1 Introduction

Despite gains in the overall education of women over the last half century, certain fields remain heavily male-dominated, including the field of economics. A growing research literature focuses on how choice of major contributes to gender differences in the labor market trajectories of college educated men and women. High-earning majors tend to be male-dominated. For example, those who graduate with a bachelor's degree in economics have on average higher earnings than other majors (Hershbein and Kearney, 2014; Bleemer and Mehta, 2022; Black et al., 2003). Despite these high returns, the proportion of undergraduate economics degrees going to women has been at around 30 percent since the early 1990s (Bayer and Rouse, 2016, Lundberg and Stearns, 2019). This under-representation of women in economics begins early in students' academic careers: at the University of Wisconsin-Madison – the site of our study – although about 45% of students who take introductory economics courses as undergraduates are women, only 30% of students in the intermediate level of courses are women.

Past research has explored a number of different factors that could contribute to women's under-representation in certain high earnings majors, such as gender differences in preferences for the subject matter; expectations of job characteristics and earnings associated with the major; traits such as risk tolerance, patience and competitiveness; peer and family influences; mathematical aptitude; and sensitivity to grades in introductory courses (see Bayer and Rouse (2016) and Patnaik et al. (2020) for an overview of this work). Other research (Porter and Serra, 2020; Breda et al., 2023; Kofoed et al., 2019; Carrell et al., 2010; Bettinger and Long, 2005; Rask and Bailey, 2002; Canes and Rosen, 1995; Hoffmann and Oreopoulos, 2009; Fairlie et al., 2014; Kato and Song, 2022) has focused on the role of same-gender or same-race mentors, professors, and role models, or lack thereof, in encouraging students to pursue the same majors. It is this last explanation that our current research focuses on.

In our study, we combine an in-class intervention with administrative data on student course-taking to test how economics alumni speakers impact students' future course-taking in a large US public university. We model our intervention after the experiment described in Porter and Serra (2020), with the goal of exploring whether the gender of the speaker is a key aspect of encouraging female students and whether an intervention can scale up to the setting of a large public university. Porter and Serra (2020) find that short presentations from female alumni to small introductory economics courses significantly increase the likelihood that women take future courses in economics and the likelihood that women major in economics.

We conduct a similar guest speaker or role model intervention as in past work but include both male and female speakers, allowing us to examine differential role model effects on male and female students. In six out of the eight introductory economics courses in the 2018-19 school year, alumni speakers spoke for 15 minutes with three lectures receiving a talk from a female speaker and three receiving a talk from a male speaker. Given a lecture-level intervention and the few lectures available, we could not randomize the treatment. Instead we exploit our administrative transcript data on the full population of economics undergraduate student across several years and evaluate various constructed control groups comprised of students from contemporaneous and past untreated lectures, testing for balance using student demographic and test score information. We estimate the effect of the treatment on future course-taking using two primary specifications: a two-period difference-in-difference model and a multi-period two-way fixed effects model which include professor-level and semester-year-level fixed effects and student-level covariates.

Overall we estimate that the alumni intervention increases the likelihood that students continue in economics by taking intermediate microeconomics by 1.7-2.1 percentage points or 9-12% more than the baseline level. We find that these effects are much larger when we look at the effects separately by gender of the speaker and gender of the student. A male speaker increases the likelihood that male students take intermediate microeconomics by 8.1-8.8 percentage points or 36-38% (from the base rate of 22.5%) and has no significant effect on women. Female speakers increase the likelihood that female students take intermediate microeconomics by 4.9-5.0 percentage points or 37-40% (from the base rate of 12.4%) and have no significant effect on men. These results demonstrate that the findings in Porter and Serra (2020) generalize across university settings and extend their findings to suggest that there is a specific effect of *same-gender* speakers for both male and female students.

Additionally, we analyze the effects of the speakers across different demographic dimensions and find suggestive evidence for racial homophily for male students. Our speakers were all White and two worked for a Wisconsin-based company.[1] We find that the effects of the intervention on course take-up are larger for White male students, with the male speaker increasing take-up more for White male students than non-White male students. Unlike with gender, we do not have a non-White or non-Wisconsin based treatment arm, but these findings are suggestive that the similarity of the speaker to a student impacts how effective the intervention will be.

Last, we show that the impacts of the intervention varied across the skill distribution. Past research has shown that one of the contributors to gender gaps in quantitative majors such as economics is that women seem to be more sensitive to grade signals in introductory courses than men and less confident in their own-skills conditional on grades than men (Goldin, 2015; Rask and Tiefenthaler, 2008, McEwan et al., 2021; Bordón et al., 2020; Saltiel, 2023; Owen, 2023; Kugler et al., 2021). Other studies (Owen, 2021; Astorne-Figari and Speer, 2019) find that while lower grades are predictive of switching out of quantitative majors, there is no statistically significant difference in switching due to grades by gender. We see similar patterns to the former set of studies for students at UW-Madison; while men are equally likely to continue on to intermediate microeconomics regardless of grade in Econ 101 or pre-college math test scores, women's likelihood of continuing is decreasing in both Econ 101 grades and math test scores. Using ACT/SAT math

[1] The company was, in fact, a cheese company.

score terciles as a proxy for pre-college ability, we show that while the same-gender speaker had similar effects regardless of skills, the effects of the opposite-gender speaker differed across skill levels. For women, the female speaker had similar positive effects on women at all math test score levels, but the male speaker significantly decreased take-up for high-skilled women and had null effects on low-skilled women. For men, the male speaker had similar positive effects across all math test scores, but the female speaker increased course-taking for men with high math test scores and the male speaker had null effects on low-skilled men. Taken together with the overall findings about the effectiveness of same-gender speakers, these results demonstrate the importance of considering the marginal student affected by these interventions and how to design policy interventions to target these students.

In the next section, we review the current literature. We then describe the study setting, data collection, descriptive evidence on gendered patterns in economic course-taking, and our intervention. Next, we discuss the identification of the treatment effects in our study design and present various balance tests. We then present various estimates of overall and sub-group specific effects of the intervention. Lastly, we show that the results are robust to various changes in the estimation sample and estimator.

2 Literature Review

A large body of work studies the determinants of college majors and gender differences in choices of field of study. This literature has identified that earnings expectations play a relatively small role and has attempted to deconstruct what are the underlying factors in "unobserved tastes" such as differences in non-wage considerations or behavioral traits (Zafar, 2013; Buser et al., 2014; Bronson, 2014; Wasserman, 2023; Wiswall and Zafar, 2021, 2018; Gemici and Wiswall, 2014; Patnaik et al., 2020; Reuben et al., 2017). Many of these past papers rely on small-scale survey based data; our study uses both a policy intervention and large-scale transcript data from one of the largest public universities to analyze student's realized choices.[2] Our study also focuses on a specific channel through which these unobserved tastes for major may evolve – exposure to same-gendered role models – and extends the existing literature on role model impacts in two ways.

First, our study allows us to test whether male and female students respond similarly to short-term exposure to same-gendered role models by varying the type of 'role model' treatment, including both male and female alumni. Though past studies find that long-term exposure to male versus female mentors or teachers increases interest in a major (e.g., Kofoed et al., 2019; Lim and Meer, 2017), studies that involve brief exposure to successful individuals in a field have typically focused on exposure to a female role model vs. no role model (e.g., Porter and Serra, 2020; Breda et al., 2023; Li, 2018). In these studies, it is less clear whether it

[2]Carlson et al. (2022); Li (2018); Patnaik (2021); Owen (2023) are other recent papers taking advantage of administrative data in a large public university setting. University of Wisconsin-Madison is the fourth largest source of bachelor's degrees in economics, making it an ideal setting to study determinants of interest in economics.

was merely the exposure to any successful role model that resulted in increased interest in the field or if the gender of the role model made them particularly effective. For example, if there are gender differences in information about the career opportunities available to economics majors, a speaker of any gender may increase women's interest in an economics major through an information provision channel.[3] By comparing the effects of female versus male role models, we are able to speak to the relative benefits of same-gender role models versus the information or inspiration that any successful person in the field provides.

Second, our study contributes to our understanding of how effective college major 'nudge' interventions are and how well these interventions generalize across universities. If the findings from Porter and Serra (2020) replicate in other university settings, this low-cost intervention has a very high benefit-to-cost ratio and has the potential to make a significant difference in the gender gap in economics. A number of recent studies have attempted to replicate low cost interventions to reduce the gender gap in economics, such as letters encouraging high-performing students to continue in economics, providing information about earnings, or providing information about performance relative to the class average. These studies have had mixed results ranging from large positive impacts on high-performing women's take-up of economics majors (Li, 2018), non-significant or small effects on the gender gap (Bayer et al., 2019; Halim et al., 2022; Chambers et al., 2021), or effects that increased the gender gap by increasing male take-up (Pugatch and Schroeder, 2021; Bedard et al., 2021) or reducing female take-up (Antman et al., 2020). We might expect that the effectiveness of 'nudge' style interventions would vary across university types or based on the demographics of the students receiving the intervention. Our study tests whether alumni 'role models' are effective in a different university setting than past light-touch role model studies – a large public university – relative to high schools (Breda et al., 2023) or smaller private universities (Porter and Serra, 2020). Our results indicate that the findings from Porter and Serra (2020) replicate across settings, demonstrating that this type of 'nudge' may be more generalizable across settings than other light touch interventions such as encouragement letters.

3 Setting and Data

We conduct our study at the University of Wisconsin-Madison (UW-Madison), a large public university that is the flagship school in the state's university system. Economics is consistently one of the most popular majors. From 2014-2019, the average annual number of graduates with an economics major was 551 or 7.9 percent of all degrees conferred.

Introductory Microeconomics (Econ 101) is the first course in the economics sequence and is taught in lecture-style courses, with 300-400 students in each course. There are typically four total lectures of Econ 101 offered in each semester (Fall and Spring), with an enrollment of around 1500 in the Fall and 1000 in

[3]For example, past research (Reuben et al., 2017) suggests that women and men have different expectations about the earnings associated with different majors, with men expecting higher earnings on average than women.

the Spring.[4] The course's popularity is in part because this is a requirement for application to Wisconsin's business school and is a common choice for fulfilling social science requirements. Econ 101 is a prerequisite for the next course in the economics sequence, introductory macroeconomics (Econ 102), and students can not proceed with the sequence until they receive credit for Econ 101.

Our primary data source is administrative transcript data for all students who attended UW-Madison between the years of 2010/11 and 2019/20. The data contains detailed transcript information including courses taken and grades received in each semester. Using unique anonymous identifiers we link students across semesters and are therefore able to identify all students who took Econ 101 at any point during the period, as well as identifying if they took future courses in economics, such as introductory macroeconomics (Econ 102), intermediate microeconomics (Econ 301), or intermediate macroeconomics (Econ 302), and their completed major or majors.[5] The administrative data from Wisconsin also contains rich student level information including gender, race, international student status, state residency at admission, and measures of student ability including standardized admission test scores such as ACT and SAT scores.

4 Gender Gaps in Course-Taking

Though the economics major is male-dominated at UW-Madison, Econ 101 is nearly gender-balanced. Approximately 45 percent of Econ 101 students are women in comparison to the 25-30 percent of graduating majors who are women. This gender gap opens up almost immediately following Econ 101; the percent of women who take introductory macroeconomics is 40 percent and the percent of women who take the intermediate microeconomics course is 30 percent. This indicates that these introductory microeconomics and macroeconomics courses are the point at which women drop out of the economics pipelines and suggests they are therefore the point at which an intervention may be most effective.

The linked administrative data allows us to understand what demographic characteristics are associated with continuing economics coursework following Econ 101. Figure 1 provides the take-up of Econ 301 for both men and women, conditional on a number of student characteristics. Panel A provides the take-up based on Econ 101 grades.[6] Although male students are about as likely to continue to Econ 301 at every grade level, the proportion of female students who continue is strongly increasing in Econ 101 grades. This implies that the gender gap is much larger among B students than among A students, although a sizable gender gap exists at each grade level. A similar pattern exists using admission tests scores (Panel B).

Panels C and D break the students taking Econ 101 into demographic groups. Panel C includes only American

[4] There are also a number of Summer courses–discussed below–and we largely exclude these from the analysis. Results are robust to their inclusion.

[5] Undergraduates are not required to declare a major until they have reached 86 course credits and typically declare their major in the fall of junior year.

[6] UW-Madison uses a non-standard grading regime which combines +/- grades. For example, rather than having A- and B+, that letter grade is combined into the grade AB.

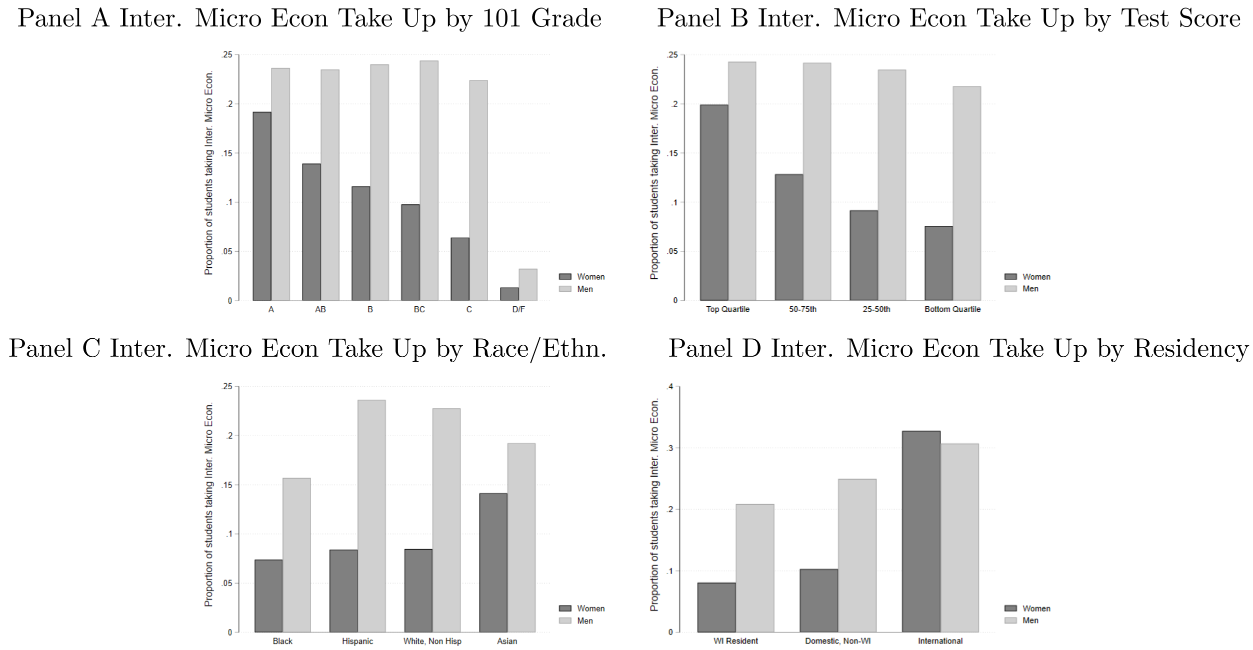

Figure 1: Take-up of Intermediate Microeconomics, by Econ 101 Grade, Math Test Scores and demographic characteristics
Notes: This figure tabulates the proportion of students for each gender who continued on to take Intermediate Microeconomics by grade in Econ 101 (Panel A), quartile of the ACT/SAT score (Panel B), race (Panel C) and residency status (Panel D). The sample includes all students whose first semester at UW-Madison fell between Fall 2010 and Spring 2016.

students and shows the gender gap is about the same for Black, Hispanic, and White/Non-Hispanic American students, but the gap is smaller but still sizable for Asian American students. Panel D shows the gender gap in continuing to Econ 301 is similar for Wisconsin resident and non-resident American students, but there is a female *advantage* among the international students.[7]

Continuing this analysis of who continues, Table 1 reports the mean and standard deviation of demographic characteristics for men and women among students who took Econ 101 between 2010 and 2016 and those who continued on to Econ 301. We also report the difference across groups, as well as results of a t-test of equality of means. Women who take Econ 101 are less likely to be White Americans than their male counterparts and instead more likely to be Asian/Pacific Islander (API) Americans or international students, though the differences are small in magnitude. This gap grows significantly among students who continue on to Econ 301. While the racial make-up of men who continue on to Econ 301 mirrors their demographics in Econ 101, White American women are much less likely to continue with the sequence and API American and International students are disproportionately likely to continue in economics, with the proportion of White Americans dropping from 77% in Econ 101 to 55% in Econ 301.

There are small differences in the timing of course-taking by gender. Women are slightly more likely to be sophomores when they take Econ 101 than men, but among those who continue on to Econ 301, men are

[7] Race is not reported for international students, and they are not included in the Panel C analysis.

Table 1: Comparison of Men and Women who Take Economics Courses

	Intro Microeconomics			Intermediate Microeconomics			Major		
	Men	Women	Difference	Men	Women	Difference	Men	Women	Difference
White Non-Hispanic	0.795	0.767	-0.0274***	0.770	0.549	-0.221***	0.782	0.575	-0.207***
	(0.404)	(0.423)	[0.007]	(0.421)	(0.498)	[0.018]	(0.413)	(0.495)	[0.021]
Asian/Pacific-Islander Category	0.0582	0.0671	0.0088*	0.0484	0.0833	0.0349***	0.0560	0.0921	0.0360**
	(0.234)	(0.250)	[0.004]	(0.215)	(0.277)	[0.009]	(0.230)	(0.289)	[0.012]
Non-White/Non-Asian Category	0.0635	0.0600	-0.0034	0.0606	0.0498	-0.0108	0.0623	0.0460	-0.0163
	(0.244)	(0.238)	[0.004]	(0.239)	(0.218)	[0.009]	(0.242)	(0.210)	[0.011]
International Student	0.0966	0.121	0.0242***	0.136	0.340	0.204***	0.116	0.314	0.198***
	(0.295)	(0.326)	[0.005]	(0.343)	(0.474)	[0.016]	(0.321)	(0.465)	[0.017]
First-years	0.724	0.711	-0.0137	0.798	0.814	0.0158	0.787	0.805	0.0174
	(0.447)	(0.454)	[0.007]	(0.402)	(0.390)	[0.016]	(0.409)	(0.397)	[0.019]
Sophomores	0.190	0.216	0.0260***	0.147	0.159	0.0115	0.155	0.171	0.0163
	(0.392)	(0.412)	[0.006]	(0.354)	(0.365)	[0.014]	(0.362)	(0.377)	[0.017]
Juniors	0.0558	0.0509	-0.00488	0.0264	0.0150	-0.0113	0.0313	0.0143	-0.0170*
	(0.230)	(0.220)	[0.004]	(0.160)	(0.122)	[0.006]	(0.174)	(0.119)	[0.008]
ACT score	28.34	27.85	-0.489***	28.35	28.27	-0.0851	28.35	28.36	0.00937
	(2.969)	(2.819)	[0.051]	(2.926)	(2.830)	[0.145]	(2.873)	(2.793)	[0.164]
Econ 101 Grade, B or Higher	0.712	0.610	-0.102***	0.733	0.742	0.00916	0.732	0.754	0.0216
	(0.453)	(0.488)	[0.007]	(0.443)	(0.438)	[0.018]	(0.443)	(0.431)	[0.021]
Observations	8965	7112	16077	2129	864	2993	1573	630	2203

Notes: This table reports mean and standard deviation (in parentheses) of demographics for men and women in Econ 101 (col. 1 and 2), men and women in Econ 301 (col. 4 and 5), and men and women who major in economics (col. 7 and 8) for the years prior to 2016. Column 3, 6, and 9 report the results of a balance test in which the demographic outcome is regressed on an indicator for being female. Standard errors are in brackets. $p < 0.05$ * $p < 0.01$ ** $p < 0.001$ ***

slightly more likely to have taken Econ 101 in their junior year.

In terms of test scores, men have significantly higher ACT scores than women, but these differences no longer remain among those who continue on to Econ 301. We see a similar pattern in terms of performance in Econ 101: while men are more likely to have received a B or higher in Econ 101 than women (71% of men versus 61% of women), there is no difference in performance in Econ 101 among those who continue on to future coursework. Comparing those in Econ 301 to those who major in economics, we see fairly similar demographic patterns, suggesting that the selection into Econ 301 is a good proxy for the types of students who will continue on to the major.

These descriptive analyses suggest an important contrast between male and female economics course-taking. The men who take Econ 101 look similar demographically and in terms of test scores and grades to the men who take Econ 301, suggesting that there are no systematic patterns by demographics for men who continue on to future economics coursework. In contrast, women who continue on in economics look substantively different than those who took Econ 101 – they are more likely to be an Asian American or an international student than those who do not continue, and they are more likely to be a high-performing student both overall in terms of admissions scores (ACT score) and in economics in particular (Econ 101 grade).

5 Intervention Design

For our role model intervention, we invited alumni who graduated from the department to give a fifteen minute presentation to the Econ 101 courses in the sixth week of classes. The study took place during the 2018-19 academic year, encompassing the Fall 2018 and Spring 2019 semesters.

To recruit alumni in Fall 2018, we relied on the assistance of department staff to compile a list of possible alumni speakers who had careers that undergraduates may be interested in and had graduated within the past ten years. We reached out to 20 women who had graduated from UW-Madison between 2003 and 2008. The majority of alumni did not respond to emails. There was ultimately one women among the 20 inquiries who was available in the week of our intervention to speak to the students. She (RM1) worked in supply chain management with past jobs in manufacturing, transportation, and cheese production firms. Later we were able to identify from among alumni board members for the department a second female alumna (RM2) who was a marketing executive with experience in publishing and education sectors. For Spring 2019, we invited a male alumnus (RM3), who worked at the same company in a similar job as the first female alumna.

The alumni speakers were given a series of questions as prompts, including questions about their first jobs out of college, their experiences in economics course work, the skills they think they gained in the economics courses, and how an economics degree helps them in the work force. In the talks, each speaker discussed their current job responsibilities and how they found the job that they are currently in. All three alumni mentioned that they still use concepts they learned in Econ 101 in their jobs today; for example, RM1 discussed how understanding tariffs is important for a cheese company that exports their products.

To allocate these speakers, we were limited by the number of possible lectures. Over the academic year Fall 2018-Spring 2019, there were eight total Econ 101 lectures. Given the intervention was at the lecture level (and not the student level), as well as the small number of possible lectures, randomization of treatment was not possible. In total six lectures received an alumni speaker treatment, and two lectures did not. We intentionally assigned the no treatment lectures to the two professors who teach Econ 101 in both semesters to allow these two professors to have one treated and one untreated lecture over the course of the year. The remaining assignment of alumni speakers was based on their availability and scheduling constraints: RM1 presented to one lecture in Fall 2018; RM2 presented to two lectures in Fall 2018; and RM3 presented to three lectures in Spring 2019. One lecture in each semester did not receive the intervention. Because both students and professors were unaware of which lecture would receive the intervention at the time of course selection, it is not possible that students selected into specific lectures in anticipation of the intervention.Table 2 delineates which professors were teaching in each of the semesters included in our analysis and indicating whether that lecture was included as a control lecture, omitted, or had a treatment. Lectures are omitted if an alternative intervention took place (professor 1 in Fall 2015-16 and Spring 2016-17) or if we were piloting our intervention (Spring 2017-18). RM1 and RM2 refer to the two female speaker interventions; RM3 refers to the male speaker intervention.

Table 2: Treatment and Control lectures, by Semester and Professor

	Fall 2015-16	Spring 2015-16	Fall 2016-17	Spring 2016-17	Fall 2017-18	Spring 2017-18	Fall 2018-19	Spring 2018-19
Prof. 1	Omitted	Control	Did not teach	Omitted	Did not teach	Omitted	RM2	Did not teach
Prof. 2	Control Control	Control Control	Control	Control	Control Control	Omitted	RM1	Control
Prof. 3	Did not teach	Control	Control Control	Control	Control	Omitted	RM2 Control	RM3 RM3
Prof. 4	Did not teach	Did not teach	Did not teach	Did not teach	Control	Omitted	Did not teach	RM3
Prof. 5	Did not teach	Did not teach	Control	Did not teach	Did not teach	Omitted	Did not teach	Did not teach

Notes: This table reports the treatment status of lectures in our sample, by professor (row) and semester(column). RM1, RM2, and RM3 refer to treatment semesters. Control refers to semesters in our control group. Semesters in which the professor was not teaching are 'Did not teach' and semesters omitted due to alternative internvetions are 'Omitted.' If a professor taught two lectures, both are listed in the same column.

To provide some additional context for the setting of the intervention, the professor make-up is 60% female, with Professor 1, 2, and 4 being women. All professors are dedicated teaching professors who are not tenure-track or research professors, but are instead in semi-permanent contracts in which they teach multiple undergraduate courses each semester. Professor 1, 2, 3, and 5 each had over 15 years of teaching experience at the time of the treatment; Professor 4 had less than 5 years of teaching experience. Professor 1, 2, 3, and 4 are all white; Professor 5 is non-white.

6 Sample, Design, and Balance Tests

In this section we describe our treatment sample, control groups, and conduct tests of balance between the groups.

6.1 Analysis Sample

Our primary sample is restricted to students who completed Econ 101 between the years 2015/16 and 2018/19 in the Fall or Spring semesters.[8] We exclude the small number of students who were seniors (in their fourth year of study or more) when we observe them taking Econ 101. We observe each student at least through the 2020/21 academic year, allowing at least 2 additional years following Econ 101 to take additional economics courses.

Column 1 of Table 3 presents descriptive statistics for our primary sample of 7,730 total Econ 101 students in 24 total lectures, 6 treated and 18 untreated. 43 percent of the total students were female. In terms of race and origin, the vast majority, 75 percent, were White, non-Hispanic American students. 12 percent of the total were international students, and the remaining 14 percent were non-White.[9] About 74 percent of the students were First year students, 18 percent Sophomores, and 5 percent Juniors, and, as noted above,

[8] In 2015/16, a new lecturer was hired to teach Econ 101 and became a permanent part of the teaching roster for the course. For the primary sample we omit earlier years of data in our analyses to ensure that our control sample is as similar to the intervention years as possible.

[9] About 2 percent of students report multiple races. We allow the dummy variables for race to equal one for all races reported in all analyses. We also re-run our results dropping multi-racial students; it does not substantively change the results.

the small number of Seniors were excluded entirely from this sample.[10]

Table 3: Sample Characteristics and Balance Tests

		Two-Way FE			Diff-in-Diff		
	Full Sample	Treated	Control	Balance Test	Treated	Control	Balance Test
Female	0.429	0.413	0.434	-0.0389***	0.420	0.450	-0.0281**
	[0.495]	[0.493]	[0.496]	(0.0100)	[0.494]	[0.498]	(0.0102)
White Non-Hispanic	0.750	0.688	0.769	-0.0656**	0.689	0.754	-0.0636*
	[0.433]	[0.463]	[0.422]	(0.0242)	[0.463]	[0.431]	(0.0266)
Asian/Pacific-Islander Category	0.0728	0.0799	0.0707	-0.00834	0.0821	0.0753	-0.00597
	[0.260]	[0.271]	[0.256]	(0.00580)	[0.275]	[0.264]	(0.0065)
Non-White/Non-Asian Category	0.0781	0.102	0.0708	0.0177*	0.102	0.0735	0.0125**
	[0.268]	[0.302]	[0.257]	(0.00875)	[0.303]	[0.261]	(0.0047)
International Student	0.118	0.148	0.108	0.0537*	0.144	0.115	0.0563*
	[0.322]	[0.355]	[0.311]	(0.0238)	[0.351]	[0.320]	(0.0226)
First-year	0.737	0.756	0.731	-0.0412*	0.735	0.728	-0.0622***
	[0.440]	[0.429]	[0.444]	(0.0162)	[0.442]	[0.445]	(0.0172)
Sophomore	0.182	0.159	0.189	0.0376*	0.176	0.192	0.0520*
	[0.385]	[0.366]	[0.391]	(0.0163)	[0.381]	[0.394]	(0.0203)
Junior	0.0501	0.0449	0.0517	0.00686	0.0519	0.0481	0.0147**
	[0.218]	[0.207]	[0.221]	(0.0077)	[0.222]	[0.214]	(0.0045)
ACT score	28.58	28.81	28.52	-0.345	28.89	28.60	-0.329
	[3.037]	[3.140]	[3.004]	(0.2270)	[3.112]	[3.068]	(0.2080)
χ^2 , Joint Test of Null			2946.2			11655.6	
Predicted Diff. in 301 Takeup			0.0067			0.0074	
Observations	7730	1828	5902	7730	1194	2244	3438

Notes: This table reports mean and standard deviation (in brackets) of demographics for the full sample (col. 1), treatment group in the two-way FE specification sample which includes all years between 2015/16 and 2018/19 (col. 2), the control group in the two-way FE specification (col. 3), the treatment observations in the D-inD specification which only includes Prof. 1 and Prof. 3 (col. 5), and control observations in the D-in-D specification (col. 6). Column 4 and 7 report the results of balance tests in which the demographic outcome is regressed on an indicator for the treatment, semester-year FEs, and professor FEs for the two-way FE sample and the D-in-D sample. We report the test-statistic for the test of the joint null hypothesis that the coefficient on the treatment is equal to zero for all characteristics, $\chi^2(10)$. Standard errors (in parentheses) are clustered at the lecture-level. $p < 0.05$ * $p < 0.01$ ** $p < 0.001$ ***

6.2 Empirical Design

Ideally, we would be able to randomly assign students to see the alumni speaker presentation, and assignment to this treatment would therefore be independent of all unobservable components that contribute to the student's interest in economics. With a lecture-level treatment and only eight possible lectures spread over two semesters, simple randomization is not possible. Moreover, not all professors teach the exact same number or time slot of lectures each semester, meaning that there is not a specific prior year's courses that can be used as a pre-period comparison group. We therefore need to identify a plausible control group for the treated lectures.

We construct and evaluate various control groups, considering both the contemporaneous untreated lectures in the 2018-19 academic year and past untreated lectures. Our analysis focuses on two primary specifications. In our first specification, we include a treatment group comprised of all treated lectures and a control group made up of untreated lectures from the 2015/2016 through 2018/19 academic years. Second, we use a single

[10]3 percent of students started taking courses at UW-Madison off-cycle (i.e., the first time they are observed is a summer semester). Because our designation of class-year at time of Econ 101 is based off of time relative to start-date, these off-cycle students have a missing class year in our analysis. Omitting these students does not substantively change our results.

year – 2016-17– as the control group and restrict the sample to the two professors for whom we have both a control and a treatment observation during the year of the intervention (Panel B), as described in Table 2. This mirrors the two-period difference-in-difference set up in Porter and Serra (2020) in which we are able to match the lectures in the treatment year to the lectures taught by the same professors in an earlier year. Though this allows for greater comparability to past work, we are not using the full information set available to us in this analysis.

Although we could not practically randomize our intervention, we may nevertheless have been able to achieve independence of the treatment with respect to relevant student characteristics if students largely register for lectures based on factors unrelated to future interest in economics. Two institutional factors help us in this regard. First, students register for lectures with limited information and no anticipation of our future intervention, which was not announced. Second, the lectures are typically capacity constrained, meaning students who have preferences may not be able to select the lecture of their choice, especially first-years who have lower registration priority.

6.3 Balance Tests

We begin our test for balance in Table 3 using the two specification samples: one using past pre-intervention lectures from the 2015/16 to 2017/18 academic years (all of which are untreated) and the other using the untreated lectures for the professors who are in the difference-in-difference sample which includes the 2016/17 and 2018/19 academic years. Columns 3 and 4 provide the mean and standard deviation of student characteristics for these definitions of control lectures, and these statistics can be compared to the same statistics for the treatment lectures in Column 2.

Columns 5 and 6 provide the difference in means between treatment and control for each characteristic, derived from a regression pooling treatment and control samples in which each characteristic is regressed on an indicator for treatment, including professor and semester-year fixed effects (as we have in our main specifications presented below). The standard errors for these coefficients are estimated assuming clustering at the lecture level. In addition, at the bottom of the table, we compute a chi-squared test for the joint hypothesis that all treatment-control differences are 0 across all 10 characteristics. Lastly, as our speakers' genders varied across semesters and may impact students differently by gender, Table 4 also reports comparable balance tests separately by gender and semester.

Although there are statistically significant treatment-control differences in many of the characteristics, the magnitudes are small. For both definitions of the control group, we see that the treatment lectures have about 3-4 percentage point fewer female students, 6 percentage point fewer White/non-Hispanic students, 4 to 6 percentage point fewer First-year students (and nearly correspondingly higher Sophomore students). There is about a 1/3 of an ACT point average difference between treatment and control, a small difference

Table 4: Sample Characteristics and Balance Tests, By Gender and Semester

	Women						Men					
	Fall, Treated	Fall, Control	Balance	Spring, Treated	Spring, Control	Balance	Fall, Treated	Fall, Control	Balance	Spring, Treated	Spring, Control	Balance
White Non-Hispanic	0.683	0.749	-0.109*	0.751	0.775	0.0553	0.668	0.749	-0.136***	0.751	0.775	-0.0619
	[0.466]	[0.433]	(0.05)	[0.433]	[0.418]	(0.06)	[-0.471]	[0.433]	(0.04)	[0.433]	[0.418]	(0.05)
Asian/Pacific-Islander	0.0803	0.0821	0.0269	0.0545	0.0744	-0.0207	0.0924	0.0821	-0.0171	0.0545	0.0744	-0.0152
	[0.272]	[0.275]	(0.03)	[0.227]	[0.263]	(0.04)	[0.29]	[0.275]	(0.02)	[0.227]	[0.263]	(0.03)
Non-White/Non-Asian	0.102	0.065	-0.00207	0.101	0.0658	0.0348	0.0996	0.065	0.0082	0.101	0.0658	0.0473
	[0.303]	[0.247]	(0.03)	[0.302]	[0.248]	(0.04)	[0.300]	[-0.247]	(0.03)	[0.302]	[0.248]	(0.03)
International Student	0.149	0.126	0.0876*	0.117	0.0971	-0.0505	0.162	0.126	0.125***	0.117	0.0971	0.0158
	[0.356]	[0.332]	(0.04)	[0.322]	[-0.296]	(0.04)	[0.368]	[0.332]	(0.03)	[0.322]	[0.296]	(0.04)
First-year	0.705	0.722	-0.0309	0.809	0.748	-0.00088	0.775	0.722	0.00127	0.809	0.748	0.0214
	[0.457]	[0.448]	(0.05)	[0.394]	[0.435]	(0.06)	[-0.418]	[0.448]	(0.04)	[0.394]	[0.435]	(0.06)
Sophomore	0.189	0.21	-0.0315	0.121	0.179	0.0398	0.143	0.21	0.0152	0.121	0.179	-0.0535
	[0.392]	[0.408]	(0.04)	[0.326]	[0.384]	(0.05)	[0.35]	[0.408]	(0.03)	[0.326]	[0.384]	(0.05)
Junior	0.0502	0.0515	0.014	0.0428	0.0367	-0.00847	0.0491	0.0515	0.0135	0.0428	0.0367	0.001
	[0.219]	[0.221]	(0.02)	[-0.203]	[0.188]	(0.03)	[0.216]	[0.221]	(0.02)	[0.203	[0.188]	(0.03)
ACT Score	28.54	28.36	-0.373	28.25	27.98	7.43E-05	29.47	28.36	0.495	28.25	27.98	-1.243**
	[3.099]	[2.843]	(0.36)	[3.126]	[3.026]	(0.48)	[3.091]	[2.843]	(0.31)	[3.126]	[3.026]	(0.45)
Observations	498	1632	2130	257	927	1184	693	1632	2971	257	927	1445

Notes: This table reports mean and standard deviation (in brackets) of demographics for the treatment the control group in the primary specification sample which includes all years between 2015/16 and 2018/19, separately by gender and semester. Columns labelled 'Balance' reports the results of a balance test in which the demographic outcome is regressed on an indicator for the treatment, semester-year FEs, and professor FEs for the primary specification sample. Standard errors (in parentheses) are clustered at the lecture-level. $p < 0.05$ * $p < 0.01$ ** $p < 0.001$ ***

relative to the average score of 28 points and not statistically significant from 0.[11] When split by semester and gender, we see that these differences between treatment and control are similar across gender and concentrated in the Fall semester.

To provide a better measure of how important these differences in covariates are between treated and untreated lectures, we conduct an additional analysis combining all characteristics together in an "importance index," weighting each characteristic by their estimated marginal contribution to the dependent variable. We estimate a model using only the untreated lectures prior to Fall 2018 (our intervention) and including all of the covariates we consider for balance. This provides an estimate of the influence of each covariate on future course-taking. We then predict course-taking using the estimate treatment - control difference in covariates multiplied by these estimates of each covariates influence. The results are reported in Table 3. We find that the difference in covariates between treatment and control lectures would increase the probability of taking Econ 301 by just 0.007 (less than 1 percentage point) and would result in an even smaller increase in course-taking when estimated separately by gender. Of course, we condition on all of these variables in our main specifications, and this exercise cannot provide a definitive answer to the key question of the extent and direction of selection on unobserved variables. However, this exercise suggests that the imbalance in these important covariates would only lead to a small upward bias in the unconditional estimate of the effect of the treatment, suggesting that even without randomization, the treatment and control lectures are not too different.

7 Results

In this section we report the effects of the treatment on various measures of economics course-taking. We begin by discussing the treatment effects for the sample as a whole and then go on to analyze whether this differs by the gender of the role model or by other characteristics of the students.

7.1 Empirical Specification

Our primary empirical specification is given by

$$Y_{iptg} = \beta_g D_{p(i)t} + \theta_{pg} + \tau_{tg} + X_i'\delta_g + \epsilon_{iptg}, \tag{1}$$

where Y_{iptg} is the outcome for student i who completed Econ 101 taught by professor p in semester and year defined by t (e.g., Spring semester 2018), g indexes student sex (male or female), $D_{p(i)t}$ is an indicator equal to 1 for student i being in a treated lecture taught by professor p's in semester t, θ_{pg} are professor

[11] Approximately 20% of the sample is missing ACT scores. For some of these students, we observe SAT scores, and in our regression analysis, we use these scores in addition to ACT scores (as separate variables), including dummy variables for missing scores. In our main specifications, we find that treatment and control lectures are not significantly different in terms of the proportion of students with missing ACT scores.

fixed effects and τ_{tg} are semester-year fixed effects, each allowed to vary freely by gender.[12] The vector X_i includes an intercept and indicators for race and residence, entry cohort, ACT/SAT scores, and age at college entrance. We do not control for grade in Econ 101 due to concerns that this is endogenous to the treatment; see Appendix Section A.1 for discussion of the impacts of the treatment on performances in economics coursework.

We estimate all specifications separately by gender. This allows treatment effects, covariates, and professor and time fixed effects to freely vary by student gender. This specification permits for the possibility that some professors lead to more women or men continuing on to take further courses (e.g. female professors have a role model effect on female students). We identify the effect of the treatment separately from professor effects because the professors in our sample teach multiple years and lectures, both treated and untreated. We also allow for time varying gender-specific semester effects, allowing for the possibility of overall gender-specific trends in economic course-taking. We identify these separately from the treatment effect because in the same semester, there are both treated and untreated lectures.

The estimating equation is the same for both of our primary specifications. The first specification, referred to as the 'two-way fixed effects' model, includes all treatment and control observations from 2015-16 through 2018-19 as described in Table table: professors. The second specification, referred to as the 'difference-in-difference' model, only includes lectures in the 2016-17 school year and the 2018-19 school year for Professors 2 and 3.

We estimate these models first with the treatment defined as any intervention. We then test whether the gender of the speaker/ role model impacts male or female students differently by including an indicator for the gender of the speaker, and testing equality of the coefficients across the female and the male samples. Standard errors are clustered at the lecture-level, and we also show pairwise bootstrapped errors.

7.2 Overall Effects

Table 5 reports the average effect of the treated lectures on various measures of course-taking, overall and by gender. Panel A reports results for the two-way fixed effects model that includes all treated lectures and Panel B reports results for the difference-in-difference specification that only includes lectures for Professor 2 and 3 in the years 2018/19 and 2016/17.[13] The outcomes that we focus on are whether the student enrolled in the next course in the economics sequence (i.e., Introductory Macroeconomics) or the next level of economics courses (i.e., Intermediate Microeconomics). Standard errors are clustered at the lecture level (in parentheses) and adjusted using pair-wise bootstrapping (in brackets). We find no statistically significant

[12] Note also that each semester course was taught by 1 professor (no co-teaching).

[13] Our preferred specification is the two-way FE model due to the difficulty in separately identifying whether the effects are driven by the gender of the speaker or the gender of the professor in the difference-in-difference specification. Professor 2 is male and professor 3 is female, meaning that if we do not also include the other lectures which are taught by women, the positive effect of the female speaker could be attributable to speakers being more encouraging to women being taught by a male

Table 5: Effects of Treatment on Future Economics Course-taking

	(1) Econ 102	(2) Econ 102: Men	(3) Econ 102: Women	(4) Econ 301	(5) Econ 301: Men	(6) Econ 301: Women
Panel A: Full Sample	0.00509	0.0199	-0.00418	0.0213**	0.0365+	0.0123
	(0.0106)	(0.0362)	(0.0392)	(0.0069)	(0.0198)	(0.0199)
	[0.0171]	[0.0720]	[0.0755]	[0.0161]	[0.0412]	[0.0410]
N	7729	4415	3314	7729	4415	3314
N. Lecture Clusters	24	24	24	24	24	24
Panel B: D-in-D	0.00109	0.0155	-0.00691	0.0171+	0.0285	0.0121
	(0.0112)	(0.0340)	(0.0492)	(0.00858)	(0.0189)	(0.0189)
	[0.0166]	[0.0743]	[0.0889]	[0.0195]	[0.0429]	[0.0416]
N	3438	1927	1511	3438	1927	1511
N. Lecture Clusters	11	11	11	11	11	11
Professor FE	Y	Y	Y	Y	Y	Y
Semester FE	Y	Y	Y	Y	Y	Y
Covariates	Y	Y	Y	Y	Y	Y
Control Mean	0.47	0.51	0.42	0.18	0.23	0.13

Note. This reports the results of regressions of taking intro macro (Econ 102) or intermediate micro (Econ 301) on an indicator for receiving the alumni speaker treatment, professor fixed effects, semester fixed effects, and controls for race, gender (in full sample), ACT/SAT score, age at entrance to college, class year during Econ 101, international student status, and state residence. Col. 1 and 4 are the full sample; Col. 2 and 5 are for men only; Col.3 and 6 are for women only. The sample is drawn from administrative transcript data from UW-Madison. Panel A includes all students who took Econ 101 between the years 2015/16 and 2018/19, excluding eight lectures which had other experimental interventions. Panel B includes all students who took Econ 101 in 2018/19 and 2016/17 from Professor 2 or Professor 3. Standard errors clustered at the lecture level in parentheses with indicators $p < 0.10$ + $p < 0.05$ * $p < 0.01$ ** $p < 0.001$ ***; Pair-wise bootstrapped standard errors in brackets.

effects of the treatment on take-up of the next course in the sequence, Introductory Macroeconomics (Econ 102). The small and non-significant effects on Econ 102 may be due to the fact that a large proportion of students take Econ 102 as a prerequisite for a business major and may have already been planning to take this course even in the absence of the treatment.

We do, however, see a somewhat sizable main effect of the treatment on take-up of the first course that indicates potential interest in an economics major: intermediate economics (Econ 301). This course is a requirement for the major and is almost exclusively taken by students majoring or considering majoring in economics.[14] We see that the treatment increased take-up of this course by 1.7 to 2.1 percentage points or a 9-12% increase off the base take-up rate of 18 percent. While it is significant at the 5-percent level when standard errors are clustered at the lecture level in the full sample, it is only marginally significant in the difference-in-difference specification and no long significant when we adjust standard errors using pair-wise bootstrapping. The magnitude of the result is robust to a series of alternate sample specifications, discussed in more detail in section 8.

While we cannot reject the null that the effects of the treatment are equal for male and female students, the effects are positive for both genders and slightly larger for men. This provides some suggestive evidence of a stronger effect of the policy on men than women, which differs from past findings in Porter and Serra (2020).

professor in the difference-in-difference specification.

[14]Of those who take Econ 301 between 2010 and 2016, 93% eventually major in economics, agricultural economics, or business, with 79% majoring in economics alone. 99.5% of Econ 101 students who eventually major in economics take Econ 301, compared to only 7% in the next most popular major among Econ 101 students, business.

However, their experiment only included female speakers, meaning the larger effects for male students seen in the current study may be due to the inclusion of a male speaker in this intervention.

7.3 Effects by Speaker Gender

Because we included both female and male speakers, we are able to test if the female students responded more to the female speaker or vice versa. The results of these estimates are reported in Table 6, along with standard errors are clustered at the lecture level (in parentheses) and adjusted using pair-wise bootstrapping (in brackets). We find that the average effects of the intervention differ by the gender of the role model and

Table 6: Effects of Treatment on Future Economics Course-taking, by Gender of Speaker

	Male Alumni		Female Alumni	
	(1)	(2)	(3)	(4)
	Male Student	Female Student	Male Student	Female Student
Panel A1	0.107***	-0.0680	-0.0487***	0.0639***
Econ 102	(0.0133)	(0.0565)	(0.0060)	(0.0139)
Full Sample	[0.0238]	[0.0724]	[0.0134]	[0.0289]
Panel A2	0.0806***	-0.0308	0.00113	0.0498**
Econ 301	(0.0114)	(0.0235)	(0.0116)	(0.0155)
Full Sample	[0.0201]	[0.0385]	[0.0161]	[0.0237]
N.Obs.	3723	2816	4035	3057
N. Clusters	21	21	21	21
Panel B1	0.126***	-0.121	-0.0541***	0.0804***
Econ 102	(0.0167)	(0.0809)	(0.0062)	(0.00881)
D-in-D	[0.0221]	[0.0921]	[0.0150]	[0.0165]
Panel B2	0.0876**	-0.0510+	-0.0117	0.0487*
Econ 301	(0.0204)	(0.0253)	(0.0078)	(0.0163)
D-in-D	[0.0256]	[0.0337]	[0.0164]	[0.0271]
N.Obs.	1464	1163	1697	1358
N. Clusters	9	9	9	9
Econ 102, Control Mean	0.51	0.42	0.51	0.42
Econ 301, Control Mean	0.23	0.13	0.23	0.13
Professor FE	Y	Y	Y	Y
Semester FE	Y	Y	Y	Y
Covariates	Y	Y	Y	Y

Note. This reports the results of regressions of taking intro macro (Econ 102) and intermediate micro (Econ 301) on an indicator for receiving the alumni speaker treatment conditional on the speaker's gender, professor fixed effects, semester fixed effects, and controls for race, gender, ACT/SAT score, age at entrance to college, class year during Econ 101, international student status, and state residence. Columns 1 and 3 are restricted to male students; Columns 2 and 4 are restricted to female students. The sample is drawn from administrative transcript data from UW-Madison. Panel A includes all students who took Econ 101 between the years 2015/16 and 2018/19, excluding eight lectures which had other experimental interventions. Panel B includes all students who took Econ 101 in 2018/19 and 2016/17 from Professor 2 or Professor 3. Standard errors clustered at the lecture level in parentheses with indicators $p < 0.10$ + $p < 0.05$ * $p < 0.01$ ** $p < 0.001$ ***; Pair-wise bootstrapped standard errors in brackets.

specifically that the male role model was more effective at encouraging take-up for male students whereas the female role model was more effective at encouraging take-up for female students. First, we see that the null effects overall for Econ 102 were masking the fact that each speaker had a positive impact on their own gender and a negative impact on the opposite gender. The male speaker increased the likelihood that male students took Econ 102 by 10.7-12.6 percentage points, and the female speaker increased the likelihood that female students took Econ 102 by 6.4-8.0 percentage points. However, the female speaker significantly reduced the likelihood of male students continuing on to Econ 102 by 4.9-5.4 percentage points; the male speaker had similarly negative though noisily estimated effects on female students.

Turning to Econ 301, we see the same evidence of homophily, and the effects are stronger relative to the base rate of course take-up. The male speaker increased male take-up of Econ 301 by 8.1-8.8 percentage points (36-38%) and the female speaker increased female take-up by 4.9-5.0 percentage points (37-40%).[15] There is no longer a negative impact of opposite gender speakers, suggesting that the students who were deterred from taking Econ 102 were more casual students who would not have continued on to later courses even in the absence of an intervention. These results are robust to pair-wise bootstrapping, as reported in the table.[16]

These results confirm the findings of Porter and Serra (2020) that same-gender speakers increase interest in the economics. By including a male speaker as comparison, we are able to take their findings one step further to rule out the possibility that the benefit of a speaker is purely an non-gendered information or inspiration channel that closes gender gaps in knowledge about job opportunities in economics. If this were true, the male and female speaker should have similar effects on both genders' interest in economics. By showing that male speakers increase male students' take-up and vice versa for female students, we provide more concrete evidence that this is a "role model" or homophily story.

7.4 If Not Economics? Alternative Course-Taking Analysis

Given that one motivation for encouraging women to take more economics courses is to improve their earnings potential, we also explore the majors from which we are attracting students. Interventions that improve women's representation in economics have value beyond their potential to steer women into higher earning majors, but we might consider an intervention particularly desirable if they are drawing women into economics who would have majored in a lower paying field. In contrast, if we are drawing women into economics from business or STEM – fields which depending on sub-field have similar or higher earnings – the intervention may have unintended negative impacts on student's long term economic outcomes.

To test this, we explore whether the intervention reduced the likelihood of taking the core courses required

[15] Specifications run separately by speaker find positive significant impacts of both female speakers on women's take-up of future course work and null impacts on male student's course work, with the younger speaker who works in supply chain management (RM1) having an impact of 8 p.p. on female take-up of Econ 301 and the older speaker who works in marketing (RM2) having an impact of 3 p.p. While this is suggestive that younger speakers are more appealing, we cannot reject the null that the effects are equal across the two female speakers. This does, however, suggest that speakers in occupations which may be less commonly associated with an economics degree such as marketing in the publishing sector do not have a stronger effect on women's interest.

[16] We have also tested whether these results are robust to Wild bootstrapping; the positive impacts for men remain statistically significant at the $p < 0.05$ level but the female estimates, while of the same magnitude, are now more noisily estimated and no longer significant at the p=0.10 level.

for a business major. We choose business as a comparison for two reasons. First, of the high-earning majors, we think students considering a business major have a greater likelihood of being on the margin of choosing economics relative to majors such as engineering or computer science. A majority of students taking introductory economics enter the course intending to apply to the business school, and for students who get a B or higher in the course, over 40% go on to take the business core courses. Thus, if we are concerned about drawing students from a high-paying major, business is the natural comparison point. Second, regardless of sub-field, a business major at UW-Madison must take four specific 300-level courses in finance, marketing, and management. This allows us to identify a set of intermediate level courses required for the major that are of comparable level to the intermediate economics courses.

To test whether the intervention is drawing students away from a business major, we re-run the full-sample specification in Table 6 with two new outcomes: an indicator for taking any of the four business core courses and an indicator for taking all of the four business core courses. Panel A and B in Table 7 report the results from these regressions. The results suggest that the intervention did in part draw students away from the business course sequence. For example, female students exposed to the female speaker are about 2.7 to 2.9 percentage points less likely to take business courses. Male students exposed to the male speaker are 2 to 4 percentage points less likely to take business courses. When we test the effects of the intervention on taking *either* intermediate microeconomics or the business core courses (Panel C), we see that the effect sizes are, in fact, smaller than our main specification and now only marginally significant. The relative magnitudes in Panel C compared to the main findings suggest that about two-thirds of the women we are drawing into economics are students who would have otherwise majored in business. Nonetheless, the total effect of the intervention on students taking economics or business is a noisily estimated positive effect, suggesting that these speakers did induce some students on non-business or economics paths to switch to economics.

These analyses also show a consistent negative impact of the female speakers on male students' interest in a business major. The female alumni reduces the likelihood that male students take the business core courses by around 5 percentage points. This also suggests an explanation for the negative impacts of the female speaker on male interest in Econ 102 and the null impacts on male interest in Econ 301 as shown in Table 6. Because Econ 102 is both a prerequisite for an economics major and a business major, if the female speaker intervention reduced interest in majoring in business but not economics, this would explain the decline in students taking Econ 102, but not Econ 301.

In addition to our focus on business courses, we also consider the total number of STEM classes and the total number of arts, humanities, and social science classes that were not economics that each student took.[17]

[17]We used the internal college subject code of each course to categorize classes based on within-university classifications of majors within the College of Letters and Sciences (UW-Madison, 2024). STEM is defined as courses within the categories 'computer, data, & information sciences'; 'natural sciences'; and engineering, where engineering is defined as courses taken in the College of Engineering. We defined Arts, Humanities, and Social Sciences as the 'social sciences' or 'arts & humanities' categories that UW-Madison uses.

Table 7: Effects of Treatment on Future Business Course-taking, by Gender of Speaker

	Male Alumni		Female Alumni	
	(1)	(2)	(3)	(4)
	Male Student	Female Student	Male Student	Female Student
Panel A: Any Business Core	-0.0444*	0.00571	-0.0473**	-0.0285
	(0.0198)	(0.0463)	(0.0149)	(0.0200)
	[0.0421]	[0.0660]	[0.0282]	[0.0419]
Control Mean	0.36	0.40	0.36	0.40
Panel B: All Business Core	-0.0203	-0.0260	-0.0548***	-0.0273*
	(0.0136)	(0.0175)	(0.0104)	(0.0121)
	[0.0252]	[0.0348]	[0.0122]	[0.0204]
Control Mean	0.16	0.16	0.16	0.16
Panel C: Econ or Business Core	0.0204^{+}	-0.0133	-0.0448**	0.0159
	(0.0115)	(0.0350)	(0.0139)	(0.0226)
	[0.0237]	[0.0501]	[0.0216]	[0.0455]
Control Mean	0.51	0.48	0.51	0.48
Professor FE	Y	Y	Y	Y
Semester FE	Y	Y	Y	Y
Covariates	Y	Y	Y	Y
N.Obs.	3723	2816	4035	3057
N. Clusters	21	21	21	21

Note. This table reports the results of regressions of taking any Business Core Classes in panel A , all four Business Core Classes in panel B, and any Business Core Classes or Intermediate Microeconomics in panel C on an indicator for receiving the alumni speaker treatment conditional on the speaker's gender, professor fixed effects, semester fixed effects, and controls for race, gender, ACT/SAT score, age at entrance to college, class year during Econ 101, international student status, and state residence. Columns 1 and 3 are restricted to male students; Columns 2 and 4 are restricted to female students. The sample is drawn from administrative transcript data from UW-Madison for all students who took Econ 101 between the years 2015/16 and 2018/19, excluding eight lectures which had other experimental interventions. Standard errors clustered at the lecture level in parentheses with indicators $p < 0.10$ + $p < 0.05$ * $p < 0.01$ ** $p < 0.001$ ***; Pair-wise bootstrapped standard errors in brackets.

Similar to the analysis of business classes, we re-run the full sample specification in Table 6 with outcomes being the total number of each class type taken. Panel A of Table 8 shows the results from regressions estimated with the number of STEM classes taken and Panel B shows results from regressions estimated with the number arts, humanities, and social science classes taken.

As shown above, the alumni speakers were successful in encouraging students of the same sex to take more economics classes. However, this is at the expense of STEM classes (Panel A, col. 1 and 4). This is similar to our results from the business core classes. Specifically, we find that male alumni reduced the number of STEM classes that male students took by 0.359 courses (9.6%) and female alumni reduce the number of STEM classes that female students took by 0.209 classes (10.5%). We also found that female alumni speakers caused male students to be less likely to take Econ 102. As an alternative, this group of male students takes more STEM classes (0.234, or 6.3%), as shown in Panel B, column 3. We also found that this group (males who had a female alumni speaker) took 0.155 (3.2%) fewer arts and humanities classes overall (Panel B, col 3). Although we didn't find evidence that male alumni speakers caused female students to take more economics classes, we do find this group of students took fewer arts, humanities, and non-economics social sciences classes (Panel B, column 2).

Table 8: Effects of Total Number of Classes Taken, by Gender of Speaker

	Male Alumni		Female Alumni	
	(1)	(2)	(3)	(4)
	Male Student	Female Student	Male Student	Female Student
Panel A	-0.359**	-0.00687	0.234*	-0.209+
STEM	(0.119)	(0.111)	(0.0906)	(0.107)
Classes	[0.188]	[0.196]	[0.171]	[0.189]
Control Mean	3.71	2.88	3.71	2.88
Panel B	0.209	-0.258*	-0.155*	0.120
Arts, Humanities,	(0.161)	(0.0963)	(0.0609)	(0.142)
& Social Sciences	[0.245]	[0.203]	[0.0964]	[0.250]
Control Mean	4.70	4.85	4.70	4.85
Professor FE	Y	Y	Y	Y
Semester FE	Y	Y	Y	Y
Covariates	Y	Y	Y	Y
N Obs.	3723	2816	4035	3057
N. Clusters	21	21	21	21

Note. This table reports the results of regressions of the total number of Arts, Humanities, and Social Science classes ever taken in Panel A, and the total number of STEM classes ever taken in Panel B on an indicator for receiving the alumni speaker treatment conditional on the speaker's gender, professor fixed effects, semester fixed effects, and controls for race, gender, ACT/SAT score, age at entrance to college, class year during Econ 101, international student status, and state residence. Columns 1 and 3 are restricted to male students; Columns 2 and 4 are restricted to female students. The sample is drawn from administrative transcript data from UW-Madison for all students who took Econ 101 between the years 2015/16 and 2018/19, excluding eight lectures which had other experimental interventions. Standard errors clustered at the lecture level in parentheses with indicators $p < 0.10$ + $p < 0.05$ * $p < 0.01$ ** $p < 0.001$ ***; Pair-wise bootstrapped standard errors in brackets.

Overall, these findings suggest that many of students drawn into an economics major by the intervention would likely have otherwise majored in business or taken STEM classes. This differs from the findings from Porter and Serra (2020)'s experiment, in which they found that female alumni speakers primarily induced humanities students to switch to majoring in economics. One explanation for our different findings may be that the career paths of our speakers were more traditional business career paths (i.e., supply chain management and marketing). Alternatively, the students in UW-Madison's introductory economics course may be more likely to be potential business majors or STEM majors than those in the setting of Porter and Serra (2020); with the majority of Econ 101 students typically majoring in business in the absence of an intervention, it is perhaps not surprising that a little more than half the students we induce to choose economics would have been business majors. This results highlight a potential limitation of this type of intervention. The relatively high-pay of economics as a major is often cited as a reason to be concerned about low rates of female interest. If this intervention draws women from other high-paying majors, it will presumably have a limited impact on post-college earnings gaps driven by gender differences in major choice.

7.5 Heterogeneous Effects

In addition to gender differences in the impacts of this intervention, there may be other demographic groups that are more or less impacted by a treatment such as this. For example, all speakers were White and worked in Wisconsin, meaning that if homophily applies across multiple dimensions, we might expect stronger impacts of such speakers on White students or in-state residents. In addition, it is important to understand which types of students are on the margins when it comes to major choice, i.e., the subgroups for whom the alumni speaker is most effective in increasing interest in economics. In this section, we discuss the heterogeneity in the effects of the alumni speaker treatment. We focus on heterogeneity by race, year in school, state residency status, and pre-college math test scores. For all heterogeneity analyses, we use our first primary specification which includes all treated and control lectures rather than the single-year control group.[18]

Figure 2 shows the effects of the alumni treatment by race, class year, and residency on take-up of intermediate microeconomics, for male students (left) and female students (right). Panel A reports results for both men and women. Panel B and C report results overall; Panel D and E report results for the male role model; Panel F and G report results for the female role model. The effect sizes are consistent with the main results: for most groups, there is a significant positive effects on take-up of Econ 301.[19]

For Econ 301, we find the largest differential effects of the treatment for White students relative to non-White students. We test whether coefficients are equal across sub-groups and cannot reject the null that the effects are equal by class year or residency, but can by race. In regressions for the full sample, the

[18]Heterogeneity results are similar in the second specification but more imprecisely estimated due to smaller sample size in the sub-groups.

[19]We report results for Econ 102 in Appendix Figure A-1. We cannot reject the null that impacts differ by any sub-group for take-up of Econ 102

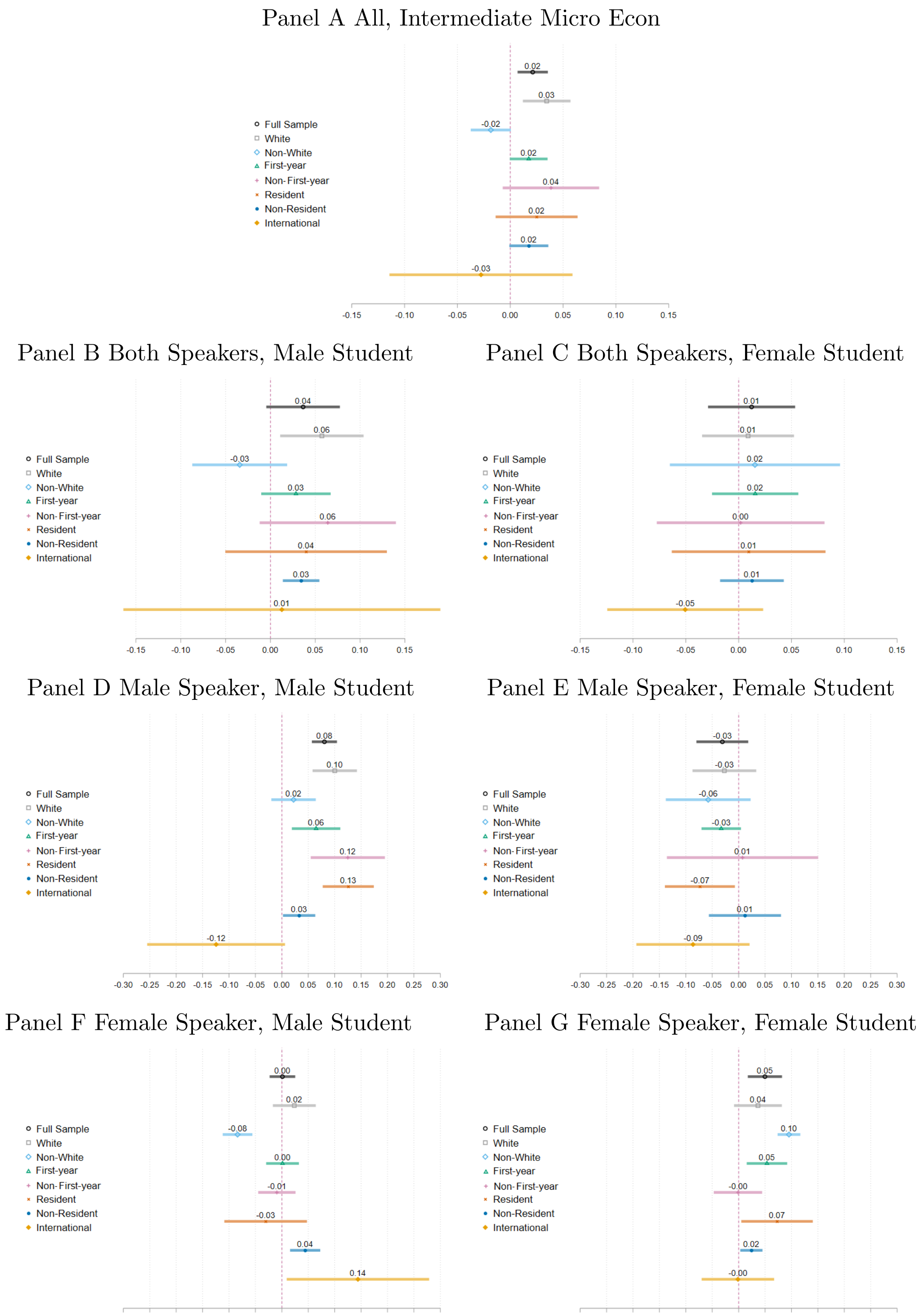

Figure 2: Heterogeneous effects of receiving the alumni speaker treatment

Notes. This figure reports the coefficient from a regression of Econ 301 take-up on an indicator for receiving the alumni speaker treatment for various subsamples of student. Regressions also include an indicator for receiving the alumni speaker treatment, professor fixed effects, semester fixed effects, and controls for race, gender, ACT/SAT score, age at entrance to college, class year during Econ 101, international student status, and state residence. The sample is drawn from administrative transcript data from UW-Madison for all students who took Econ 101 between the years 2015/16 and 2018/19, excluding eight lectures which had other experimental interventions. 95 percent confidence intervals shown; standard errors clustered at the lecture level.

treatment increased take-up of Econ 301 for White students by 4 percentage points on average relative to a non-significant decline in take-up for non-White students.[20] The difference is coefficients is significant at a $p < 0.001$ level in a t-test of equality of coefficients (Clogg et al., 1995). As shown in Figure 2, this effect is clearly driven primarily by men for whom there is a strong positive impact of the treatment for white men (6 percentage points increase) and a negative, though non-significant effect for non-White men (-3 percentage points). For women, there is no significant difference in treatment impacts by race. For other demographic groups (first-year vs. non-first-year, state resident vs. non-state resident, high vs. low math scores), we cannot reject the null that the effects of the treatment are the same across group for Econ 301. Split by gender of the speaker, it is clearer that the racial differences in effects for men are caused by two different, but reinforcing effects. While the male speaker caused no effect for non-White men and positive significant effects for White men, the female speaker is associated with negative significant effects for non-White men and no effects for White men. In both cases, we can reject the null that coefficients are equal for White and non-White students. However, for female students, we see that the female speaker increased take-up for non-White women significantly more than for White women. Notably, prior to the treatment, non-White women were more likely to continue with economics, suggesting they may be closer to the margin in terms of interest.

Next, we explore heterogeneity based on student ability. We use pre-college test scores instead of grades in Econ 101 due to concerns that performance in Econ 101 could be impacted by the treatment itself, either through impacting effort of the student or by reducing learning time in the class. Appendix section A.1 discusses impacts of the intervention on performance in both Econ 101 and in later economics courses; there is some suggestive evidence that the female speaker is associated with increased performance in Econ 101 and future economics classes whereas the male speaker has no impact on performance in Econ 101 but is associated with lower grades for male students in later economics courses.

In our sample, we observe ACT scores for 7302 students, SAT scores for 935 students, and no pre-college test scores for the remainder of the sample.[21] For the math component of the ACT, we calculate the within-sample percentiles and define a 'high-scoring' student as one in the top tercile of the ACT within the sample which corresponds to a score of 30 or the 95th percentile of the national distribution. To allow for comparability with the SAT, we define students who took the SAT as 'high-scoring' if they scored at the same percentile within the national distribution, which corresponds to a 730 on the math SAT.[22] For students who took both tests, we assign the tercile based on the higher percentile (i.e., if a student was in the 70th percentile on the ACT but scored below 730, they would be assigned to the top tercile based on

[20]Note that due to the demographics of University of Wisconsin-Madison as a whole and Econ 101 in particular, non-White students are primarily Asian and Pacific Islander American students, rather than Black or Hispanic students.

[21]53 percent of students missing test-scores are international students and 93 percent are non-Wisconsin residents. We omit all international students from the test score heterogeneity analyses for this reason.

[22]Results are similar if we instead use the 67th percentile of SAT scores within the sample which corresponds to a score of 750.

their ACT score).

Table 9 reports the effects of the alumni treatment by math skills and gender on take-up of intermediate microeconomics for both speakers (Panel A), the male speaker (Panel B), and the female speaker (Panel C). In the overall sample, we see a stronger positive impact of the speaker on high skill male students and low skill female students, though we cannot reject the null that the effects are equal across groups. When split by speaker, we can see that this is driven by interactions between speaker gender and skill. For male students, both speakers had positive impacts on take-up for high-skill students but there was only a significant positive impact on take-up of the male speaker on low-skill male speakers. We can reject that the female and male speaker effects are equal for low-skill male students at the $p < 0.001$ level. For women, the opposite is true: while both speakers had a positive impact on take-up for low-skill students, the male speaker had strong negative impacts on high skill female students wheras the female speaker increased take-up. We can reject that the effects of the male speaker are equal for high and low skill female students at the $p < 0.05$ level and can reject that effects of male and female speakers are equal for high-skill female students as well. This suggests that same-gender speakers impact men and women differently by skill level, with gendered effects being stronger for less-skilled men and for more-skilled women. This is also consistent with the selection patterns suggested by the effects of the treatment on performance in later economics courses: the female speaker is associated with marginally higher performance in later classes for women, and the male speaker is associated with lower performance in later classes for men (see Appendix Table A-1).

Table 9: Effects of Treatment on Econ 301 by Math Scores, Overall

	(1) Male High Math	(2) Male Low Math	(3) Female High Math	(4) Female Low Math
Panel A: Both Speakers	0.0737*	0.0499	-0.0232	0.0465^{+}
	(0.0268)	(0.0352)	(0.0652)	(0.0241)
Num. Obs.	1611	1855	792	1836
Panel B: Male Speaker	0.113*	0.124***	-0.149*	0.0321^{b}
	(0.0468)	(0.0181)	(0.0696)	(0.0461)
Num. Obs.	1390	1618	709	1590
Panel C: Female Speaker	0.0454*	-0.0153	0.104^{+}	0.0578***
	(0.0162)	(0.0264)	(0.0519)	(0.0122)
Num. Obs.	1518	1696	739	1694
Professor FE	Y	Y	Y	Y
Semester FE	Y	Y	Y	Y
Covariates	Y	Y	Y	Y
P-Value, $H0 : \beta_F = \beta_M$	0.174	<0.001	0.003	0.596

Note. This reports the results of regressions of taking intermediate micro (Econ 301) on an indicator for receiving the alumni speaker treatment by student gender and math skills, where 'High Math' indicates an ACT/SAT score in the top tercile. All regressions include professor fixed effects, semester fixed effects, and controls for race, gender (in full sample), ACT/SAT score, age at entrance to college, class year during Econ 101, international student status, and state residence. The sample is drawn from administrative transcript data from UW-Madison for all students who took Econ 101 between the years 2015/16 and 2018/19, excluding eight lectures which had other experimental interventions. Standard errors clustered at the lecture level; Significance for H0: $\beta = 0$ given by $p < 0.10$ + $p < 0.05$ * $p < 0.01$ ** $p < 0.001$ ***. Significance for H0: $\beta_{high} = \beta_{low}$ within gender given by $p < 0.10$ a $p < 0.05$ b $p < 0.01$ c $p < 0.001$ d. We also report the p-value for the t-test of equality of coefficients within skill group across gender of the role model, $\beta_F = \beta_M$.

8 Robustness and Alternative Specifications

We chose our primary sample to reduce differences in observed student characteristics across our treatment and control group, as well as to avoid including lectures in which alternative experimental treatments were taking place.[23] However, there are alternative samples we could have chosen as our control semesters, each with their own pros and cons. We therefore estimate a number of different specifications to ensure that our estimates are robust to these different dimensions of potential bias. Our robustness checks can be categorized in two ways: i) robustness to alternative mechanisms and ii) robustness to sample selection based on characteristics of the students.

8.1 Alternative Mechanism: Semester-Specific Effects

Thus far, we have framed the prior results as indicating that gender of the speaker matters: female speakers increase interest in economics for female students and vice versa for male speakers. However, because the female speakers both presented in the Fall semester and the male speaker in the Spring semester, an alternative explanation for these results is that female students who take Econ 101 in the Fall are more susceptible to alumni speakers whereas Spring semester male students are more susceptible. Alternatively, because the control group professor differs in the Fall and the Spring, the results could be attributable to an interaction between professor and semester. Notably, in the Fall semester, first-year students are unlikely to have any knowledge about the teaching style of professors and cannot as easily select into a certain professor's class based on their skills or interest in economics. In the Spring, however, they may select into a particular professor on the basis of that professor's reputation (e.g., a more difficult class). If female students less likely to continue on in economics select into the treatment professor's classes and vice versa for male students, this could explain our results rather than homophily.

To address these concerns, we do two exercises. First, we define treatment lectures based not just on professor, but on professor crossed with semester. First, we run a specification in which we replace the professor fixed effects with professor X [Fall, Spring] fixed effects, allowing the effect of a professor to vary by time of year. This addresses concerns that it is not just a professor who influences likelihood of continuation with economics, but that professors differentially influence likelihood in Fall versus Spring. Panel A of Table 10 reports the results of this regression; the results are comparable in magnitude to the primary two specifications.

Next, we explore pre-trends in course-take-up. If it is the case that students who take economics in Fall versus Spring are differentially interested in continuing with economics, we would see these differences in pre-treatment semesters. One downside to our two-way fixed effects model is that all professors with lectures in the control group are also at some point in the treatment group, making it difficult to cleanly examine pre-

[23]A separate experiment involving letters of encouragement was run in two lectures in Fall 2015 and two lectures in Spring 2016. Our pilot was run in Spring 2018. We therefore omit these eight lectures from all analyses.

Table 10: Robustness Checks: Semester-Specific Effects

	(1) Overall	(2) Men	(3) Women	(4) Men, Male RM	(5) Women, Male RM	(6) Men, Female RM	(7) Women, Female RM
	0.0213**	0.0365+	0.0123	0.0806***	-0.0308	0.0011	0.0498**
Original	(0.0069)	(0.0198)	(0.0199)	(0.0114)	(0.0235)	(0.0116)	(0.0155)
N	7099	4034	3065	3399	2620	3681	2826
Panel A:	0.0071	0.0197	0.0084	0.0802***	-0.0639*	-0.0112*	0.0506**
Prof. X Semester FE	(0.0087)	(0.0193)	(0.0254)	(0.0193)	(0.0255)	(0.0053)	(0.0164)
N	7729	4415	3314	3723	2816	4035	3057
Panel B:	0.0265*	0.0582***	0.0174	0.0537***	0.0255*	0.00332	0.0163+
Event Study	(0.0102)	(0.0105)	(0.0147)	(0.0098)	(0.0107)	(0.0010)	(0.0093)
N	7028	3996	3032	3304	2534	3846	2928

Note. This table reports the results of regressions of taking intermediate micro (Econ 301) on an indicator for receiving the alumni speaker treatment, professor fixed effects, semester fixed effects, and controls for race, gender, ACT/SAT score, age at entrance to college, class year during Econ 101, international student status, and state residence. Each panel uses different sample selection criteria described in-text. All analyses exclude eight lectures which had other experimental interventions during the study period. Standard errors clustered at the lecture level in parentheses; $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

treatment patterns relative to post-treatment. We can use the professor crossed with semester specification to address this shortcoming. We have four professors in our treatment group: Professor 1 is treated in Fall, Professor 2 is treated in Fall and not in Spring, Professor 3 has one lecture treated in Fall and one control in Fall and two lectures treated in Spring, and Professor 4 is treated in Spring. If we define Treated as Professor 1, Professor 2 X Fall, Professor 3 X Spring, and Professor 4, we can look at the pre-trends in 'treated' vs 'control' semesters. Specifically, the 'treatment' group includes all courses between 2015-16 and 2018-19 taught by Professor 1 and Professor 4, all fall courses taught by Professor 2, and all spring courses taught by Professor 3. We drop the one lecture that is treated for Professor 3 in a fall semester. This then allows us to create a 'never-treated' control group containing all spring courses taught by Professor 2, all fall courses taught by Professor 3, and the course taught by Professor 5.

Figure 3 plots the average proportion of students in each semester for these groups for the full sample, estimated by regressing take-up of Econ 301 on indicators by semester separately by treatment group without the inclusion of additional controls. Panel B of Table 10 reports the post-period average effects from the event study specification, in which we re-run our main specification using these definitions of treatment and control including the full set of controls including professor FE and semester-year FE. In the overall sample, we cannot reject the null that the pre-trends are equivalent. Looking at the post-period average effects, they look broadly similar to the main specification.

These results are suggestive that semester of the course is not the primary driver of future course-taking behavior. Nonetheless, these analyses cannot fully rule out that there is some unobservable characteristic about male students and female students that make them susceptible in different ways to alumni speakers during different semesters.

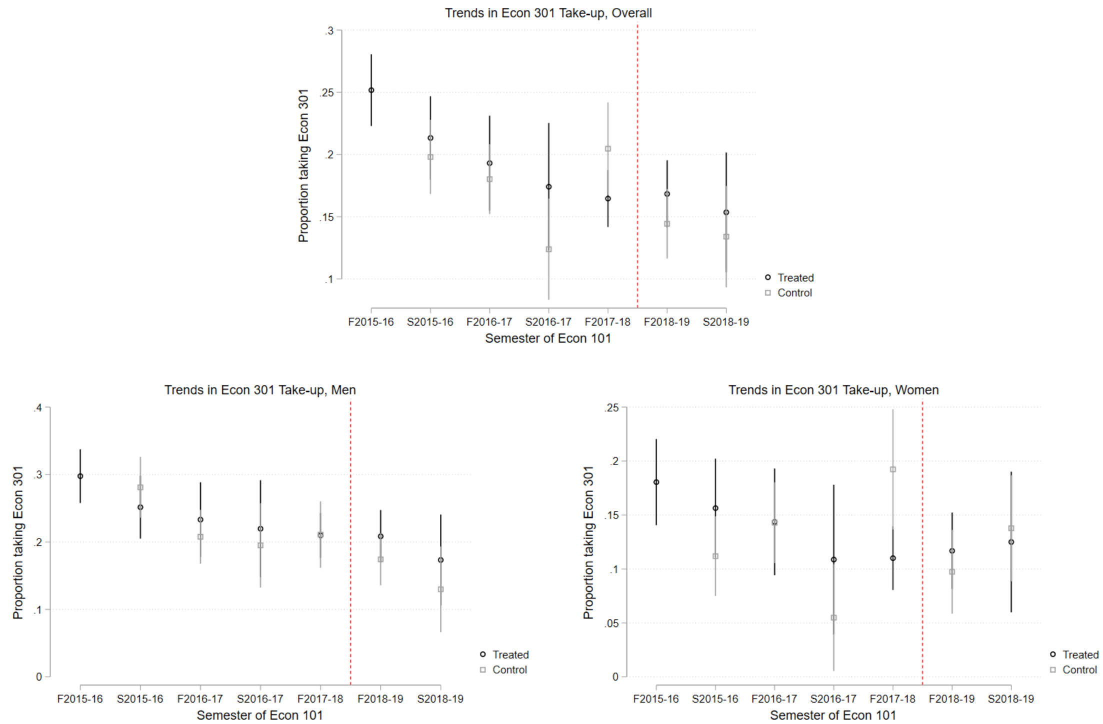

Figure 3: Time Trends in Econ 301 Take-up, By Treatment and Control
Notes: This figure plots the average proportion of students who go on to take Econ 301 in the 'Treatment' and 'Control' group as defined in the robustness checks, as estimated by regressing an indicator for taking Econ 301 on indicators for semester separately by treatment and control.

8.2 Alternative Samples

In the second set of robustness checks, we vary the student demographic characteristics used to create the control group. In our primary specifications, we rely on the assignment of treatment by lecture and the plausibly random selection of students into each lecture to justify our identification assumptions. We can relax this assumption by testing specifications in which we match on observable traits of the students.

Specifically, we use two methods which adjust for differences in observables. To address concerns that the treated and control group differ on observable characteristics, one alternative for balancing our samples on observables is to reweight observations using a inverse propensity score weighting method (Hirano et al., 2003; Fortin et al., 2011). A second alternative is to use nearest neighbor matching on student observables and professor fixed effects Abadie et al. (2004). More detail on these specifications is reported in Online Appendix Section A.2 and results from these regressions are reported in Panel A and Panel B, respectively of Appendix Table A-3. In both specifications, we still see strong positive and statistically significant effects of same-gender speakers.

Lastly, we run a series of specifications removing demographic groups that are less likely to be impacted by the treatment. These non-marginal students should not contribute much to estimate of the treatment as they are unaffected, but will add noise to the estimates of our effects. For example, students who fail Econ 101 are unable to take the intermediate courses unless they re-take Econ 101; they are thus almost certain to have zeroes for the outcome regardless of treatment and thus reduce the precision of our estimates. Panel C of Appendix Table A-3. removes international students who have substantively different take-up patterns from the majority of UW-Madison students. Panel D removes students who failed the course (D or F grades). Panel E removes both international students and those who failed the course. Panel F removes juniors who are more likely to be taking the course as a distributional requirement and thus not be responsive to the treatment. Across all specifications, the results are of similar magnitude to the primary specifications.

9 Conclusion

This paper adds to the growing literature that tries to both understand the gender gap in interest in economics courses and intervene to close this gap. We explore whether exposure to same-gender alumni speakers increases interest in taking additional economics course work, extending past research on the topic by introducing variation in the gender of the speakers. This allows us to understand if speakers increase interest merely by engaging students about real-life applications to an economics major – in which case we would expect speakers of any gender to have similar impacts – or if there are instead gender-specific impacts of the speaker. We confirm past findings that alumni speakers increase interest in economics courses, showing that the overall effect of our intervention was to increase take-up of intermediate economics courses by 1.7-2.1 percentage points. We then show that these effects are masking heterogeneous effects by gender of the speaker. We can reject that male and female speakers have similar effects on students' interest and instead show that students respond positively to same-gender speakers, but are unaffected or even possibly deterred from future economics course-work by opposite-gender speakers.

While we can rule out that the impact of speakers works through a non-gendered channel such as information about job types, we cannot speak to *why* a same-gender speaker is more effective. Are these effects due to a 'role model' effect in which students see someone of their demographic group succeeding in economics and are inspired to continue in the field? Does seeing the success of someone similar to oneself increase self-efficacy, improving students' belief that they can succeed in economics courses? Are alumni speakers providing information that is in some way gendered? While we chose speakers who worked in the same firm and gave them the same guidelines on what to speak about, we did not restrict the topics they discussed, meaning that they may have discussed topics that were more relevant to students of the same-gender. Future work should explore further the mechanisms through which same-gendered speakers influence students.

Our results also suggest that efforts to address the gender gap in economics should consider what demographic or ability groups might be most responsive to their interventions. Our results show that the gender gap in economics interest is largest for lower-skilled students, and the alumni speaker intervention differs in it's impact on high and low skill students as measured by pre-college standardized test scores. A number of recent 'nudge'-style interventions have used nudges targeted at students in the upper half of the grade distribution (e.g., letters of encouragement to students with a B or higher in the course) and found that these efforts had non-significant or counterproductive effects on the gender gap in economics (Halim et al., 2022; Bedard et al., 2021; Antman et al., 2020; Chambers et al., 2021). Our analyses suggest that given where in the skill distribution these gender gaps are, interventions must either increase the number of lower skilled women who want to continue on in economics or reduce the number of lower skilled men who major in economics.[24] Future interventions to target the gender gap in economics should consider the effectiveness of universal versus targeted interventions, both in terms of whether they will impact the students who are on the margin of choosing economics as a major and in terms of how they may reinforce or work against existing gender gaps conditional on skill levels.

[24] That said, it is not a priori clear that it would be desirable for women to act as men do and disregard grades as signals of success in a field. To the extent to which men are over-confident in their ability within the field, interventions meant to encourage lower-scoring men to find majors that are better match for their skills are an alternative approach.

References

Abadie, A., D. Drukker, J. L. Herr, and G. W. Imbens (2004). Implementing matching estimators for average treatment effects in stata. *The stata journal 4*(3), 290–311.

Antman, F. M., N. E. Flores, and E. Skoy (2020). Can better information reduce college gender gaps? the impact of relative grade signals on academic outcomes for students in introductory economics.

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

Bayer, A., S. P. Bhanot, and F. Lozano (2019). Does simple information provision lead to more diverse classrooms? evidence from a field experiment on undergraduate economics. In *AEA Papers and Proceedings*, Volume 109, pp. 110–14.

Bayer, A. and C. E. Rouse (2016). Diversity in the economics profession: A new attack on an old problem. *Journal of Economic Perspectives 30*(4), 221–42.

Bedard, K., J. Dodd, and S. Lundberg (2021). Can positive feedback encourage female and minority undergraduates into economics? *AEA Papers and Proceedings 111*, 128–32.

Bettinger, E. P. and B. T. Long (2005, May). Do faculty serve as role models? the impact of instructor gender on female students. *American Economic Review 95*(2), 152–157.

Black, D. A., S. Sanders, and L. Taylor (2003). The economic reward for studying economics. *Economic Inquiry 41*(3), 365–377.

Bleemer, Z. and A. Mehta (2022). Will studying economics make you rich? a regression discontinuity analysis of the returns to college major. *American Economic Journal: Applied Economics 14*(2), 1–22.

Bordón, P., C. Canals, and A. Mizala (2020). The gender gap in college major choice in chile. *Economics of Education Review 77*, 102011.

Breda, T., J. Grenet, M. Monnet, and C. Van Effenterre (2023). How effective are female role models in steering girls towards stem? evidence from french high schools. *The Economic Journal 133*(653), 1773–1809.

Bronson, M. A. (2014). Degrees are forever: Marriage, educational investment, and lifecycle labor decisions of men and women. *Unpublished manuscript 2*.

Buser, T., M. Niederle, and H. Oosterbeek (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics 129*(3), 1409–1447.

Canes, B. J. and H. S. Rosen (1995). Following in her footsteps? faculty gender composition and women's choices of college majors. *ILR Review 48*(3), 486–504.

Carlson, D., A. Schmidt, S. Souders, and B. Wolfe (2022). The effects of need-based financial aid on employment and earnings: Experimental evidence from the fund for wisconsin scholars. *Journal of Human Resources*, 0121–11458R1.

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

Chambers, A., S. Dickert-Conlin, C. Elder, S. J. Haider, and S. Imberman (2021). Info. econ: Increasing diversity among economics majors. *AEA Papers and Proceedings 111*, 133–37.

Clogg, C. C., E. Petkova, and A. Haritou (1995). Statistical methods for comparing regression coefficients between models. *American journal of sociology 100*(5), 1261–1293.

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika 96*(1), 187–199.

Fairlie, R. W., F. Hoffmann, and P. Oreopoulos (2014, August). A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review 104*(8), 2567–91.

Fortin, N., T. Lemieux, and S. Firpo (2011). Decomposition methods in economics. In *Handbook of labor economics*, Volume 4, pp. 1–102. Elsevier.

Gemici, A. and M. Wiswall (2014, 2). Evolution of gender differences in post-secondary human capital investments: College majors. *International Economic Review 55*(1), 23–56.

Goldin, C. (2015). Gender and the undergraduate economics major: Notes on the undergraduate economics major at a highly selective liberal arts college. *manuscript, April 12.*

Halim, D., E. T. Powers, and R. Thornton (2022). Gender differences in economics course-taking and majoring: Findings from an rct. In *AEA Papers and Proceedings*, Volume 112, pp. 597–602.

Hershbein, B. and M. Kearney (2014). Major Decisions: What Graduates earn over their lifetimes.

Hirano, K., G. W. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica 71*(4), 1161–1189.

Hoffmann, F. and P. Oreopoulos (2009). A professor like me: The influence of instructor gender on college achievement. *Journal of human resources 44*(2), 479–494.

Kato, T. and Y. Song (2022). Advising, gender, and performance: Evidence from a university with exogenous adviser–student gender match. *Economic Inquiry 60*(1), 121–141.

Kofoed, M. S. et al. (2019). The effect of same-gender or same-race role models on occupation choice evidence from randomly assigned mentors at west point. *Journal of Human Resources 54*(2), 430–467.

Kugler, A. D., C. H. Tinsley, and O. Ukhaneva (2021). Choice of majors: Are women really different from men? *Economics of Education Review 81*, 102079.

Li, H.-H. (2018). Do mentoring, information, and nudge reduce the gender gap in economics majors? *Economics of Education Review 64*(C), 165–183.

Lim, J. and J. Meer (2017). The impact of teacher–student gender matches random assignment evidence from south korea. *Journal of Human Resources 52*(4), 979–997.

Lundberg, S. and J. Stearns (2019, February). Women in economics: Stalled progress. *Journal of Economic Perspectives 33*(1), 3–22.

McEwan, P. J., S. Rogers, and A. Weerapana (2021). Grade sensitivity and the economics major at a women's college. *AEA Papers and Proceedings 111*, 102–06.

Owen, S. (2021). Ahead of the curve: Grade signals, gender, and college major choice. *unpublished manuscript*.

Owen, S. (2023). College major choice and beliefs about relative performance: An experimental intervention to understand gender gaps in stem. *Economics of Education Review 97*, 102479.

Patnaik, A. (2021). *Pricing, Income and College Major Choice*. Ph. D. thesis, University of Wisconsin-Madison.

Patnaik, A., J. Venator, M. Wiswall, and B. Zafar (2020). The role of heterogeneous risk preferences, discount rates, and earnings expectations in college major choice. *Journal of Econometrics*.

Patnaik, A., M. J. Wiswall, and B. Zafar (2020, August). College majors. Working Paper 27645, National Bureau of Economic Research.

Porter, C. and D. Serra (2020). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics 12*(3), 226–254.

Pugatch, T. and E. Schroeder (2021). Promoting female interest in economics: Limits to nudges. *AEA Papers and Proceedings 111*, 123–27.

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

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

Reuben, E., M. Wiswall, and B. Zafar (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal 127*(604), 2153–2186.

Saltiel, F. (2023). Multi-dimensional skills and gender differences in stem majors. *The Economic Journal 133*(651), 1217–1247.

UW-Madison (2024). Areas of study: College of letters & science *https://ls.wisc.edu/areas-of-study*.

Wasserman, M. (2023). Hours constraints, occupational choice, and gender: Evidence from medical residents.

Wiswall, M. and B. Zafar (2018). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics 133*(1), 457–507.

Wiswall, M. and B. Zafar (2021). Human capital investments and expectations about career and family. *Journal of Political Economy 129*(5), 1361–1424.

Zafar, B. (2013). College Major Choice and the Gender Gap. *Journal of Human Resources 48*(3), 545–595.

A Additional Analysis

A.1 Effects on Grade Performance

In addition to take up on economics courses, we also study the effects of these interventions on grade performance, where grades are measured on a 4.0 scale with a 4.0 indicating an A, 3.0 indicating a B, etc. This helps us understand whether these interventions might have an effect on the performance of students. While these interventions may increase the motivation to perform well in the course, it also takes time out of class that otherwise would be used for conveying course material. This may result in a direct effect of the intervention on grade performance for the introductory economics course. For later courses, the effects are not just due to the direct effect of the intervention on student's work effort or knowledge from introductory economics, but also includes a selection component due to compositional changes in who takes economics. A counter argument to having these interventions is that these interventions may push or nudge people into fields which may not be optimal for them. Earlier, we showed that the intervention reduced the gender gap in course-take up for those in the bottom and middle tercile of ACT/SAT math scores. This may mean that the intervention encourages lower performing individuals to continue on in economics, resulting in lower performance in later courses.

Table A-1 provides the estimates for grade performance in economics courses. We do not see evidence of the treatment resulting in lower performance, however, with the treatment having no economically significant effect on performance in upper-level courses. There is no effect of the treatment on performance in Econ 101, as we'd expect given that a fifteen minute talk should not substantively changing the amount of course material that students were exposed to or their skill levels in economics. We also do not see any large effects on performance in later courses. While there is a statistically significant positive impact of the intervention on women's grades in Econ 102, the magnitude is small and not significant in a practical sense– equivalent to 0.08 GPA points or one-sixth of a grade level. There is a small, non-significant negative relationship between the intervention and grades in Econ 301, but even if these effects were not noisily estimated, the magnitudes are small.

When we split the effects on grade by gender of speaker, we see that the male speaker is associated with lower grades in later courses and the female speaker is associated with higher grades in later courses, though the effects are still small in magnitude. Table A-2 reports the same regressions as previously, split by gender of the speaker. We see that the male speaker is associated with a -0.256 lower grade in Econ 301 (equivalent to half a grade level) whereas the female speaker is associated with a marginally significant 0.163 higher grade in Econ 301 for female students (equivalent to one-third of a grade level). The selection patterns into Econ 102 are similar, though of smaller magnitude.

There remains no effect on performance in Econ 101 in response to the treatment, suggesting that these gendered grade effects are indicative of sorting on ability in response to the treatment rather than the speakers impacting student's ability to succeed in economics. These results are consistent with the heterogeneity patterns by pre-college math scores in which the female role model had a greater impact on high-achieving students whereas the male role model had a greater impact on those in the middle or the bottom of the skill distribution.

Table A-1: Effect of the Treatment on Grades

	(1) All	(2) Men	(3) Women
Panel A: Econ 101	0.0157	0.00110	0.0404
	(0.0240)	(0.0234)	(0.0398)
N	7729	4415	3314
Panel B: Econ 102	0.0148	-0.0126	0.0801**
	(0.0444)	(0.0589)	(0.0257)
N	3600	2200	1400
Panel C: Econ 301	-0.102	-0.100	-0.0437
	(0.0608)	(0.0881)	(0.0947)
N	1367	957	410

Note. This reports the results of regressions of grade points in introductory micro (Panel A: Econ 102), intro macro (Panel B: Econ 102), and intermediate micro (Panel C: Econ 301) on an indicator for receiving the alumni speaker treatment, professor fixed effects, semester fixed effects, and controls for race, gender, ACT/SAT score, age at entrance to college, class year during Econ 101, international student status, and state residence. The sample is drawn from administrative transcript data from UW-Madison for all students who took Econ 101 between the years 2015/16 and 2018/19, excluding eight lectures which had other experimental interventions.Standard errors clustered at the lecture level; $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A-2: Effects on Grades by Gender of the Speaker

	Male Alumni		Female Alumni	
	(1) Male Student	(2) Female Student	(3) Male Student	(4) Female Student
Panel A: Econ 101	0.0232	-0.0181	-0.0229	0.0636+
	(0.0330)	(0.0765)	(0.0340)	(0.0337)
N	3723	2816	4035	3057
Panel B: Econ 102	-0.174*	0.0170	0.0902*	0.107+
	(0.0705)	(0.0416)	(0.0346)	(0.0593)
N	1863	1189	2029	1294
Panel C: Econ 301	-0.256**	-0.253*	0.0160	0.163+
	(0.0870)	(0.111)	(0.0785)	(0.0787)
N	822	353	885	381

Notes. This reports the results of regressions of grade points in introductory micro (Panel A: Econ 102), intro macro (Panel B: Econ 102), and intermediate micro (Panel C: Econ 301) on an indicator for receiving the alumni speaker treatment (separated by gender of speaker), professor fixed effects, semester fixed effects, and controls for race, gender, ACT/SAT score, age at entrance to college, class year during Econ 101, international student status, and state residence. The sample is drawn from administrative transcript data from UW-Madison for all students who took Econ 101 between the years 2015/16 and 2018/19, excluding eight lectures which had other experimental interventions.Standard errors in parentheses; $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.2 Robustness Checks: Varying the Student Composition of the Control Group

In the robustness checks described in Section 8.2, we vary the student demographic characteristics used to create the control group. In our primary specification, we rely on the assignment of treatment by lecture and the plausibly random selection of students into each lecture to justify our identification assumptions. We can relax this assumption by testing specifications in which we match on observable traits of the students.

Specifically, we calculate the following weight, W_i, for each student i grouped according to whether they were treated ($j_i = 1$) or not ($j_i = 0$):

$$W_{it} = \frac{P(j_i = T|X_i)}{P(j_i = T)} \frac{P(j_i)}{P(j_i|X_i)}$$

To calculate the probability of j_i conditional on X_i, we run a Probit regression of an indicator for being treated on demographic characteristics including gender, race, international student status, class-year during Econ 101, professor, ACT scores, and indicators for missing the ACT score and then predict the probability that a student is in that group given their observable characteristics. For the unconditional means, we take the proportion of the sample treated versus untreated. Thus, for those with $j_i = 1$ or the treated, the weight will equal one whereas for students in other lectures, they will be weighted more or less in the regression depending on their predicted likelihood of being treated given their observables.

We then ensure that we have overlap of the propensity scores using the method proposed in Crump et al. (2009) which creates a threshold for dropping units with extreme (i.e.,close to zero or one) values for the estimated propensity score. We thus omit any observations predicted to have a propensity score lower than 0.09 or greater than 0.91. Results from the sample re-weighted by propensity score are reported in Panel A of Appendix Table A-3. Though the results for the full sample are now insignificant, we still see strong positive and statistically significant effects of same-gender speakers.

In addition to exploring results when students are re-weighted to look similar on observables, we also conduct nearest neighbor matching based on the same set of student observable characteristics and professor. We use the nearest neighbor matching estimator from Abadie et al. (2004), which determines the "nearest" by using a weighted function of the covariates for each observation and the Mahalonobis distance metric. These results are reported in Panel B of Appendix Table A-3. This is the only robustness check in which we see a negative impact of the treatment overall. This is driven largely by the fact that when we use nearest neighbor matching to form the control group, we not only see strong positive effects of same gender role models, but also strong negative effects of exposure to opposite gender speakers. Thus, the broader conclusions of the main specification hold: same gender role speakers increase interest in the subject. However, they do point to the possibility that departments considering implementing such programs may have to weight the benefits of a speaker encouraging one gender's interest in the subject against the cost of losing enrollment from the other gender.

A.3 Additional Tables and Figures

Table A-3: Robustness Checks: Alternative Comparison Groups

	(1) Overall	(2) Men	(3) Women	(4) Men, Male RM	(5) Women, Male RM	(6) Men, Female RM	(7) Women, Female RM
	0.0213**	0.0365+	0.0123	0.0806***	-0.0308	0.00113	0.0498**
Original	(0.00692)	(0.0198)	(0.0199)	(0.0114)	(0.0235)	(0.0116)	(0.0155)
N	7729	4415	3314	3723	2816	4035	3057
Panel A:	0.00492	0.00573	0.0201	0.0553***	-0.0453	-0.0138	0.0519**
PSW	(0.00793)	(0.0147)	(0.0198)	(0.00938)	(0.0306)	(0.00935)	(0.0178)
N	6050	3361	2689	2716	2213	2985	2434
Panel B:	-0.0311*	-0.0359+	0.0216	0.0721+	-0.0541*	-0.0677***	0.0825***
Matching	(0.0147)	(0.0185)	(0.0160)	(0.0390)	(0.0231)	(0.0200)	(0.0202)
N	7729	4415	3314	3723	2816	4035	3057
Panel C:	0.0263**	0.0389	0.0155	0.102***	-0.0218	-0.00755	0.0473*
Omit Int'l	(0.00928)	(0.0266)	(0.0199)	(0.0135)	(0.0294)	(0.0155)	(0.0173)
N	6677	3840	2837	3267	2419	3524	2622
Panel D:	0.0235**	0.0408*	0.0122	0.0833***	-0.0319	0.00663	0.0502**
Omit D/F	(0.00699)	(0.0192)	(0.0201)	(0.0120)	(0.0233)	(0.0119)	(0.0162)
N	7582	4350	3232	3666	2740	3979	2987
Panel E:	0.0290**	0.0438	0.0159	0.106***	-0.0217	-0.00201	0.0477*
Omit Int'l and D/F	(0.00923)	(0.0263)	(0.0200)	(0.0138)	(0.0294)	(0.0162)	(0.0177)
N	6677	3840	2837	3267	2419	3524	2622
Panel F:	0.0218*	0.0317+	0.0213	0.0703***	-0.0117	0.000835	0.0491**
Freshman and Soph.	(0.00788)	(0.0181)	(0.0176)	(0.0121)	(0.0246)	(0.0136)	(0.0173)
N	7099	4034	3065	3399	2620	3681	2826

Standard errors clustered at the lecture level in parentheses; $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note. This reports the results of regressions of taking intermediate micro (Econ 301) on an indicator for receiving the alumni speaker treatment, professor fixed effects, semester fixed effects, and controls for race, gender, ACT/SAT score, age at entrance to college, class year during Econ 101, international student status (omitted if not in sample), and state residence. Each panel uses different sample selection criteria described in-text. All analyses exclude eight lectures which had other experimental interventions during the study period.

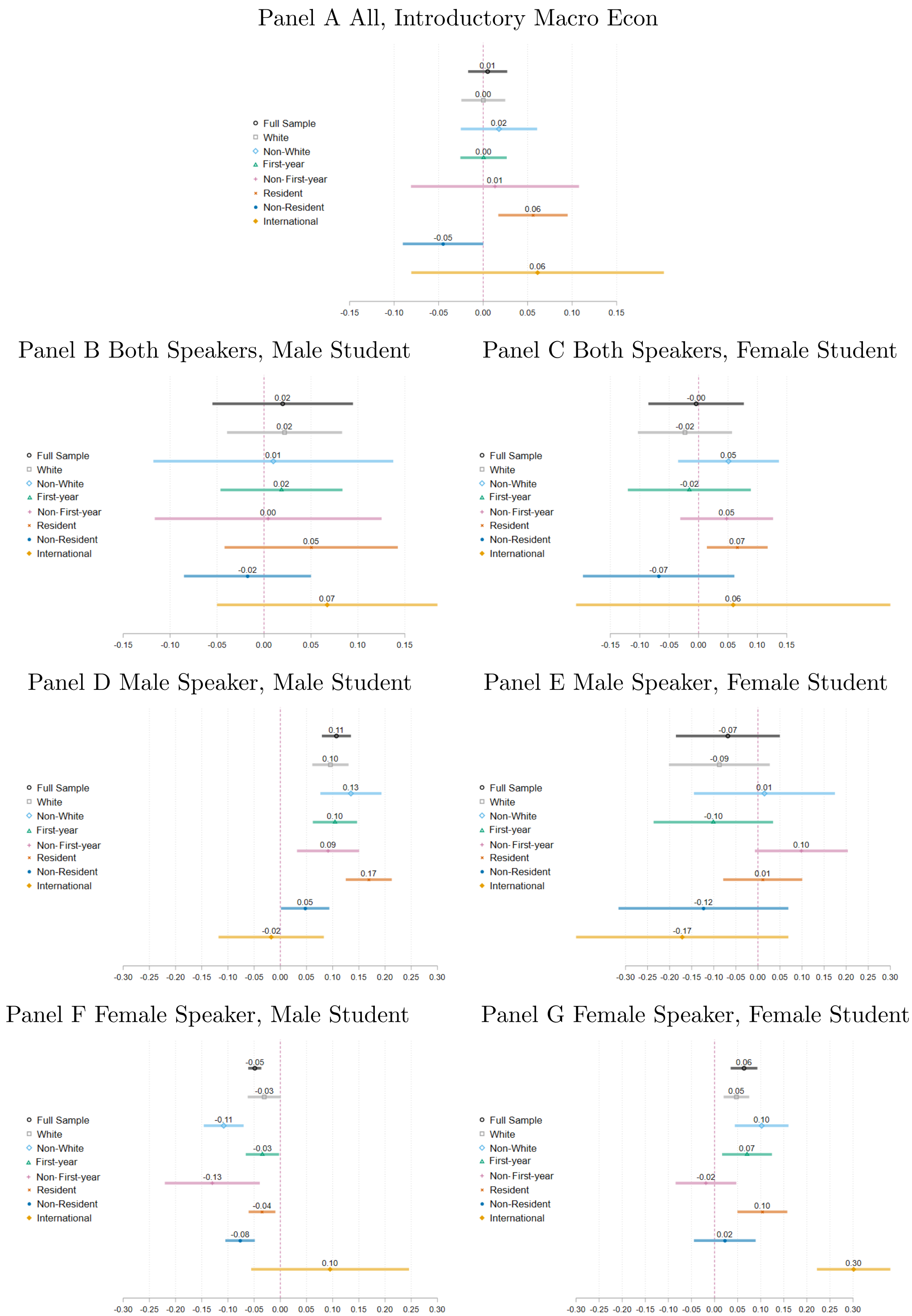

Figure A-1: Heterogeneous effects of receiving the alumni speaker treatment

Notes. This figure reports the coefficient from a regression of Econ 102 take-up on an indicator for receiving the alumni speaker treatment for various subsamples of student. Regressions also include an indicator for receiving the alumni speaker treatment, professor fixed effects, semester fixed effects, and controls for race, gender, ACT/SAT score, age at entrance to college, class year during Econ 101, international student status, and state residence. The sample is drawn from administrative transcript data from UW-Madison for all students who took Econ 101 between the years 2015/16 and 2018/19, excluding eight lectures which had other experimental interventions. 95 percent confidence intervals shown; standard errors clustered at the lecture level.