the UBA students and ACT exam scores for 71% of the students. In total, 24% of the students had taken both

exams; less than 0.5% of students did not report either exam score.

Consistent with the literature on performance differences in high school, we note significant differences across male and female students in prior academic achievements by subject. On average, female students had a higher high school GPA and a higher language proficiency, as indicated by their ACT English and AP English literature scores. However, male students, on average, had a higher proficiency in more quantitative subjects, as indicated by their ACT math, ACT science, SAT math, and AP calculus, microeconomics, macroeconomics, and statistics test scores.

Instructor characteristics. For each class in which a student enrolls in the undergraduate business program, we have information on each instructor's name, gender, ethnicity, type of appointment, and teaching experience. Table 2 summarizes these data by gender. Over our period of study, the average instructor taught for a little over five terms in the fixed-core program. Of the instructors teaching fixed-core classes, 34% are female.

Table 2: Faculty Characteristics

	Female Instructors (N=60)		Male Instructors (N=115)		Male - Female Difference (Total = 175)	
	Mean	S.D.	Mean	S.D.	Mean	p-value
Tenure-track or Tenured	0.23	0.43	0.30	0.46	0.07	(0.31)
Not tenure-track	0.28	0.45	0.26	0.44	-0.02	(0.75)
Graduate student	0.48	0.50	0.43	0.50	-0.05	(0.54)
Number of terms taught in core	5.27	7.08	4.17	4.99	-1.10	(0.29)
U.S. citizen	0.58	0.50	0.50	0.50	-0.09	(0.27)
White	0.48	0.50	0.62	0.49	0.13	(0.09)
Asian	0.37	0.49	0.34	0.48	-0.03	(0.72)
Black	0.10	0.30	0.03	0.18	-0.07	(0.13)

Female instructors are marginally less likely to be white. We do not find other differences across instructor genders. However, representation of instructor demographics varies greatly across subjects, and a lack of correlation in the aggregate should not be taken as a lack of correlation within a subject area. An important feature of the data that may not be apparent from the aggregate statistics is the lopsided distribution of female instructor across subjects. Only 17% of students are taught business economics by

female instructors. In contrast, 43%, 31%, 68%, and 85% of students are taught marketing, management and organizations, business law, and strategy, respectively, by female instructors. Our empirical analyses that investigate the impact of instructor gender also control for the unequal chances of students being taught by a female professor across different courses, because the discrepancies across courses in female instructor representation may be correlated with disparities in GPD, leading to an aggregation bias in the estimates of interest (impact of instructor gender on student performance). We may expect such a correlation because academic success in undergraduate programs in a field is a strong predictor of pursuing graduate work in that field (Sax, 2001).

Syllabi Data

We obtained the syllabi for 278 course–instructor combinations out of the 308 in our sample from fall 2006 to winter 2017.⁵ We coded each syllabus to indicate whether and what percent the following components contributed to a student’s grade in the course: class participation; individual in-class assignments, exams, and quizzes; individual take-home assignments, exams, and quizzes; and group take-home assignment and projects. The syllabi data confirm that all classes offered in the same term (by different professors) have the same graded components, the same distribution of points across those components, and the same grading rules, due to the coordinated nature of the UBA core program.

Table 3 reports the average percentage of the grade each of these components accounts for in the UBA core across the 278 syllabi. As we expected, the largest part of the grade is determined by individually completed in-class exams, such as midterms (28%) and finals (31%). These components are followed closely by group term projects (15%), take-home exams (10%), and class participation (10%). While most variation in the weight of these components are across courses, there is also within-course variation over terms. We use the syllabi data to test whether our results regarding the impact of instructor gender vary by the weight of different grading components in a course.

⁵ For the remaining 30 syllabi covering 19 course–term combinations, we conducted a robustness check of results by imputing missing terms based on data in adjacent terms. The results are robust.

Table 3: Syllabi, Graded Components

	Mean	S.D.	Min	Max
<i>Individual-based Components</i>				
Class participation	.103	.058	0	.25
Assignments during term	.043	.104	0	.55
In-class exam during term	.281	.123	0	.67
In-class final exam	.308	.113	0	.6
Take-home exam during term	.106	.112	0	.5
Final project	.003	.038	0	.45
<i>Group-based Components</i>				
Assignments during term	.150	.155	0	.5
Final project	.003	.031	0	.3

Survey of UBA Students

We conducted a survey with 102 junior UBA students. These students were in their last semester of taking introductory core classes, and were enrolled in a mandatory core class that required them to complete research studies for course credit. First, we asked them to indicate the core classes they took or were currently taking and their career interests. Then, we asked the students to select the professor they had for each core class from a drop-down menu. In the next screen, we asked them to think back on each of the core classes and rate their excitement/interest, their initial probability of getting an A in the course (to measure initial performance expectations), and their effort in each course. All three items were assessed on a 0–100 scale, with larger numbers indicating greater interest, a higher probability of getting an A, and more effort. On the next screen, for each course, we asked the students to rate how they felt they were treated by the professor and to what extent the professor was a source of inspiration or a role model (1–7 Likert scale, with larger numbers corresponding to more positive sentiments). These questions were followed by an incentive-compatible elicitation of their beliefs about the average grade difference between male and female students (GPD beliefs). We chose to do this for 8 of the 11 core classes due to time constraints of the survey. The response range was set between –1.1. and 1.1, with 0 indicating no performance difference. Positive (negative) values indicate that the respondents believed men (women) perform better than women (men) in a given course. For example, 0.5 meant that the respondent thought

that the average grade of men is 0.5 grade points higher than that of women in that course. Finally, we asked the students to rate their perception of the quantitateness of each course (on a 0–100 scale, where 0 is not quantitative at all and 100 is extremely quantitative). At the end of the survey, the respondents indicated their ethnicity and gender. The program randomly selected one of the eight GPD belief elicitation questions, and respondents received \$5 for guessing within 0.005, \$2 for guessing within 0.1, and \$1 for guessing within 0.2 of the performance differences in grades in the core classes offered in the 2013–2014 academic year. Further details about the survey appear in the Web Appendix.

Table 4: Survey Average Responses

	Quantitativeness perception	GPD beliefs	Interest in course	Initial prob (A)	Effort in course	Treated fairly	Role model
Accounting 1	75.51	.10	58.24	62.27	73.56	5.59	3.54
Accounting 2	73.34		52.45	62	72.39	5.98	4.11
Bus. Economics	64.92	.09	48.22	67.08	68.92	5.31	3.22
Bus. Law	26.12		73.45	63.58	70.07	5.87	4.78
Finance	82.79	.27	65.64	58.35	66.23	5.65	3.59
Marketing	24.35	-.22	59.42	72.26	73.15	5.86	3.56
Org. Behavior	22.72	-.20	50.18	65.10	59.56	6.03	4.33
Strategy	34.98		60.71	63.81	65.44	5.34	3.44
Bus. Inf. Sys.	53.64	.02	50.12	66.01	63.75	5.35	3.96
Statistics	81.97	.10	57.85	65.38	70.61	5.86	4.22
Operations	77.88	.11	55.47	60.33	68.03	5.68	3.96

Table 4 reports the average responses. The differences across courses in terms of student interest, perceived initial probability of getting an A, and student effort are not large. Students felt that they were mostly treated fairly and had overall positive sentiments toward the professors in the core courses. However, the average responses reveal larger differences across core courses in terms of student perceptions of course quantitateness. Students reported the following seven courses to be mostly quantitative in nature: finance (82.79), statistics (81.97), operations (77.88), accounting (level 1: 75.51 and level 2: 73.34), business economics (64.96), and business information systems (53.64). They rated the following four courses to be mostly nonquantitative in nature: strategy (34.98), business law (26.12), marketing (24.35), and organizational behavior (22.72). Therefore, in some of our analyses, we will be

referring to these groups of courses as quantitative (or perceived to be quantitative) and nonquantitative (or perceived to be nonquantitative) courses, respectively.

We also observe that students' GPD beliefs are positive for quantitative courses and negative for nonquantitative courses, suggesting that students expect men (women) to perform better in quantitative (nonquantitative) courses. We examine these patterns and respondents' reasons for their assumptions in our discussion of the *actual* GPD estimates obtained from the administrative data in the next section

DOCUMENTING GENDER PERFORMANCE DIFFERENCES

In this section, we document systematic grade disparities across female and male students in the introductory courses in the fixed-core business curriculum. To quantify the gender performance differences after controlling for differences in other demographics and academic backgrounds, we estimate the following:

$$\text{Grade}_{\text{spt}} = \alpha + \psi_c 1(g_s = F) + \delta_{k(c)} X_{\text{st}} + \eta_{\text{ct}} + \varepsilon_{\text{spt}}, \quad (1)$$

where $\text{Grade}_{\text{spt}}$ is the standardized grade of student s in course c with instructor p in semester-year t , and $1(g_s = F)$ is an indicator for whether the gender (g_s) of student s is female. The coefficient of interest (ψ_c) captures the difference in mean performance between male and female students in course c after we control for their other demographics and their academic backgrounds. We refer to ψ_c estimates as the GPD. Recall that standardized grades have a mean of 0 and a variance of 1 within each course, semester, and year for ease of comparability. Therefore, the ψ estimates reflect the GPD in terms of a percentage of a standard deviation. GPD is defined to be negative when male students perform better than female students and positive when female students perform better than male students.

The vector X_{st} includes all student characteristics noted in Table 1. These include demographics such as ethnicity, whether English is the student's native language, household income, maximum parental education level, previous academic aptitude variables (e.g., high school GPA; a high school calculus

indicator; and their ACT, SAT, and AP test scores), and indicators for the availability of these variables.⁶ Furthermore, we allow SAT and ACT scores to have different coefficients for the group of students who took both exams, so that the incremental impact of the SAT score for a student who also took the ACT is captured separately from the impact of the SAT score for a student who took only the SAT. As a control for general academic performance as a university student, we also control for the student's cumulative GPA at the university by term t and excluding course c , which varies by term, necessitating the term (t) subscript in X_{st} . These controls enable us to compare the academic performance of students in the fixed-core curriculum of the undergraduate business program who had similar demographics and academic aptitudes but differ in gender. We recognize that X_{st} may predict academic success in each subject k differently. For example, a student's SAT math score may be a better predictor of his or her grade in the finance core class, while the SAT verbal score may be a better predictor of the student's grade in the business law course. Coefficient $\delta_{k(c)}$ allows the impact of X_{st} to vary by the subject k of course c . The specification also includes course-term fixed effects, η_{ct} . Robust standard errors are clustered at the course–instructor level.

Table 5 presents the results. The estimated GPD coefficient is the most negative (female students lagging relative to comparable male students) for finance and is most positive (male students lagging relative to comparable female students) for management and organizations. The magnitudes of GPD are substantial for these courses; women's (men's) grades in finance (management and organizations) are 34% of a standard deviation lower than the grades of men (women) with comparable academic and demographic backgrounds. Finance is followed closely by business economics ($\psi = -.23$), accounting ($\psi = -.14$ and $-.07$), and statistics ($\psi = -.09$) in having negative GPD coefficients. Management and organizations is followed closely by marketing ($\psi = .21$) and business law ($\psi = .18$) in having positive GPD coefficients.

⁶ To keep all students in the regression sample, we include indicators for whether each of the academic aptitude measures were available. We do not have information for 540 students' high school GPAs, possibly because these students came from nontraditional high school programs. There are students who took only the ACT (2,897 students), those who took only the SAT (1,835 students), those who took both (1,525 students), and 55 students for whom we do not have either test score. We identify each of these groups with separate indicator variables. Finally, we allow for an indicator of whether the student took a particular AP exam, along with a dummy variable for each of the possible five performance levels if he or she took it.

In the introductory core courses in operations, business information systems, and strategy, we do not find a significant GPD.

Table 5: Male - Female Gender Performance Gap in Fixed-Core Classes

	(1)	(2)
Female Student x Quantitative		-.110*** (.000)
Female Student x Non-quantitative		-.232*** (.000)
Female Student x Finance	-.337*** (0.000)	
Female Student x Business Economics	-.226*** (.000)	
Female Student x Accounting 1	-.144*** (.000)	
Female Student x Statistics	-.094*** (.000)	
Female Student x Accounting 2	-.071* (.078)	
Female Student x Operations	-.014 (.463)	
Female Student x Business Information Sys.	-.002 (.918)	
Female Student x Strategy	-.039 (.589)	
Female Student x Business Law	.182*** (.000)	
Female Student x Marketing	.207*** (.000)	
Female Student x Organizational Behavior	.335*** (.000)	
	48,511	48,511

Notes: The dependent variable is the individual student's normalized grade. Control variables are course by semester fixed effects and student-term control variables X_{st} interacted with each subject. Robust standard errors are clustered at the course-instructor level. p-values in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

It may be apparent that there is a relationship between the estimated GPD coefficients and the average student beliefs about GPD elicited by the survey. A formal rank-test confirms that the GPD rank of courses as suggested by students' beliefs is congruent with the GPD rank of courses based on our results (Spearman's $\rho = -.78, p = .022$). In addition, GPD in a course is related to its quantitiveness. We find

a significant rank correlation between mean perceived quantitateness of a course from the survey results and the estimated GPD coefficient in these regressions (Spearman's $\rho = -.81, p = .003$).

To summarize, and for ease of communication, we repeat the analysis with the binary grouping of courses based on perceived quantitateness results from our survey. Recall that based on the survey results, accounting, business economics, finance, business information, and operations courses are categorized as quantitative, while marketing, business law, strategy and management and organizations courses are categorized as nonquantitative. Keeping the control variables unchanged, we estimate ψ for quantitative and nonquantitative courses with the following regression:

$$\text{Grade}_{\text{scept}} = \alpha + \psi_{\text{q(c)}} \mathbf{1}(g_s = F) + \delta_{\text{k(c)}} X_{\text{st}} + \eta_{\text{ct}} + \varepsilon_{\text{scept}} \quad (2)$$

where subscript q(c) indicates which of the quantitative/nonquantitative binary classification course c belongs to. The rest of the specification remains the same.

Column (2) of Table 5 presents the results. On average, in quantitative courses, female students lag behind male students. On average, they score 11% of a standard deviation lower than male students who are academically and demographically similar. In contrast, in nonquantitative courses, male students lag behind female students. In these courses, on average, female students earn higher grades than comparable male students by about 22% of a standard deviation.

The magnitude of the discrepancies is substantial. For comparison, prior studies focusing on STEM coursework at the college level report average GPDs of -15% of a standard deviation in STEM course grades at the Air Force Academy (Carrell et al., 2010), and -10% difference in absolute letter grades for STEM courses (Koester et al., 2016). In most STEM programs, women are in the minority. Given that business schools have paid a lot of attention to equity and representation in their programs, the differences are all the more interesting and beg the question of what may be driving them.

POTENTIAL DRIVERS OF GENDER PERFORMANCE DIFFERENCES

Why are gender performance differences occurring? Can educational institutions attenuate them? In this section, we explore three hypotheses for GPD. Two hinge on gender stereotypes: GPD may arise from gender stereotypes in two ways—gender stereotypes held by students (we call this “stereotype bias”) or gender stereotypes held by instructors (we call this “instructor bias”). There is much literature to support these first two stereotype hypotheses. The third hypothesis we propose is related to innate differences in interest across genders; this hypothesis is more exploratory in nature.

Stereotype Bias (Student-Based)

Quantitative courses tend to rely more on math skills and are typically considered a male-stereotyped domain, whereas nonquantitative subjects tend to rely more on communication skills and are typically considered a female-stereotyped domain (Fennema and Sherman 1977; Hyde et al. 1990). Stereotypical beliefs can hamper academic performance through “stereotype threat,” or the idea that a person’s actual performance may suffer when a negative performance stereotype connected to his or her identity is evoked (Steele 1997; Steele and Aronson 1995). Performance deterioration after evocation of a negative stereotype has been demonstrated in the context of math test performance and gender identity (e.g., Cadinu et al. 2005; Spencer, Steele and Quinn 1999). For example, when the Asian identity of Asian women was evoked before a math test, their performance was better (compared with the control); however, when their female identity was evoked, they performed worse (Shih, Pittinsky, and Amabady 1999).

Stereotype threat may hamper performance through competency and self-efficacy beliefs (e.g., Bordalo et al. 2019; Bouchey and Harter 2005), thus affecting motivation, the selection of activities, and focus (Bandura, 1977; Bussey and Bandura, 1999). In the university context, such beliefs can influence students’ beliefs about the career they are best suited for and their likelihood of success in that field, thus affecting their interest and motivation to do well in courses related to that career path. In the context of business school, female (male) students may be less (more) interested in and less (more) motivated to perform well in the finance core course than in the marketing core course due to their beliefs about their eventual success in that

field. Given that students juggle several courses and activities in a single term, male students may underperform in the marketing core class, and female students may underperform in the finance core class relative to what their academic aptitude would predict because of their differences in motivation and interest.

It has been proposed that salient examples contradicting educational stereotypes can be powerful in changing gendered beliefs (Solanki and Xu, 2018). These examples can help nonstereotypical students shape and maintain an identity related to that field (Gilmartin et al. 2007; Oyserman, 2007), become confident that a future in that field is attainable, and increase their interest in the field. In particular, having a female instructor in quantitative courses in which GPD is negative can change the stereotype by providing a powerful counterexample (Spencer, Steele, and Quinn 1999). This argument can also hold for male instructors in a nonquantitative courses in which GPD is positive.

In the next section, we test instructor gender as an intervention that challenges gendered beliefs. If stereotypes are the reason that students have diverging interests and expectations about success, and if instructor gender influences those stereotypes, grades and the interests of female (male) students should increase in quantitative (nonquantitative) courses when female (male) instructors teach them. We expect a stronger effect of instructor gender on GPD in quantitative courses compared with the effect of instructor gender on GPD in nonquantitative courses because female instructors in quantitative courses are rarer and thus more likely to change gendered beliefs. Therefore, we expect the interest and grades of female students to increase in quantitative courses when female versus male instructors teach those courses. If instructor gender indeed changes stereotype beliefs by providing a salient counterexample to the stereotype, we also expect female students to be more likely to rate female instructors as inspirational or good role models compared with male students in quantitative courses.

Instructor Bias

Instructors of different genders can also hold different beliefs about male and female ability, and these gendered beliefs may drive instructor behaviors, expectations, and evaluations, thus indirectly impacting

students' academic success (Lavy and Sand 2015; Leinhardt, Seewald and Engel 1979). Research has shown that male and female instructors may differ in their perception and treatment of male versus female students (Krieg, 2005; Rodriguez, 2002; Stake and Katz, 1982).

To elaborate in the context of our study, if instructors hold beliefs that female students will perform worse than male students in quantitative courses and that male students will perform worse than female students in nonquantitative courses, the instructor's behavior may help realize these beliefs. For example, the instructor may give different levels of homework to male versus female students, thus facilitating their subject proficiency differently. The instructor may also grade male students and female students differently in subjective grade components or call more on male students than on female students in class. While all students in fixed-core classes get the same assignments due to the coordinated nature of the program, class participation and other subjectively graded components may influence grades in business school classes.

In the next section, we provide two empirical tests for the possibility of instructor bias. First, using the survey data, we test whether the way students feel treated by their professor varies by the student's and the professor's gender in quantitative and nonquantitative courses. Second, using data from course syllabi, we test if the impact of the interaction between instructor and student gender on the student's grade varies with the importance of subjective performance evaluations and class participation.

Innate Preference Differences Across Genders

One may also consider innate preference differences between men and women that can drive interest and motivation in different academic subjects and subsequently affect academic performance. A stream of literature in psychology claims that women are more people-oriented and men are more thing-oriented and that this dichotomy explains both college majors and vocational preferences (e.g., Lippa 1998, 2010; Su, Rounds, and Armstrong 2009). A difference between genders is also evident in educational and occupational choices (Xie and Shaumann, 2003; Zafar, 2013). Zafar (2013) suggests that men care more about money than women and thus pursue more lucrative careers. If there are innate preference differences across genders, GPD

may simply be a result of students optimally allocating time and effort into courses based on their interests.

If innate student preferences alone can explain GPD, GPD may be a substantial problem. After all, if male students inherently prefer quantitative subjects more than female students do, and if female students inherently enjoy and prefer nonquantitative courses and/or careers, what does it matter if these preferences are reflected in associated GPD? If this is the case, it would be an open question as to whether interventions are needed to change these preferences.

Note that if innate preferences drive GPD, there is no reason to believe that the instructor's gender would change these innate preferences for some courses and not for other courses. More specifically, if innate preferences drive GPD, there is no reason to believe that female instructors would increase the interest of female (but not male) students in quantitative (and not in nonquantitative) courses.

EVIDENCE FOR DRIVERS OF GENDER PERFORMANCE DIFFERENCES

We use survey, administrative, and syllabi data to assess which of the hypotheses are most likely to be the primary drivers of the GPD differences seen in the data.

Student Effort, Performance Expectations, Interest, and GPD Expectations Across Courses

Recall that we asked students in the survey to think back to the beginning of each of their core courses and evaluate how likely they thought it would be for them to get an A in that course, how interested/excited they were about the course, and how much effort they put into the course. We also asked them about their GPD expectations, which we report in Table 4.

Effort. We do not find differences in how women and men expected to allocate effort across types of courses. However, women reported more effort overall (men = 64.01, women = 75.33; $p < .01$).

Performance expectations. When assessing their own performance capacity, women reported lower expectations of getting an A in quantitative courses compared with nonquantitative courses (71.72 vs. 58.69, $p < .01$), whereas men were equally confident across course types (64.04 vs. 64.28, $p > .8$). These differences support the conjecture that female students' expectations about their performance competency vary across

stereotypically male and female subjects.

Interest. Women reported being more interested in nonquantitative courses than quantitative courses (64.18 vs. 54.83, $p < .01$); in contrast, men reported being more interested in quantitative courses than nonquantitative courses (56.23 vs. 52.22, $p = .042$). Interest differences can be driven by gender stereotypes or by innate preferences. Recall that the survey asked about students' career interests. We find that 64% of male students versus 48% of female students in the survey indicated an interest in pursuing a career in finance or consulting, the two highest-paying jobs after a UBA program. The administrative data also corroborates gendered differences in career paths: the percentage of female students among those who took three or more electives in that subject is 17% in finance, 34% in accounting, and 60% in marketing (the top three subjects of interest in electives). Taken together, these differences provide credence to the idea that student interests vary by gender. However, it is unclear whether the interest differences are driven by stereotypes or innate preferences.

GPD expectations. Recall that the students' GPD expectations in the survey were in line with our findings for the actual GPD at the business school we studied. The survey also asked students to explain the reasons behind their GPD guesses. Interestingly, students' lay theories correspond to theories put forth in the literature. Approximately 36% of the students made statements in line with gender stereotypes: "I also think girls are more creative and better at soft skills so those classes favor them," and "I think that gender stereotypes in regard to numbers and the finance field is what made me choose certain answers. I think it varies across classes for that exact reason. I feel that because of these stereotypes and them being so prevalently spoken about at the business school that girls would probably do better in marketing and boys in statistics and finance."

In addition, 38% of students mentioned gendered differences in course interest driven by differences in career interests. As one student stated, "the variation across classes could be because of the differences in career interests between males and females and the effort they put in [sic:courses] because of these interests." As another said, "I think men tend to concentrate more in areas such as finance and consulting, and put more effort into classes related to those fields." A total of 13% of students mentioned lack of representation (e.g.,

low number of women in finance, of men in human resources) as a reason they expected women (men) to outperform men (women) in nonquantitative (quantitative) courses, for example stating “male to female professor ratio” or “finance is a male-dominated field” as a support for expecting GPD in certain courses. Four students reported the importance of feeling connected to the professor and the professor having an impact on female students’ performance and participation in the course. For example, one student stated, “I have noticed that some of my female classmates are more willing to participate/engage with the material depending on the type of professor. Therefore, in the courses that I believe females do better than males, it is because of the professors’ ability to make everyone more engaged and feel comfortable.” Another student referred to the gender of the professor as an important factor in explaining her GPD guesses: “Female professors can connect to the females in the class better, leading them to do better.”

Testing for Stereotype Bias (Student-Based)

Given the possibility that gendered stereotypes may be contributing to the GPD differences we find, we test instructor gender as an intervention that challenges gendered beliefs. Recall that we expect our test to be particularly strong for female instructors in quantitative courses. We argue that gender of the instructor in quantitative courses is likely to be a stronger manipulator of gender stereotypes because female instructors in quantitative courses are rarer.

We present three pieces of evidence in support of stereotype bias (student-based) driving GPD. First, using the administrative data, we provide evidence for the impact of instructor gender on GPD. Second, using the survey data, we provide evidence for the impact of instructor gender on student interest. Third, using the survey data, we provide evidence for the impact of instructor gender on the extent to which instructors are viewed as a role model.

Impact of instructor gender on GPD. In documenting the impact of instructor gender on GPD, we simultaneously address identification challenges that arise from students’ self-selection into courses or instructors and from the aggregation bias stemming from unequal instructor gender representation across courses that may have confounded some prior work in this area (e.g., Carrell et al., 2010; Griffith, 2014). To estimate the causal impact of instructor gender on GPD, we rely on the random instructor–student

assignment and estimate the following regression:

$$\begin{aligned} \text{Grade}_{\text{spt}} = & \alpha + \gamma_{\text{q(c)}} 1(\text{g}_s = \text{F}) \cdot 1(\text{g}_p = \text{F}) + \beta_{\text{q(c)}} 1(\text{g}_p = \text{F}) \\ & + \delta_{\text{k(c)}} X_{\text{st}} + \kappa_{\text{g}_s} X_{\text{pt}} + \eta_{\text{ctg}_s} + \varepsilon_{\text{spt}} \end{aligned} \quad (3)$$

where $1(\text{g}_s = \text{F})$ is an indicator that the student is female and $1(\text{g}_p = \text{F})$ is an indicator that the instructor p is female. The coefficient $\beta_{\text{q(c)}}$ captures the average effect of having a female instructor teaching quantitative or nonquantitative courses compared with having a male instructor. We do not have a priori expectations of differences in the teaching effectiveness of male and female instructors. To ensure that the differences we document across instructor genders are not explained by differences in instructor descriptors (e.g., race, experience), we allow for student-gender-specific responsiveness to instructor-term-specific control variables (X_{pt}) by permitting the slope κ_{g_s} to vary with student gender g_s . The control variables X_{pt} include the instructor's cumulative experience in years, the number of terms taught in the UBA core program, an indicator for whether the instructor is white, a U.S. citizenship indicator, and indicators for each possible instructor position type (i.e., tenure-track or tenured, non-tenure-track, and graduate student). Recall that in Table 2, we did not find systematic differences in X_{pt} (except for race) across female and male instructors in the aggregate. However, because instructor demographics and the correlational patterns may vary across courses, it is important to include all controls.

Note that equation (3) is a difference-in-differences specification where the student gender main effect is subsumed in the vector of intercepts for each term-course-student gender combination (η_{ctg_s}). We allow student-gender intercept to vary by course to allow for gender differences in academic performance to vary by course. This variation is important to account for. Johnson (2014) discusses the potential for an aggregation bias arising from unequal exposure of students to different instructor genders across courses/contexts. Across different fields in the business school, we may expect an unequal allocation of instructor gender across courses that is correlated with GPD in the undergraduate program, because undergraduate success in a field paves the way for a Ph.D. and then an academic job in that field. To make inferences that are not confounded by such a correlation, it is necessary to control for the average GPD by

student gender that would be observed regardless of the gender of the instructor teaching the course. In this manner, our empirical specification improves upon the related prior literature (e.g., Carrell et al., 2010; Griffith, 2014).

Given this specification, the coefficients of interest, $\gamma_{q(c)}$, reflect the difference in grades female students achieve compared to male students in the same type of course when they are taught by a female instructor compared with their relative (to male students') performance when they are taught by a male instructor. If female students' relative academic performance in quantitative courses are better with female instructors than it is with male instructors, we expect γ to be positive, closing the negative gap we have documented between female and male student performance in quantitative courses. Since male students lag behind female students and the GPD is positive in non-quantitative courses, a positive γ in non-quantitative courses would suggest that the gap between female and male students when non-quantitative courses are taught by female (male) rather than male (female) instructors is larger (smaller). Therefore, a positive γ in either type of course would suggest that the gender performance gap is decreases in absolute value when the instructor gender matches the gender of the students lagging behind on average (and therefore goes against the gender stereotype).

The estimates of interest appear in the first column of Table 6. We find that in quantitative courses, female instructors have no effect on the grades of male students ($\beta=-.011$, p-value=.559), but they have a differential positive impact on the performance of female students ($\gamma=.077$, p-value=.002). Focusing on the estimated coefficient on the female student–female instructor interaction, we observe that the estimate is of substantial magnitude (7.7% of a standard deviation). Recall that in Table 5, GPD was –11% for quantitative courses. Therefore, having a female instructor teach quantitative courses substantially reduces GPD, closing a majority of the original gap. In addition, this finding is entirely driven by female students doing significantly better when taught by a female instructor and not by a decline in male students' academic achievement, as the estimate of the $\beta_{q(c)=1}$ coefficient (female instructors' main effect on quantitative course grades) is small in magnitude and statistically insignificant. Results at the course level

are presented in Appendix B and are in line with the results in Table 6.

Table 6: Impact of Instructor Gender on Grades, Interest, Performance and Effort

	(1) Standardized Grade	(2) Interest	(3) Performance Expectations	(4) Effort
<i>Quantitative courses</i>				
γ (Female instructor * Female Student)	.077*** (.002)	24.74*** (.000)	11.64** (.031)	-0.59 (.868)
β (Female Instructor)	-.011 (.559)	4.50 (.138)	1.62 (.498)	-1.33 (.493)
<i>Non-quantitative courses</i>				
γ (Female instructor * Female Student)	-.001 (.987)	-7.48* (.077)	-.74 (.872)	3.27 (.534)
β (Female Instructor)	-.011 (.727)	-1.27 (.679)	.76 (.771)	.38 (.927)
Observations	48,498	859	851	864

Notes: Column headings correspond to the dependent variable evaluated by the regression. Control variables in (1): Course by semester by student-gender fixed effects, student-term control variables X_{St} interacted with each subject and instructor-term control variables X_{Pt} interacted by student gender. Control variables in (2)-(4): Course by student-gender fixed effects. Robust standard errors are clustered at the course-instructor level. p-values in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In nonquantitative courses, we do not find any impact of instructor gender, regardless of the gender of the students taking these courses. This finding suggests that male students do not benefit from having a gender-matched instructor in courses in which they lag behind. This null result suggests that male students' underperformance in nonquantitative classes may not be primarily driven by stereotypical beliefs and/or that having a male instructor in nonquantitative classes is not effective in challenging these stereotypes.

How Instructor Gender Effect on GPD Varies by Student's Math Aptitude. We explore whether the instructor gender impact on GPD varies by the initial math skills of students. We run a regression that extends Specification 3 by interacting student math skill group (high, medium, low) and course type (quantitative, nonquantitative) with student gender, instructor gender, and the interaction of student and instructor gender.

Table 7: Impact of Instructor Gender, By Students' Initial Math Aptitude

	Low	Middle	High
<i>Quantitative courses</i>			
γ (Female instructor x Female Student)	.026 (.405)	.131*** (.001)	.073** (.016)
β (Female Instructor)	-.008 (.787)	-.032 (.234)	-.008 (.688)
<i>Non-quantitative courses</i>			
γ (Female instructor x Female Student)	-.026 (.694)	.031 (.525)	-.008 (.904)
β (Female Instructor)	-.008 (.854)	-.003 (.949)	-.023 (.547)
Observations	48,511		

Notes: The regression includes interactions of student's initial aptitude level and course type with student gender, instructor gender, and the interaction of student and instructor gender. The table organizes coefficients from this regression across columns by student's initial aptitude level. The regression also includes the following control variables: course by term by student-gender fixed effects, student-term control variables X_{st} interacted with each subject and instructor-term control variables X_{pt} interacted by student gender. Robust standard errors are clustered at the course-instructor level. p-values in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7 reports the main coefficients of interest by course-type and student-type. We find that female instructors have a positive impact on female students in quantitative courses, but this effect is mostly driven by the grade lift of students with math skills that are in the middle of the distribution, followed by the grade lift of students with top math skills.⁷ As expected, there is no heterogeneity of instructor gender impact based on initial math skills in the GPD in non-quantitative courses.

Taken together and interpreted through the lens of the additional process evidence put forth in this paper, these results suggest that female students who have middle to high aptitude for quantitative subjects are the most disadvantaged by gender stereotypes and benefit the most from role models. Importantly, it is these women that the professional world is missing out on due to the misallocation of talent driven by stereotyped beliefs.

⁷ Carrell et al. (2010) has found that the relative importance of having female faculty is the highest for the female students in the Air Force Academy with the highest initial math skills, as measured by their SAT scores. On the other hand, Griffith (2014) suggests that the impact is the largest for university students who are in the middle of the math skill distribution. Our results are closer to that of Griffith (2014), but also show that the female students with the highest math skills still benefit from having a female professor.

Impact of instructor gender on student interest. Turning to the survey data, we examine the impact of instructor gender on student interest and performance expectations in fixed-core courses. In particular, we estimate the following:

$$Y_{scp} = \alpha + \gamma_{q(c)}1(g_s = F) \cdot 1(g_t = F) + \beta_{q(c)}1(g_t = F) + \eta_{cgs} + \varepsilon_{scp}$$

where $1(g_s = F)$ is an indicator that the student is female and $1(g_t = F)$ is an indicator that the instructor t is female. Note that because we survey one cohort of students who take fixed-core classes at the same time, there is no variation at the term t level that we need to account for. Again, student gender main effect is subsumed in η_{gsc} and robust standard errors are clustered at the course–instructor level.

First, we study the impact of instructor gender on student interest. The coefficients appear in the second column of Table 6. Female students' interest in a quantitative course relative to those of men's is higher ($\gamma = 24.72$, $p\text{-value} < .001$) when the instructor is female, but instructor gender does not substantially influence the gendered interest difference in non-quantitative courses ($\gamma = -7.48$, $p\text{-value} = .077$). The instructor's gender does not influence men's interest in quantitative ($\beta = 4.50$, $p\text{-value} = .138$) or in nonquantitative ($\beta = -1.27$, $p\text{-value} = .679$) courses. These results are in line with the impact of instructor gender on GPD.

The third column of Table 6 reports results from the same specification when the dependent variable is students' initial expectations about performance. We find marginal evidence that female students' performance expectations relative to those of men's in quantitative courses are higher when the instructor is female rather than male ($\gamma = 11.64$, $p\text{-value} = .031$). Instructor gender does not influence men's performance expectations in quantitative ($\beta = 1.62$, $p\text{-value} = .498$) or in nonquantitative courses ($\beta = .76$, $p\text{-value} = .771$). These results are again in line with the impact of instructor gender on GPD. We do not find any significant effect of instructor gender on student effort for any group of students or courses (Column 4, Table 6, all $p\text{-values} > .49$).

In summary, we find that having a female professor challenges gendered beliefs and increases female

students' interest (and to some extent performance expectations) in quantitative courses, in which gendered beliefs hamper female student performance. Taken together with the results pertaining to GPD, these results seem to suggest that the impact of female instructors on GPD in quantitative courses may operate by changing student beliefs and interests. Echoing the null effect of instructor gender on GPD in nonquantitative courses, we also find a null effect of instructor gender on student interest and beliefs in courses in which male students lag behind. If interest is an antecedent to performance, this result would suggest that instructor gender may fail to impact GPD in nonquantitative courses, because having a male instructor teach these courses does not increase male students' interest in them.

Evaluations of the instructor as a role model. Recall that the survey asked students to rate the extent to which they considered their instructor a role model or felt inspired by the instructor on a 1–7 Likert scale. We use the same specification as in Equation 3 to explore the impact of instructor gender on these evaluations, divided by student gender and course type.

Table 8: Impact of Instructor Gender on Perceptions of Role-Models and Treatment

	(1) Instructor as Role Model and Inspiration	(2) How Instructor Treated Student
<i>Quantitative courses</i>		
γ (Female instructor * Female Student)	.911** (.018)	-.178 (.306)
β (Female Instructor)	.533* (.060)	.162 (.311)
<i>Non-quantitative courses</i>		
γ (Female instructor * Female Student)	.099 (.762)	-.111 (.691)
β (Female Instructor)	.032 (.917)	.149 (.523)
Observations	867	851

Notes: Column headings correspond to the dependent variable evaluated by the regression. The regression includes the following control variables: course by term by student-gender fixed effects, student-term control variables X_{st} interacted with each subject and instructor-term control variables X_{pt} interacted by student gender. Robust standard errors are clustered at the course-instructor level. p-values in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0$.

We report the estimates in Table 8. We find that female instructors teaching quantitative courses are seen marginally more inspirational than male instructors teaching these courses ($\beta=.533$, p-value = .060), but the positive inspirational/role-model effect is much larger for female students ($\gamma =.911$, p-value=.018). Consistent with other empirical patterns, instructor gender does affect either student gender group's perceptions in nonquantitative courses. These results provide direct evidence for the conjecture that instructors can be powerful role models by providing successful counterexamples to stereotypes. Taken together with the evidence we present regarding the positive impact of female instructors in closing the GPD in quantitative courses, this finding lends credence to the idea that role models can create meaningful performance changes by challenging stereotypical beliefs.

Instructor Bias as a Possible Explanation for GPD

Instructor bias could be an alternative explanation for the instructor gender effects we document in quantitative courses -- if female instructors have preferences or beliefs that help female students (only) in quantitative courses relative to male instructors. For example, if male instructors think that women cannot perform well in quantitative classes but female instructors do not hold this belief, we could potentially have gendered instructor differences in grading and in student treatment within quantitative courses. We test for potential instructor bias by examining differences in students' perceived treatment as measured in the survey. If instructor bias were palpable, we would expect students to report differences in how they fairly they felt male versus female instructors treated them. Of course, bias is not always noticeable, but still may lead to grade differences. We expect such grade differences to be more prominent in courses where a larger proportion of the grade is subjective. Therefore, we also test for potential instructor bias by examining whether GPD varies with how performance in a course is graded (how subjective the grade is).

Instructor gender and perceived treatment. The survey asked students to rate how fairly they thought the instructor treated them on a 1–7 Likert scale. We use the same specification as in Equation 3 to explore the impact of instructor gender on these evaluations by student gender and course type. The second column of Table 8 reports the results. We find that how students feel treated by their professor does not vary by the

student's and the professor's gender in quantitative or nonquantitative courses (all p-values $>.3$). This null effect is supportive of the conjecture that overt instructor bias is unlikely to be a main driver of results in this context.

Moderation by grading component weights. We turn to the administrative data to test instructor bias in grading. As discussed previously, if instructor bias is a main driver of the impact of instructor gender on GPD in quantitative courses, we would expect this effect to be larger in classes in which the instructors have more discretion over the grades and/or when subjective performance evaluations comprise a larger fraction of a student's grade in the course.

To test this conjecture, we code the syllabi of fixed-core courses. For each class, we denote the percentage of the grade that depends on participation, group assignments, and individual exams or assignments. Almost all individual exams and individual assignments are graded by teaching assistants and/or are standardized (e.g., multiple choice, common rubric); therefore, they are unlikely to suffer or benefit from instructor bias. However, class participation may be influenced by instructor bias.⁸ In addition, group projects are more likely to be graded by instructors; therefore, the potential for bias exists, but all students in a group receive the same grade. Using variation across terms within a course in the weight of class participation and group- and individual-graded components, we test the conjecture that the effect of instructor gender on GPD would be larger in classes in which subjective grading components (class participation and group assignments) represent a larger fraction of a student's grade in the course. In particular, we extend Equation 3 by including interactions of professor gender, student gender, and the percentage of the grade that comes from participation, group assignments, and individual exams/assignments.

Table 9 presents the results. We provide the estimates for quantitative and nonquantitative classes across two columns for ease of exposition. The impact of professor gender on GPD (captured by γ) is not moderated by the weight of the class participation and group components versus the individual component

⁸ We thank an anonymous referee for this suggestion.

of the grade. Therefore, we conclude that the positive impact of female instructors on closing the gender gap in quantitative courses in which female students lag behind is not moderated by how the students are graded in a course. Although we cannot explicitly rule out instructor bias, because we do not observe instructor beliefs, these results do not show the kind of data patterns we would expect if instructor bias were a main driver of our results.

Table 9: Moderation by Weight of Individual, Group, Participation Grade Components

	<i>Quantitative Courses</i>	<i>Non-quantitative Courses</i>
γ_0 (Female instructor * Female Student)	.084** (.041)	.093 (.372)
γ_p (Female inst. * Female St. * Participation %)	.126 (.751)	-.626 (.585)
γ_g (Female inst. * Female St. * Group %)	-.138 (.602)	-.027 (.948)
β_0 (Female Instructor)	.055* (.077)	-.061 (.498)
β_p (Female inst. * Participation %)	-.71* (.032)	.189 (.830)
β_g (Female inst. * Group %)	-.109 (.517)	.076 (.627)
Observations	37,409	

Notes: The regression includes interactions of course type, professor gender, student gender, and the percentage of the grade that comes from participation, group assignments. The columns organize the coefficients by course type. The regression includes the following control variables: course by term by student-gender fixed effects, student-term control variables X_{st} interacted with each subject and instructor-term control variables X_{pt} interacted by student gender. Robust standard errors are clustered at the course-instructor level. p-values in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Innate Preferences as a Possible Explanation for Positive GPD

Our results for instructor gender effects on GPD in quantitative courses are not consistent with innate preferences as the driver of GPD. As we stated previously, if differences in innate preferences are the main drive GPD, then female instructors should not increase interest and improve the grades of female students in these courses. Our results are more consistent with the hypothesis that gender stereotypes hinder female students' academic performance.

In contrast, we get null results for instructor effects in nonquantitative courses. We cannot rule in or

rule out gendered differences in innate preferences driving the male students' underachievement in nonquantitative courses.

CONCLUSION

We document significant academic achievement gaps among otherwise similar men and women in business education. The magnitude of our findings should be of concern to business schools. Women's grades are, on average, 11% of a standard deviation lower in quantitative courses than those of men with similar academic aptitudes and demographics, and men's grades are, on average, 23% of a standard deviation lower in nonquantitative courses than those of comparable women. In the case of women's achievement gap in quantitative courses, we also show that instructors whose identity counters gender stereotypes of performance in these fields can help close these gaps.

These results suggest that business schools should strive to combat gender stereotypes that may be hindering students' achievement. Academic achievement gaps not only create inequity in education, but also may have far-reaching consequences, because they may shape occupational choices (Ost, 2010). Bertrand, Goldin and Katz, (2010) cite gender differences in GPA and the propensity to take finance courses as a potential contributor to the gender gap in earnings among its MBA students in later stages of their careers. In the business school we study, the gender gap in wages were prominent even immediately upon graduation, and directly correlated with career choices. Analyzing the exit survey of the last three cohorts of undergraduate alumni, we find that women earn an average of \$71,186, and that men earn an average of \$75,155 upon graduation (difference of 3,969, p-value<.0001). The top starting salaries are highest for jobs in investment banking (average \$84,600K), consulting (average \$75,000), other finance roles (average \$74,100), operations and supply chain management (average \$68,600) and marketing in the technology sector (\$66,500). Jobs in human resources (\$59,100), general management (\$60,700), accounting (\$60,700), marketing in CPG/retail (\$62,100) and advertising (average \$ \$51,100) pay

significantly less. The top three highest paying tracks are the most popular in terms of percentage of students choosing their first jobs, followed by marketing jobs across different industries (tech, CPG/retail, other). Although female student representation is 41% overall, women are underrepresented among students taking jobs in investment banking (20%) and overrepresented in accounting (65%) and human resources (90%). Women are also overrepresented in marketing positions, but more so for positions in CPG and retail (84%) than for positions in the technology sector that pay more (54%).⁹ In light of early achievement gaps in quantitative courses that are basic stepping stones for an academic path that prepares students for jobs in the investment banking and technology sectors, these results add to the concern that academic achievement gaps may be influencing career-path choices and success down the line.

For both educational purposes in their own right, and to help close equity gaps in the initial placement and promotion between men and women in management jobs, it is important to look for ways to arrest achievement differences in business education. Our findings suggest the power of instructors as role models. In addition, business schools may also be able to reduce these discrepancies by providing other salient exemplars that counter gender stereotypes (e.g., speakers, student organization leaders, alumni). Given that we observe achievement gaps in the introductory curriculum, our findings also highlight the necessity of focusing these efforts on the early years of the UBA program. If providing role models for women in quantitative fields taught in business schools can close academic achievement gaps, this can have long-lasting repercussions on the extent to which early achievement begets interest and future performance in those fields.

Eliminating achievement gaps due to gender stereotypes can also help more accurately align students' abilities and talent with their beliefs and educational choices. Better alignment, in turn, will improve students' career success and satisfaction and therefore may ultimately benefit their employers in several ways. First, better alignment can increase productivity and employee satisfaction. Second, to the extent it also levels the playing field among men and women, it can help recruiters hire a more diverse workforce.

⁹ Please see Web Appendix for further details on this data and our analyses.

Considering both the student-side and the employer-side benefits, increased alignment of talent and educational choices can benefit business schools as they race to attract high-quality and diverse applicants, as well as top-paying recruiters who are looking for quantitatively qualified female applicants.

The potential misallocation of talent to careers due to stereotypical beliefs is particularly important to consider for the field of marketing as it becomes more quantitative and data-driven. Top business schools are introducing multiple “analytics” courses and specializations, and many of these courses are offered by marketing departments. To get desirable and lucrative jobs within the field of marketing and advance in the field, these courses are becoming increasingly important to take. Our results point to the importance of considering the representation of female faculty in delivering these courses to help allocate the right talent to the field, to benefit students, and to benefit future employers looking for a diverse and talented workforce.

Our results also underscore the importance of reducing gender gaps when hiring new faculty and highlight that special attention should be paid to this issue in more quantitative disciplines. We also recommend that faculty teaching assignments be linked to gender representation needs across courses, beyond the typical practice of “which new hire is willing and able to teach which elective.” Our findings indicate that differences in quantitiveness require different assignment practices—not because of the capability of the instructor in teaching the course but because of the impact on student performance. The good news is that a change in faculty gender representation in early coursework in UBA programs can be easily realized when hiring new faculty. Interestingly, this may also result in a more gender-balanced faculty in the quantitative disciplines – with more undergraduates moving to do Ph.Ds. in quantitative fields -- whereas, currently the faculty gender ratio is heavily skewed in favor of men at most schools. As such, our results bear many implications for the administration of business schools.

We hope that our results prove useful to business school administrators, faculty, and students. We also hope that our work inspires further research on the determinants of and solutions to academic achievement disparities across genders in higher education. After all, our economy is only as strong as the talent that fuels it.

REFERENCES

Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik (2014), "The effect of teacher gender on student achievement in primary school." *Journal of Labor Economics*, 33, p. 63-89.

Bandura, Albert (1977), "Self-efficacy: toward a unifying theory of behavioral change." *Psychological Review*, 84, 191.

Bertrand, Marianne, Claudia Goldin, and Lawrence F Katz (2010), "Dynamics of the gender gap for young professionals in the financial and corporate sectors." *American Economic Journal: Applied Economics*, 2, 228-55.

Blau, Francine D and Lawrence M Kahn (2017), "The gender wage gap: Extent, trends, and explanations." *Journal of Economic Literature*, 55, 789-865.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer (2019), "Beliefs about gender." *American Economic Review*, 109, 739-73.

Bouchev, Heather A and Susan Harter (2005), "Reflected appraisals, academic self-perceptions, and math/science performance during early adolescence." *Journal of Educational Psychology*, 97, 673.

Bureau of Labor Statistics (2016) "Highlights of Women's Earnings in 2015" Report (#1064), The U.S. Department of Labor.

Bussey, Kay and Albert Bandura (1999), "Social cognitive theory of gender development and differentiation." *Psychological Review*, 106, 676.

Cadinu, Mara, Anne Maass, Alessandra Rosabianca, and Jeff Kiesner (2005), "Why do women underperform under stereotype threat? Evidence for the role of negative thinking." *Psychological Science*, 16, 572-578.

Carrell, Scott E, Marianne E Page, and James E West (2010), "Sex and science: How professor gender perpetuates the gender gap." *The Quarterly Journal of Economics*, 125, 1101-1144.

Cao, Sissi (2017), "A Data-Backed Look Into Silicon Valley's Gender Equality Problem," *Observer*, November.

Ceci, Stephen J, Donna K Ginther, Shulamit Kahn, and Wendy M Williams (2014), "Women in academic science: A changing landscape." *Psychological Science in the Public Interest*, 15, 75-141.

Dahl, Darren W, Jaideep Sengupta, and Kathleen D Vohs (2009), "Sex in advertising: Gender differences and the role of relationship commitment." *Journal of Consumer Research*, 36, 215-231.

Edelman, Karen, Anya George Tharakan, Blythe Hurley, Aparna Prusty, and Nairita Gangopadhyay (2019). "Within reach? Achieving gender equity in financial services leadership" *Women in Finance*, Deloitte Insights.

Fennema, Elizabeth and Julia Sherman (1977), "Sexual stereotyping and mathematics learning." *The Arithmetic Teacher*, 24, 369-372.

Fisher, Robert J and Laurette Dube (2005), "Gender differences in responses to emotional advertising: A social desirability perspective." *Journal of Consumer Research*, 31, 850-858.

Fryer Jr, Roland G and Steven D Levitt (2010), "An empirical analysis of the gender gap in mathematics." *American Economic Journal: Applied Economics*, 2, 2, 210-240.

Gao, Huachao, Vikas Mittal, Yinlong Zhang (2019), "The Differential Effect of Local–Global Identity Among Males and Females: The Case of Price Sensitivity." *Journal of Marketing Research*, 57, 173-191.

Gilmartin, Shannon, Nida Denson, Erika Li, Alyssa Bryant, and Pamela Aschbacher (2007), "Gender ratios in high school science departments: The effect of percent female faculty on multiple dimensions of students' science identities." *Journal of Research in Science Teaching*, 44, 980-1009.

Gong, Jie, Yi Lu, and Hong Song (2018), "The effect of teacher gender on students' academic and noncognitive outcomes." *Journal of Labor Economics*, 36, 743-778.

Govind, Rahul, Nitika Garg and Vikas Mittal (2020), "Weather, Affect, and Preference for Hedonic Products: The Moderating Role of Gender." *Journal of Marketing Research*, 57, 717-738.

Griffith, Amanda L (2014), "Faculty gender in the college classroom: Does it matter for achievement and major choice?" *Southern Economic Journal*, 81, 211-231.

Grohmann, Bianca (2009), "Gender dimensions of brand personality." *Journal of Marketing Research*, 46, 105-119.

Hoffman, Florian and Philip Oreopoulos (2009), "A professor like me: The influence of professor gender on college achievement." NBER Working Paper 13182.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow. "The allocation of talent and us economic growth." *Econometrica* 87, no. 5 (2019): 1439-1474.

Huberth, Madeline, Patricia CHen, Jared Tritz, Timothy McKay (2015), "Computer-Tailored Student Support in Introductory Physics." *PLoS ONE*, 10.

Hyde, Janet Shibley, Elizabeth Fennema, Marilyn Ryan, Laurie A Frost, and Carolyn Hopp (1990), "Gender comparisons of mathematics attitudes and affect: A meta-analysis." *Psychology of Women Quarterly*, 14, 299-324.

Johnson, Iryna Y (2014), "Female faculty role models and student outcomes: A caveat about aggregation." *Research in Higher Education*, 55, 686-709.

Kahn, Shulamit and Donna Ginther (2017), "Women and STEM." NBER Working Paper 23525.

Koester, Benjamin, Galina Grom, and Timothy McKay (2016), "Patterns of gendered performance difference in introductory stem courses." arXiv preprint arXiv:1608.07565.

Kolhatkar, Sheela (2017), "The Tech Industry's Gender-Discrimination Problem," *The New Yorker*, November.

Kost, Lauren E, Steven J Pollock, and Noah D Finkelstein (2009), "Characterizing the gender gap in introductory physics." *Physical Review Special Topics-Physics Education Research*, 5.

Krieg, John M (2005), "Student gender and teacher gender: What is the impact on high stakes test scores." *Current Issues in Education*, 8, 1-16.

Lavy, Victor and Edith Sand (2015), "On the origins of gender human capital gaps: Short and long term consequences of teachers' stereotypical biases." NBER Working Paper 20909.

Leinhardt, Gaea, Andrea M Seewald, and Mary Engel (1979), "Learning what's taught: Sex differences in instruction." *Journal of Educational Psychology*, 71, 432.

Lim, Jaegeum and Jonathan Meer (2017), "The impact of teacher-student gender matches random assignment evidence from South Korea." *Journal of Human Resources*, 52, 979-997.

Lippa, Richard (1998), "Gender-related individual differences and the structure of vocational interests: The importance of the people-things dimension." *Journal of Personality and Social Psychology*, 74, 996.

Lippa, Richard A (2010), "Gender differences in personality and interests: When, where, and why?" *Social and Personality Psychology Compass*, 4, 1098-1110.

Lorenzo, Mercedes, Catherine H Crouch, and Eric Mazur (2006), "Reducing the gender gap in the physics classroom." *American Journal of Physics*, 74, 118-122.

Machin, Stephen and Sandra McNally (2005), "Gender and student achievement in English schools." *Oxford Review of Economic Policy*, 21, 357-372.

Machin, Stephen and Tuomas Pekkarinen (2008), "Global sex differences in test score variability." *Science*,

322, 1331-1332.

Marriage, Madison (2019), "Big consultancies struggle to close the gender pay gap" *Financial Times*, January 29.

Meyers-Levy, Joan and Barbara Loken (2015), "Revisiting gender differences: What we know and what lies ahead." *Journal of Consumer Psychology*, 25, 129-149.

Miyake, Akira, Lauren Kost-Smith, Noah Finkelstein, Steven Pollock, Geoffrey Cohen, and Tiffany Ito (2010), "Reducing the gender achievement gap in college science: A classroom study of values affirmation." *Science*, 330, 1234-1237.

Ost, Ben (2010), "The role of peers and grades in determining major persistence in the sciences." *Economics of Education Review*, 29, 923-934.

Oyserman, Daphna, Stephanie A Fryberg, and Nicholas Yoder (2007), "Identity-based motivation and health." *Journal of Personality and Social Psychology*, 93, 1011.

Puhani, Patrick A. (2018), "Do boys benefit from male teachers in elementary school? Evidence from administrative panel data." *Labour Economics*, 51, 340 – 354.

Rodriguez, Nixaliz (2002), "Gender differences in disciplinary approaches." ERIC WP ED468259.

Shih, Margaret, Todd L Pittinsky, and Nalini Ambady (1999), "Stereotype susceptibility: Identity salience and shifts in quantitative performance." *Psychological Science*, 10, 80-83.

Solanki, Sabrina M and Di Xu (2018), "Looking beyond academic performance: The influence of instructor gender on student motivation in stem fields." *American Educational Research Journal*, 55, 801-835.

Spencer, Steven J, Claude M Steele, and Diane M Quinn (1999), "Stereotype threat and women's math performance." *Journal of Experimental Social Psychology*, 35, 4-28.

Stake, Jayne E and Jonathan F Katz (1982), "Teacher-pupil relationships in the elementary school classroom: Teacher-gender and pupil-gender differences." *American Educational Research Journal*, 19, 465-471.

Steele, Claude M (1997), "A threat in the air: How stereotypes shape intellectual identity and performance." *American Psychologist*, 52, 613.

Steele, Claude M and Joshua Aronson (1995), "Stereotype threat and the intellectual test performance of African Americans." *Journal of Personality and Social Psychology*, 69, 797.

Su, Rong, James Rounds, and Patrick Ian Armstrong (2009), "Men and things, women and people: a meta-

analysis of sex differences in interests." *Psychological Bulletin*, 135, 859.

U.S. Department of Education (2018) "Digest of Education Statistics 2017" NCES 2018-070.

Winters, Marcus, Robert Haight, Thomas Swaim, and Katarzyna Pickering (2013), "The effect of same-gender teacher assignment on student achievement in the elementary and secondary grades: Evidence from panel data." *Economics of Education Review*, 34, 69-75.

APPENDIX

Appendix A: Exogenous student-instructor assignment

Student-instructor assignment is exogenous for the fixed-core courses. In addition to relying on the fact that the registrar performs student-instructor assignments exogenously, we also provide several types of empirical evidence for random distribution of student characteristics across female and male instructors teaching a course.

Table A: Demographics of students taught in fixed-core, by Faculty Gender

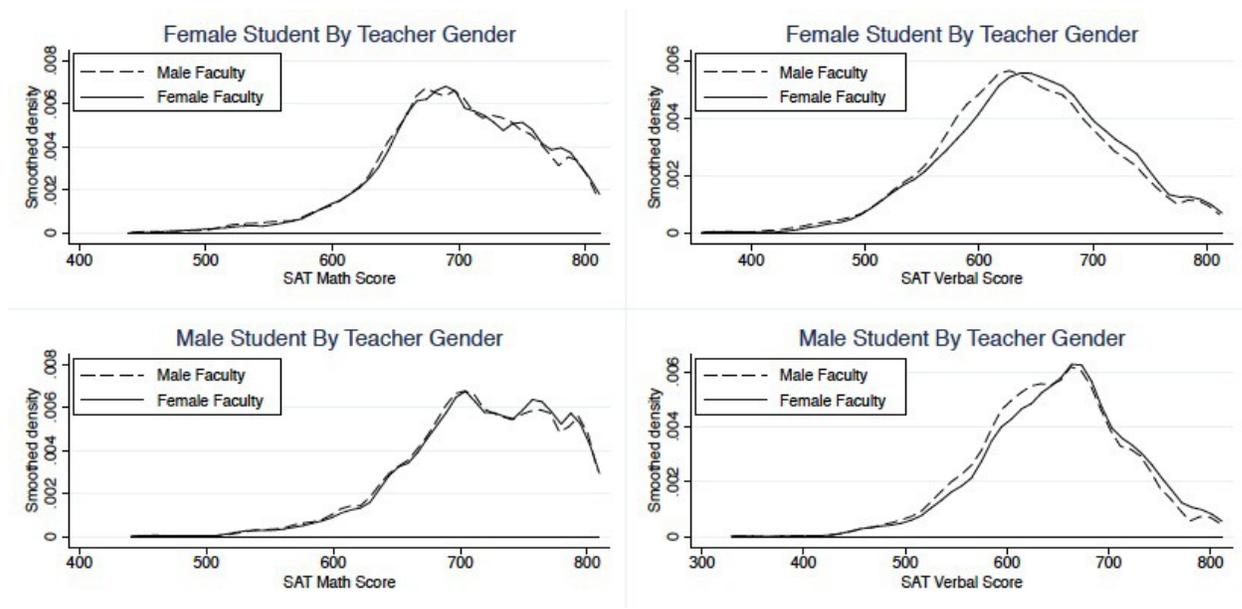
	Female Professor			Male Professor			M-F Diff.
	N	Mean	S.D.	N	Mean	S.D.	
Avg. female student ratio	60	0.36	0.08	115	0.35	0.10	-0.01 (0.57)
Avg. white student ratio	60	0.67	0.07	115	0.65	0.11	-0.02 (0.15)
Avg. hispanic student ratio	60	0.03	0.02	115	0.04	0.04	0.01 (0.16)
Avg. black student ratio	60	0.03	0.02	115	0.04	0.04	0.00 (0.25)
Avg. asian student ratio	60	0.23	0.06	115	0.25	0.12	0.02 (0.15)
Observations	60			115		175	

First, we compute summary statistics of the average student demographics, according to whether students are taught by male versus female faculty, within the fixed-core curriculum. These are given in Table A. In support of the exogenous assignment of instructors and students, we do not see any significant differences across male and female faculty in the type of students they teach.

Figure A plots the distribution of SAT test scores by student gender and instructor gender in the fixed-core curriculum. There does not seem to be any systematic difference in the aptitude of students assigned to

female and male faculty as measured by SAT scores. This evidence suggests that the registrar does a good job of balancing academic aptitude and other demographics in assigning students to sections and faculty, and the small amount of section-switching does not perturb this orthogonality.

Figure A: Distribution of SAT scores broken down by student and instructor gender.



Note: This figure plots the distribution of the SAT Math and SAT Verbal scores of students in the undergraduate business program across the gender of the students and the gender of the instructor they are assigned to.

Second, to provide a formal test of exogenous assignment of instructor gender to students taking a core class in a given term, we regress the gender of faculty member assigned to a student (dependent variable) on the gender of the student, other demographic and academic background variables (e.g., ACT score) listed in Table 1, and course-term fixed effects. The F-statistics from the joint significance test of all explanatory variables provide a test of the null hypothesis that student characteristics are not correlated with instructor gender. This null hypothesis would be rejected if students self-selected into classes based on instructor gender. As expected, the joint test cannot reject the hypothesis that instructor gender is randomly assigned to student background characteristics for fixed-core classes ($F = 1.18, p -$

$value = .205$). In contrast, the joint test strongly rejects the null hypothesis in the floating core courses ($F = 143, 483, p - value < 0.001$). Although floating-core courses are also mandatory for the UBA program, in contrast to the fixed-core courses, they provide more room for students to choose their instructors. In light of these results, we conclude that the fixed-core curriculum has both the appropriate institutional setting and the empirical support for exogenous student-instructor assignment. We therefore focus on fixed-core courses to examine the causal impact of instructor gender on the academic performance gap between male and female students.

Appendix B: Impact of Instructor Gender on GPD, By Course

In Table B, we report the results from an extension of specification (3) that allows for γ to be estimated at the course-level. During the span of 15 cohorts of UBA students, finance and business information systems core classes were each only taught by 1 female instructor. The first accounting class in the accounting sequence was only taught by 2 male instructors. As a result, although we include all observations in the regression, for these courses we are unable to report estimates because the data agreement prohibits publication of effects that could identify any individual.

Table B: Impact of Instructor Gender, By Course

<i>Quantitative courses</i>	Accounting	Bus. Econ.	Operations	Statistics
γ (Female instructor x Female Student)	.104** (.067)	.091*** (.007)	.037 (.545)	.108** (.052)
β (Female Instructor)	-.065* (.089)	-.023 (.491)	-.031 (.338)	.093*** (.002)
<i>Non-quantitative courses</i>	Bus. Law	Marketing	Org. Beh.	Strategy
γ (Female instructor x Female Student)	.014 (.738)	-.043 (.395)	.007 (.903)	-.025 (.857)
β (Female Instructor)	-.016 (.462)	.096* (.078)	-.037 (.380)	.055 (.620)
Observations	48,498			

Notes: The table reports the list of coefficients from a regression that extends specification 3 in the main text by interacting the impact of instructor gender with each course. Control variables: Course by semester by student-gender fixed effects, student-term control variables X_{st} interacted with each subject and instructor- term control variables X_{pt} interacted by student gender. Robust standard errors are clustered at the course-instructor level. p-values in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Overall, we see that in 3 out of 4 quantitative courses for which we can report estimates, the impact of female instructors on female students is positive and significant. The impact of female faculty is never negative on female students in these courses. The impact of instructor gender on student grades is insignificant in each of the non-quantitative courses. These results show that the results reported in the main text regarding the impact of instructor gender on GPD in quantitative and non-quantitative courses are reflective of the effects on the entire set of courses.

WEB APPENDIX

A: Survey

For the survey to explore the reasons for the GPD and instructor effect we find in our data, the survey sample had to be from the same population as the data. As such, the respondents were students in a core class in the same undergraduate business school where we got our data. We collected responses from 102 students, 25 of whom were female. The students completed the survey as a part of a subject pool requirement that they had for a core course. They learned about the nature of the survey after signing up for a particular study session. Although students had the option to refuse responding to the survey, none of them did. The questions were straightforward and measured simple constructs about the courses they had taken/were taking. The survey had to be kept short due to time restrictions and attention checks could not be included, but we can assess attention by looking at internal consistency. Their answers to open-ended “Why do you expect these grade differences?” question were thoughtful and sufficiently long (an average of 150 characters among responders, with only 4 subjects not responding). Although answering the career interest and GPD questions were optional, 92% of respondents completed all of these optional questions (1 person did not report their career interest, 7 people only partially completed the GPD questions). Overall, the respondents seem to be engaged and attentive; and we expect the data to be reliable.

One could argue that there may be a halo effect from female students liking female faculty that spills over to female students rating every aspect of a female faculty’s course as being higher. However, we do not get this halo effect (e.g., effort is not higher if a course is taught by a female versus a male faculty; nor is perceived treatment of student). Do students think of courses more positively if they get a better grade in a course? That is, instead of interest driving the instructor effect that we find, for example, can it merely be a consequence? The lack of a halo effect discussed above also argues against this premise -- female faculty in quantitative courses do not automatically elicit positive responses on all domains, by female students.

We coded the responses to “Why do you expect these grade differences?” questions to separate out responses: (i) Interest: responses that cited differences in interest, (ii) Stereotype: these included responses which were making a stereotypical assessment to explain differences in performance across courses (stereotype) or spoke about the existence of stereotypes in the minds of others”, (iii) Gender representation in careers -- responses that cited differences gender representation in the ultimate careers these classes prepare students for, (iv) Gender representation in business school: responses that cited differences in gender representation at the school level, either in the proportion of students or the proportion of instructors (we also included responses which cited differences within a given field), (v) Effort – responses that used

effort put in to explain grade differences, and (vi) Other -- students who cited other reasons, or offered reasons that did not differentiate between GPD across courses. Many students cited more than one reason. Thus, even with 98 students who provided reasons, 37 students (38%) cited differences in interest, 35 students (36%) cited stereotypical beliefs about abilities, 16 (16%) cited differences in career gender representation, 13 (13%) cited differences in gender representation at school, 9 cited (9%) differences in effort, and 33 cited (34%) other reasons or did not differentiate between courses.

B: Undergraduate Career Office Survey

To provide information on positions taken upon graduation and salaries, the business school conducts an exit survey. We analyzed the data collected from the cohorts graduating in 2017, 2018, and 2019. The response rate (ratio of students reporting their employment status divided by the number of students graduating that year) was 86% in 2019 and 2018 and 89% in 2017, totaling 1,000 students across the three cohorts. These data accounted for 97% of students who were seeking a job upon graduation. For each student, we have information on the one job offer that they ultimately accepted, including job title and function, industry, location, base salary, and signing bonus.

Students indicate job titles. These job titles are grouped by the school into 50 different job functions across 16 industries. For example, the job titles “Global Capital Markets Analyst” and “Investment Banking Analyst” are grouped under the job junction “Finance - Investment Banking” job function; whereas “Assurance Staff” and “Associate” are grouped under the job function “Accounting - Audit”. Other examples of job functions are “Marketing/Sales - Advertising/Public Relations”, “Marketing/Sales - General Marketing” and “Consulting - Strategy Consulting”. To better characterize the nature of each job function, we group them into “tracks” based on the job function and/or the industry in which the firm operates. For example, people taking a job in a marketing function are separated into four different tracks: Advertising, marketing in technology, marketing in retail/consumer packaged goods, and marketing in all other industries. We group job titles into human resources, accounting, consulting, operations, and general management tracks, but distinguish investment banking from other finance jobs. All the remaining job functions are grouped together under the label “Other”. At the end of this process, we are left with 13 different job tracks. For each of the 14 job tracks, Table [WA.1](#) summarizes the average starting salary, the percentage of students placed in that track, and the percentage of female students among those who took a job in that track. The tracks are sorted by percentage of total placements (any gender).

Table WA.1: Base salary and percent of students placed in each track

<i>Track</i>	<i>Base Salary</i>	<i>% students placed</i>	<i>% female among placed</i>
<i>Consulting</i>	74,969	24.0	42.0
<i>Finance - Other</i>	74,089	23.7	41.7
<i>Finance – Inv. Banking</i>	84,612	20.7	20.2
<i>Marketing – Tech.</i>	66,476	7.2	54.1
<i>Marketing -CPG/Retail</i>	62,132	6.8	84.4
<i>Marketing - Other</i>	63,678	4.2	66.7
<i>Operations</i>	68,611	2.5	44.0
<i>Accounting</i>	60,652	2.3	65.2
<i>General Management</i>	60,733	1.5	46.7
<i>Human Resources</i>	59,100	1.0	90
<i>Marketing - Advertising</i>	51,142	0.7	28.5
<i>Other</i>	69,737	5.7	33.3