The Influence of Teaching Assistants on Student Learning

by

Craig Epstein

Submitted to the Distinguished Majors Program
Department of Economics
University of Virginia
April 30, 2021
Advisor: Kenneth G. Elzinga

Abstract

I exploit the forced transition from in-person to online learning at the University of Virginia caused by the COVID-19 pandemic to study the influence of Teaching Assistant and student demographics on student learning outcomes and course evaluations. Data from two semesters of Principles of Microeconomics reveal that male students benefitted from having a male TA in both in-person and online settings. Female students benefitted from having a female TA only in the online environment. Improved course evaluations for female TAs in the online environment suggest that female students may have been aware of the significant boost in grades which female TAs provided.

Acknowledgments

I would like to thank my thesis advisor, Professor Kenneth G. Elzinga, and Daniel Harper for their guidance, encouragement, and feedback during the creation of this paper. Additionally, I would like to thank Professor Amalia Miller for her support and for including me in this program. Thank you to Annamarie Black for providing the data which made this project possible and to Professor William Johnson for his helpful comments. Finally, I am grateful to the DMP cohort for their motivation and to my family for their endless support.

Table of Contents

1. Introduction

The COVID-19 pandemic caused many universities to convert from in-person to online modes of instruction, but the effect of these transitions on student learning outcomes and course evaluation responses is unclear. Some students perform better when taught by instructors and teaching assistants with similar demographics, but few studies have explored this in online environments. The instructional value of teaching assistants (TAs), who often interact with students more than professors in courses with large enrollments, is seldom studied despite their widespread use. I aim to shed light on the intersection of these topics— online education, demographics, and TAs— by analyzing two semesters of student- and TA- level data from Econ 2010, Principles of Microeconomics, at the University of Virginia (UVA).

I address two questions: First, how do interactions between student and TA demographics affect student outcomes for in-person versus online learning? Second, do TAs' demographic characteristics affect their teaching evaluations? Using OLS and FGLS regression, I find that males benefitted from having a male TA in both in-person and online settings, whereas females benefitted from having a female TA in only the online setting. The median course evaluation response for the question regarding amount learned increased for female TAs in online learning, while the distribution of genders in discussion sections remained unchanged between in-person and online learning. This suggests that female students were aware of the boost in grades which their female TA provided through a role model effect and adjusted their course evaluations in response.

Part 2 reviews literature related to these two questions; part 3 describes the dataset; part 4 describes the methodology of my analysis; part 5 provides the results; part 6 offers an interpretation; and part 7 concludes.

2. Literature Review: Online Education, Role Models, and Course Evaluations

2.1: In-Person vs. Online Learning

It is important to understand the effects on student learning outcomes of universities' transitions from in-person to online education during the COVID-19 pandemic. While earlier studies regarding course grades and completion rates yield conflicting results, research on the effects of COVID-19 suggests that online teaching style plays a large role in student success as measured by scores on multiple-choice exams in intermediate economics courses at four R1 universities (Orlov et al. 2020).

No significant differences in test scores were found in undergraduate classes in personal finance, environmental science, statistics, and management or in a graduate level biostatistics class when controls for gender, class, major, and GPA were included (Ary and Brune 2011; Daymont and Blau 2008; Paul and Jefferson 2019; Summers, Waigandt, and Whittaker 2005; Hoffman and Elmi 2020). In an earth science course, only one exam out of 16 had a significantly different average between the otherwise identical in-person and online versions of the class (Werhner 2010).

These studies suffer from selection bias because students were able to choose which version of the course they enrolled in and were not randomly assigned. Perhaps no significant differences were found because students tailored their preferred method of learning to those offered. Gratton-Lavoie and Stanley (2009) addressed selection bias in their analysis of a Principles of Microeconomics class and found that final exam scores still were not different between the in-person and online versions of the class.

In contrast, a 43-person experiment in a graduate level education course found a positive effect of online learning on test performance. However, a study of nearly 500,000 students in

community or technical colleges in Washington state found a negative effect of online learning on course grades (Xu and Jaggars 2013; Liu 2005). Among the 500,000 students, Black students, underclassmen, and males saw particularly large gaps between grades in online compared to in-person courses (Xu and Jaggars 2013). Coates et al. (2004) found that students in Principles of Microeconomics courses at three universities performed worse on exams when taught in the online compared to the in-person format. Finally, universities' transitions to online education during COVID-19 had overall negative effects on students' grades, but these decreases were significantly offset by instructors who had prior experience with online teaching or included opportunities for inter-student interaction online (Orlov et al. 2020).

Course completion rates could be an important gauge of learning because students most often cite being unhappy with their grade and not understanding the material as reasons for dropping out of a class (Dunwoody and Frank 1995). Neither course completion rates nor test scores were significantly different for students randomly assigned to an in-person or online version of a psychology class (Waschull 2001). Conversely, online programs at the New Jersey Institute of Technology had lower student retention rates than in-person courses (Hiltz 1997; Atchley, Wingenbach, and Akers 2013; Carr 2000). At a U.S. research university, 88.8 percent of students completed their coursework in the in-person version of a business course compared to 57.3 percent in the online version (Patterson and McFadden 2009). Where significant results exist regarding course completion, online courses have lower completion rates than in-person courses. These studies, apart from Waschull (2001), suffer from the endogenous nature of online learning since students were able to choose the mode of learning.

2.2: Role Model Influences

Role models "(a) show us how to perform a skill and achieve a goal — they are behavioral models; (b) show us that a goal is attainable — they are representations of the possible, and (c) make a goal desirable — they are inspirations" (Morgenroth, Ryan, and Peters 2015). Undergraduate economics TAs fulfill all of these criteria. They show younger students that to achieve the position of TA, a student must do well in Principles classes. They show that undergraduates can attain this position of authority and status; and they make the position desirable either by their behavior or simply because experience as a TA offers a valuable line on a résumé and talking point during an interview. Some students respect a TA and envision themselves becoming one should they do well in their coursework.

It is possible that students perform better in classes with or simply prefer to be taught by professors and TAs they identify with in race or gender. An analysis of Tennessee's Student/Teacher Achievement Ratio program found that Black students in kindergarten through third grade with a Black teacher were 7% more likely to go to college and 13% more likely to graduate from high school than those without a Black teacher (Gershenson et al. 2018). These role model effects appear to benefit children of middle school age as well. A two-year study of 12-14 year old students showed that adolescents with at least one role model they identified with in race or gender reported being more active in academic pursuits and showed higher test scores than those without said role model (Zirkel 2002). Dee (2007) studied the role of same sex teacher-student dynamics among 8th graders from the National Education Longitudinal Survey of 1988 and found large, significant effects in a variety of areas: girls with a female science teacher were half as likely to say that science was not important for their future compared to those with a male teacher, and the overall effect of being taught by an opposite sex teacher was a

significant decrease in test scores by 0.042 standard deviations. Role model effects may be stronger for historically underrepresented populations; two experimental studies of college students showed that female students rated female role models more highly than male role models, whereas male students did not show a difference between in ratings according to sex (Lockwood 2006).

Whether or not these role model effects extend into college is unclear. Even if college students are able to identify a demographically similar role model, this may not lead to an enhanced female collegiate experience since faculty gender has been shown to have no effect on whether or not female students continue in their studies of economics (Robb and Robb 1999). In contrast, Rask and Bailey (2002) found that the proportion of classes in a certain major which a student takes with professors who are similar in sex or race has a significant and positive effect on the probability that a student declares said major. This effect has been supported experimentally, where female students in introductory economics courses who were exposed to females who majored in economics at the same university were 8 percentage points more likely to major in economics (Porter and Serra 2020). Female graduate student placements were not tied to the presence of a female role model in graduate school, such as a female professor or thesis advisor (Neumark and Gardecki 1998).

2.3: Teaching Assistants

While the role model effect from professors is uncertain, there is evidence that TA demographics exhibit an influence on students' performance. Female TAs in a Principles course had a positive, significant effect on the grades of the best performing students in their discussion sections (Hastedt 2008). A study at a university in California revealed that students in economics classes who are paired with TAs of the same race (grouped by Asian and non-Asian) attended

TA office hours more frequently, performed better on exams, were more likely to enroll in further economics classes, and major in economics. The authors of this study attributed their results to a role model effect— students may respond better to TAs' teaching styles when the TA matches their race (Lusher, Campbell, and Carrell 2018).

2.4: Demographic Biases and Preferences in Course Evaluations

Evidence of demographic biases in collegiate course evaluations have been documented in observational and experimental settings. At the University of Oregon, a study of six years of course evaluations found that female instructors received evaluations an average of 0.0578 points lower than male colleagues (Marcham et al. 2020), and a study at another major university found that male professors received significantly higher evaluations related to competence and effectiveness compared to female professors (Sidanius and Crane 1989; Andersen and Miller 1997). Boring (2017) found that males have an expected probability of receiving an "Excellent" leadership score 55% higher than female colleagues. In an experimental setting, male professors were rated higher on dimensions of instructor-group interaction, instructor-individual student interactions, dynamism/enthusiasm, and overall than female professors (Basow, Codos, and Martin 2013). Although the focus of bias has been on gender, discrepancies in evaluations extend to race as well. Anonymous instructor evaluations on RateMyProfessor.com were found to rate non-white faculty members lower on evaluations of overall quality, helpfulness, and clarity and higher on evaluations of ease than white professors (Reid 2010).

These findings are further broken down by looking at the gender and race of the students completing the evaluations, although these studies present conflicting results. A same-sex bias has been documented where female students rated female professors higher than male professors while male students did not exhibit any significant differences in their ratings of professors of

different sexes (Bachen, McLoughlin, and Garcia 1999). But same-sex bias for males has been documented in other studies, where males rate male professors higher than female professors on measures of leadership ability by an average of .35 points on a four-point scale (Boring 2017).

Mengel, Sauermann, and Zölitz (2019) found that male students rated female professors one-fifth of a standard deviation lower than male professors and female students rated female professors one-fourteenth of a standard deviation lower than male professors. Latino students rated white female professors as more capable than Latino and white male professors; compared to Latino students, African American students rated white female professors as less capable (Anderson and Smith 2005). In an experimental setting in South Africa, Black students rated Black lecturers lower than white lecturers (Chisadza, Nicholls, and Yitbarek 2019).

Although there is conflicting evidence in the direction of some of these effects, the takeaway is that some students, consciously or unconsciously, take the sex and race of their professor into account when filling out course evaluations. This is particularly salient because course evaluations generally affect faculty promotions and tenure decisions.

3. Data on Students, TAs, and Course Evaluations

3.1: Student and TA Data

Two rich datasets were gathered on student and TA performance and demographics from the UVA Student Information System (SIS)— one for Fall 2019 and one for Fall 2020. Data on students include each student's academic year, major(s), minor(s), Econ 2010 class section, official Econ 2010 final letter grade, grading basis, cumulative GPA, Virginia residency status, age, gender, and ethnicity. TA demographic information includes academic year, major(s), minor(s), gender, and ethnicity. Both graduate and undergraduate students can serve as TAs; around 70% of TAs are undergraduate students.

Student demographic data were then matched to official Econ 2010 grading spreadsheets, updated by a student's TA throughout the course. This dataset includes scores on the first and second midterms and final exam, the number of points earned through their discussion section, and the number of iClicker points students received from questions during lectures.

Lastly, student performance on the math section of the SAT or ACT were matched to each observation. The majority of students took the SAT, and so ACT scores were converted into SAT equivalents. Forty-three students did not have data on standardized test scores; these students were excluded from the study. Even though each student's cumulative GPA is controlled for, in prior research on introductory economics courses and grades, SAT scores have been shown to provide a useful and necessary control (Ballard and Johnson 2004; Coates et al. 2004).

Because of COVID-19, UVA offered students the option to take their courses on a Credit/General Credit/No Credit grading scheme (a letter grade of a C or above received Credit, D- through C- received General Credit, F received No Credit) instead of the typical letter grading scheme (A through F). In 2020, 18% of students chose the C/GC/NC scheme. To avoid complications of the introduction of the new grading scheme and to avoid using student letter grades, which are in part determined by the curve set at the end of the course, I calculate each student's final numeric grade based on each student's scores in the grading spreadsheets. I omit the number of points awarded to a student from the student's TA to exclude TA favoritism towards any students if present. Students' grades are weighted according to the syllabi.

These student and TA data include sensitive information and were collected with the permission of the UVA Institutional Review Board. To prevent any biases or revelation of

private information, all names and identifiers were removed and replaced with a randomly generated study ID. Table 1 shows descriptive statistics.

3.2: Course Evaluation Data

To maintain student trust and willingness to provide unfiltered answers on course evaluations, student-level course evaluations were unable to be acquired. Instead, I analyze TA-level course evaluations. In the Fall 2019 and Fall 2020 course evaluations, each question is listed with the number of students who chose each answer. For example, included with the question "The average number of hours per week that I spent outside of class preparing for Econ 2010 was:" are the number of students who answered, "less than 1 hour", "1-3 hours", "4-6 hours", and so on. This allows for analysis of the distribution of answers across TAs and across years. The questions and answer choices are seen in Tables 2 and 3.

3.3: *Ceteribus Paribus* Met and Unmet

Econ 2010 was remarkably consistent between Fall 2019 and Fall 2020. The same professor, Kenneth G. Elzinga, taught the course in both years. Mr. Elzinga has taught Econ 2010 to large audiences for over fifty years. Not only does this mean that his quality of teaching is very high, but it also minimizes the likelihood of a substantial difference in teaching quality between these two semesters. In both semesters, Econ 2010 had two sections, followed the same readings from the same edition of the same textbook, had the same format of discussion sections, had two short-answer midterms and a multiple choice final, had the same Head TA, was offered through the College of Arts and Sciences, had the same number of lectures, and had the same grading scheme. This grading scheme, dubbed "Dutch Knockout", renders a student's final grade the higher of two options: the regular weighted average grade in the class from the midterms, final exam, discussion section, and iClicker points, or only the student's final exam grade.

Discussion sections met synchronously in both semesters, though the Fall 2019 discussions met in-person whereas the Fall 2020 discussions met online.

Differences do exist between Fall 2019 and Fall 2020. Unlike in Fall 2019, students did not have a Fall Break in 2020. Fall 2019 in-person courses met, of course, synchronously, whereas the Fall 2020 version of Econ 2010 met asynchronously via lecture recordings posted online. Students had three days to watch a lecture after it was posted, and these professionally recorded lectures were made available to students at the end of the semester for review in preparation for the final exam. Finally, mid-lecture iClicker questions were replaced with end-of-lecture quizzes in Fall 2020. Fourteen of the 75 questions on the final exam differed between the two semesters. Both exams were proctored, and the three confirmed cases of cheating were removed from the dataset.

One might be concerned with the significant difference between the average GPA of students across years seen in Table 1. However, this should not change the interpretation of role-model effects. Given that the average GPA of both male and female students increased by similar amounts between years, we would expect grades in Econ 2010 to increase for all students equally, regardless of TA-student gender interactions. This does not turn out to be the case.

Additionally, there are two possible explanations for the increase in GPA not related to "smarts." First, students in Fall 2020 had the opportunity to choose the C/GC/NC option for any of their classes up to ten weeks into the semester, when they may have a reasonable assessment of their final grades. Changing those classes where students believe they will do poorly from graded to C/GC/NC would likely increase their GPA (current GPA data were collected in Spring 2021). Second, because Econ 2010 is populated largely by first-years, many of whom have not "found their major" yet, they may be taking classes which they find particularly difficult. Indeed,

second semester GPAs tend to be the lowest in a college student's education (Grove and Wasserman 2004).

Prior research on the effect of online courses on student learning outcomes have suffered from non-random student selection bias, found to incorrectly estimate effects in OLS (Olitsky and Cosgrove 2014). The COVID-19 pandemic, however, solved this problem in that students no longer had a choice between an in-person and an online course. This introduces a new problem: did students non-randomly withdraw from UVA prior to the Fall 2020 semester? Although data on students who withdrew from UVA prior to the Fall 2020 semester are unavailable, I analyze the frequencies of race and gender among the 20 students who withdrew from the class over the course of the Fall 2020 semester. I find that neither the gender nor the racial distribution of those students who withdrew deviates from the class's distribution.

Course evaluations faced changes in delivery provider and content between the two semesters. While both semesters' course evaluations were online, the website which hosted them changed. Presumably, the different platform would not influence students' responses. A change of potential concern is that the number of questions decreased from 33 to 16. Only four questions remain comparable between the two course evaluations (see Table 2). The reduction in the number of questions may have caused some students to spend more time reading and answering questions on the 2020 course evaluation than was the case in 2019, leading to more thoughtful responses.

One final concern regarding course evaluations independent of time is the students' understanding of who is being evaluated. It is clear that in some TA evaluations, students were accidentally evaluating Mr. Elzinga and not their TA, despite being able to fill out a separate course evaluation for Mr. Elzinga. There is no reason to believe that these mistakes would not be

randomly distributed throughout the TA evaluations, particularly across any metric included in this research, so this should not bias results.

4. Methodology

4.1: OLS Regression with Student Grades

The model used to understand the role of the TA includes more interaction terms with each successive specification. I begin with a simple model of student performance. I first estimate the OLS model as follows:

$$Grade_i = \beta_0 + \beta_1 GPA_i + \beta_2 SAT_i + \beta_3 OnlineSemester_i + \beta_4 VAResident_i + \beta_5 Class_i + \beta_6 GraduateTA_i + \beta_7 Female_i + \beta_8 Ethnicity_i + \beta_9 FemaleTA_i + \beta_{10} TAEthnicity_i \quad (1)$$

where i denotes each student. The coefficient β_3 provides insight into the overall effect of online learning on student grades, while the coefficients on TA demographics begin to provide insight into the role of TAs. To contribute to literature regarding historically underrepresented groups in higher education, I also replace the Ethnicity variables with a binary Underrepresented Minority (URM) variable, which is equal to 0 if a student or TA is white or Asian and is equal to 1 if a student or TA is Black or African American, Hispanic, or Multi-Race, similar to Lusher, Campbell, and Carrell (2018).

$OnlineSemester_i$ is equal to 0 for Fall 2019 and 1 for Fall 2020. $Class_i$ is a categorical variable equal to 0 for freshmen, 1 for sophomores, 2 for juniors, and 3 for seniors. $GraduateTA_i$ is equal to 0 if a student's TA is an undergraduate student and 1 if the TA is a graduate student. $Female_i$ and $FemaleTA_i$ are equal to 0 if a student or student's TA is male

and are equal to 1 if a student or student's TA is female, respectively. $Ethnicity_i$ and $TAEthnicity_i$ are categorical variables for a student's or TA's ethnicity, respectively.

The second model aims to answer the question of TA role model effects in both in-person and online learning environments. In this model, I include interactions between student and TA demographics.

$$Grade_i = \beta_0 + \beta_1 GPA_i + \beta_2 SAT_i + \beta_3 OnlineSemester_i + \beta_4 VirginiaResident_i + \beta_5 Class_i + \beta_6 GraduateTA_i + \beta_7 Female_i + \beta_8 FemaleTA_i + \beta_9 URM_i + \beta_{10} URMTA_i + \beta_{11} Female_i * FemaleTA_i + \beta_{12} StudentURM_i * URMTA_i \quad (2)$$

This model allows us to observe if female students benefitted from having a female TA through coefficients $\beta_8 + \beta_{11}$ and if URM students benefitted from having a URM TA through coefficients $\beta_{10} + \beta_{12}$.

The final specification puts in its crosshairs the crux of this paper: Did students in in-person or online learning benefit more from having a TA of similar demographics? Here, I add to the model an interaction between $Year_i$ and the final two terms of the previous model to separate the effects of student role models between in-person and online learning:

$$Grade_i = \beta_0 + \beta_1 GPA_i + \beta_2 SAT_i + \beta_3 Year_i + \beta_4 Residency_i + \beta_5 Class_i + \beta_6 GraduateTA_i + \beta_7 Female_i + \beta_8 FemaleTA_i + \beta_9 StudentURM_i + \beta_{10} URMTA_i + \beta_{11} GraduateTA_i * Year_i + \beta_{12} Female_i * FemaleTA_i + \beta_{13} Female_i * Year_i + \beta_{14} FemaleTA_i * Year_i + \beta_{15} URM_i * URMTA_i + \beta_{16} URM_i * Year_i + \beta_{17} URMTA_i * Year_i + \beta_{18} Female_i * FemaleTA_i * Year_i + \beta_{19} StudentURM_i * URMTA_i * Year_i \quad (3)$$

Rewritten as:

$$Grade_i = \beta_0 + \boldsymbol{\beta X} + \beta_{18} Female_i * FemaleTA_i * Year_i + \beta_{19} StudentURM_i * URMTA_i * Year_i$$

Thus, I isolate the sought-after effect with coefficients β_{18} and β_{19}.

4.2: Feasible Generalized Least Squares (FGLS) Regression with Student Grades

From visual analyses of residuals and significant Breusch-Pagan tests from regressions (1) through (3), it is clear that heteroskedasticity is a concern, and OLS is no longer BLUE. Therefore, I turn to FGLS methods laid out in Wooldridge (2013) to obtain valid standard errors and test statistics. The regression specifications remain the same; I simply add weights to each observation. The absence of significant estimates switching signs between the OLS and FGLS models provides evidence that the Gauss-Markov assumptions are satisfied (Wooldridge 2013).

4.3: Insights from TA Course Evaluations

In addition to interest in learning outcomes, I aim to find whether student biases or preferences for TAs of certain demographics differed between in-person and online environments. Because student level data are not available regarding this topic, I turn to a less precise, but nevertheless insightful method—Wilcoxon Rank-Sum tests. I group TAs together based on Year, Gender, and URM status. I run Wilcoxon Rank-Sum tests with the null hypothesis that the two samples are drawn from the same population on each of these groups to determine if there is a difference in distribution between years, genders, URM status, URM status between years, and gender between years.

Wilcoxon Rank-Sum tests are used to determine if students evaluated TAs according to different standards depending on the TA's demographics. If the distributions of course evaluation responses are significantly different for male and female TAs, this suggests that students, consciously or unconsciously, were taking the gender of their TA into consideration when evaluating them. I compare median course evaluation responses to determine the direction of these changes in the case of a significant Wilcoxon Rank-Sum test.

5. Results

5.1: How Do Student and TA Characteristics Affect Grades?

Across all specifications, GPA and Math SAT scores are significant, positive predictors of success. I first focus on Table 4. Online learning insignificantly increased students' grades. Neither students' in-state or out-of-state status nor their academic year are significant predictors of their final grade, except for third year students in the FGLS model. Graduate TAs have an insignificant, positive effect on student grades in the FGLS model. Female students tended to score around 1.2 percentage points lower than male students. Students appear to perform similarly in Econ 201 regardless of their race. In the FGLS model, the only demographic group with a significant difference in grades was American Indian or Alaskan Native, and only two individuals are in this group. Asian TAs have students with lower scores compared to white TAs by around 1.1 percentage points.

Moving to Table 5 columns (1) and (2) where students and TAs are grouped by URM status instead of the more specific ethnicities, URM TAs have a negative effect on students' grades in both the OLS and FLGS models.

5.2: How Do Student and TA Interactions Affect Grades?

The second model includes interaction terms between student and TA demographics for URM status and gender, seen in Table 5, columns (3) and (4). Online learning insignificantly decreased students' grades in the FGLS model. For those students with a male TA, females scored on average 1.5 percentage points lower than males. For males, having a female TA decreased their class grade by 1.1 percentage points compared to having a male TA, though this effect is insignificant in the FGLS model. For females, having a female TA does not significantly influence their class grade compared to having a male TA. URM TAs have an insignificant negative effect on students' grades.

5.3: How Do Student and TA Interactions Change Between In-Person and Online Learning?

Finally, I estimate equation 3 which includes year interaction terms, seen in Table 5, columns (5) and (6). Online learning had an insignificant positive effect on students' grades. Between in-person and online learning, the difference in final grades between males with a male TA and males with a female TA increased by 2.4 percentage points. For students with a female TA, the difference in the change in grades from in-person to online learning was 3.6 percentage points smaller for females than for males. In the FGLS model, all the gender interactions effects become insignificant. In the FGLS model, URM students performed better than white and Asian students in the in-person setting by 1.7 percentage points. But, they acclimated worse to online learning, performing worse than white and Asian students by 2.1 percentage points.

5.4: How Do Demographics Affect TA Course Evaluations?

Wilcoxon Rank-Sum tests have the null hypothesis H_0: $\mu_1 - \mu_2 = \alpha$, in this case $\alpha = 0$ and μ_1 and μ_2 are distribution medians. The alternative hypothesis is that one of the distributions

is shifted to the right or left of the other, therefore making the medians unequal. In the case of a significant Wilcoxon statistic, comparison of medians is used to determine which distribution is shifted to the right. Medians and Wilcoxon Rank-Sum test results are seen in Tables 6 and 7.

Table 6 shows course evaluation medians broken down by year, URM status, and gender. The transition from in-person to online learning caused students to evaluate TAs differently on every question. The median course evaluation increased for TAs' effectiveness in teaching and did not change for other questions. Combining both years together but this time focusing on TAs of URM status compared to white and Asian TAs, students evaluated white and Asian TAs higher than URM TAs on effectiveness. Now focusing on gender, students rated male TAs as more effective than female TAs.

Table 7 shows course evaluation medians broken down by year initially and then by URM status or gender. Focusing on URM status between 2019 and 2020 (columns (1) through (4)), while the distributions of evaluations were significantly different for more questions in online learning, the median score for course evaluations between URM TAs and white or Asian TAs equalized, compared to unequal ratings between these two groups on effectiveness in 2019. In contrast, focusing on gender between 2019 and 2020, the median course evaluation for amount learned, while equal in 2019, was higher for female TAs than male TAs in 2020. Students evaluated male and female TAs according to different distributions for all questions in 2020 compared to only two questions in 2019.

6. Discussion of Role Models and Biases

Both changes in grades and changes in course evaluations are used to differentiate bias from a role model effect. I define bias as a change in course evaluations without a change in grades in the same direction. I define the role model effect as an increase in grades accompanied

by either an increase in course evaluations or no change in course evaluations. Individuals may or may not be consciously aware of role models who positively influence them (Lockwood 2006; Nisbett and Wilson 1977). Therefore, even if a student experienced a boost in grades because of a demographically similar TA, they may or may not rate their TA higher on course evaluations depending on their awareness of the role model effect. If course evaluations increased following an increase in grades, then students were likely aware of the boost in grades provided by their role model TA. If course evaluations did not change following a boost in grades, then students were not aware of the role model effect despite its presence.

For example, if course evaluations for male TAs increased but male TAs did not have a positive effect on male students' grades, then male students were biased toward male TAs; male students did not benefit from having a male TA but rated them higher than female TAs. If male TAs had a positive effect on male students' grades and received higher course evaluations, then this was a role model effect; male students benefitted from having a male TA and rated them higher accordingly. If male TAs had a positive effect on male students' grades and male students did not rate them higher, this is still a role model effect. Male students experienced improved grades from having a demographically similar TA and do not necessarily need to be consciously aware of this benefit (Lockwood 2006; Nisbett and Wilson 1977).

First, I focus on URM students and TAs. URM students did not benefit from having a URM TA. Regressions in Table 5 were repeated for students of color compared to white students, and no significant effects became present. This suggests that by the time students took Econ 2010, the race-based role model effects present in pre-K and elementary school have dwindled.

When looking at course evaluations, while there was a significant difference in distributions of course evaluation responses for two out of the four questions in 2019, three of the questions had a significant difference in distribution in 2020. In other words, students evaluated URM TAs differently compared to white and Asian TAs to a greater extent in online learning compared to in-person learning. Combining this insight with the absence of significant differences in the effect of URM TAs on student grades across years is evidence of more prevalent student bias in course evaluations in online learning compared to in-person learning. There were no significant changes in the effect that URM TAs had on grades between in-person and online learning, but students evaluated them more differently compared to white and Asian TAs in the online setting.

Next, I focus on the gender of students and TAs. Results in column (3) of Table 5 indicate that, in contrast to prior research suggesting that underrepresented groups benefit from having a same-demographic instructor, it is actually males who benefitted from having a male TA in Econ 2010 in both in-person and online settings. There does not appear to be a role model effect of female TAs on female students when I do not split the observations by the semester. Again, the male role model effect becomes insignificant under the FGLS model, so there may simply be no effect of TAs on student performance regardless of gender. However, female TA role model effects emerge in OLS when I split the class by semester and compare distributions between in-person and online learning.

Results in column (5) of Table 5 indicate that male students transitioned to online learning better under male TAs and female students transitioned to online learning better under female TAs, indicative of either role model effects or of some level of meshing between a TA's method of transitioning to online teaching and a student's method of learning online. Specific

TA teaching methods are unknown, but it is possible that male TAs taught online in a manner more conducive to male styles of learning and female TAs taught online in a manner more conducive to female learning. In other words, students learning in a new setting (online) may benefit more from a same-gender role model than students who are learning in the traditional environment. Of course, because these effects become insignificant under the FGLS model, these OLS results should be interpreted cautiously.

When split by gender, course evaluations regarding amount learned increased for females in the online compared to the in-person setting, see columns (5) through (8) of Table 7. Given that the gender distribution of male TAs' discussion sections and female TAs' discussion sections did not change significantly between in-person and online learning, this change in course evaluations suggests that female students may have been aware of the role model effect which their female TAs were providing. Male students benefitted from a male TA in both the in-person and online settings, but the benefit which female students received was a new phenomenon present in the online environment only. Therefore, we would indeed expect to see higher course evaluations for female TAs than male TAs in the online compared to the in-person setting. Because course evaluations were unable to be obtained on the individual student level, this conclusion is speculative. If student-level course evaluations were ever made available, this would be a valuable datapoint.

7. Conclusion

The COVID-19 pandemic, by forcing classes at UVA to transition from in-person to online environments, created an experiment whereby discovering the effects of TA and student demographics on learning outcomes and course evaluations became feasible. Contrary to prior studies, I find that males in Econ 2010 benefitted from having a male TA in both in-person and

online environments, whereas females only benefitted from having a female TA in the online environment. URM students did not benefit from having a URM TA in either in-person or online environments.

Female TAs were rated higher on evaluations regarding amount learned in the online compared to the in-person setting. Combining this insight with the boost in grades which female students experienced from female TAs, it is possible that female students, in rating their female TAs higher in the online versus in the in-person setting, were aware of this boost.

Appendix

Table 1: Summary Statistics

Variable	2019	2020
Number of Students	909	1158
Average GPA*	3.48 [0.40]	3.60 [0.40]
Average Econ 2010 Grade	76.54 [11.64]	78.05 [11.34]
Average Math SAT	715.85 [65.12]	714.27 [66.35]
Virginia Resident	376 (41.4%)	441 (38.1%)
Non-Virginia Resident	533 (58.6%)	717 (61.9%)
Freshman	560 (61.6%)	754 (65.1%)
Sophomore	295 (32.5%)	342 (29.5%)
Junior	41 (4.5%)	34 (2.9%)
Senior	13 (1.4%)	28 (2.4%)
American Indian or Alaskan Native Students	2 (0.2%)	0
Asian Students	152 (16.7%)	220 (19.0%)
Black or African American Students	45 (5.0%)	52 (4.5%)
Hispanic Students	56 (6.2%)	73 (6.3%)
Multi-Race Students	50 (5.5%)	67 (5.8%)
White Students	604 (66.4%)	746 (64.4%)
Female Students	488 (53.7%)	606 (52.3%)
Male Students	421 (46.3%)	552 (47.7%)
Number of TAs	24	29
Asian TAs	5 (20.8%)	8 (27.6%)
Black or African American TAs	0	1 (3.5%)
Hispanic TAs	0	2 (6.9%)
Multi-Race TAs	3 (12.5%)	1 (3.5%)
White TAs	16 (66.%)	17 (58.6%)
Female TAs	8 (33.3%)	11 (37.9%)
Male TAs	16 (66.7%)	18 (62.1%)
Undergraduate TAs	17 (70.8%)	20 (69.0%)
Graduate TAs	7 (29.2%)	9 (31.0%)

Note:
* indicates significant difference (p = 0.05).
Average GPA, Average Econ 2010 Grade, and Average SAT display mean [SD].
All other variables display mean (%).

Table 2: Comparable Course Evaluation Questions in 2019 and 2020

Question Theme	Question in 2019	Question in 2020
Availability	The instructor's availability outside of class through office hours, appointments, or informal communication was:	The instructor was available to help support my learning.
Effectiveness	Overall, the instructor was an effective teacher.	Overall, the instructor was an effective teacher.
Amount Learned	I learned a great deal in this course.	Through this course I gained a deeper understanding of the subject matter.
Time Spent	The average number of hours per week I spent outside of class preparing for this course was:	The average number of hours (per week) that I spent outside of class preparing for ECON 20100 Principles of Econ: Microeconomics:

Note:
There were four comparable questions between the Fall 2019 and Fall 2020 course evaluations for TAs. Their themes were a TA's availability, effectiveness, the amount learned from the TA, and the time spent studying for the discussion section outside of class time.
The question from the course evaluation during in-person learning (Fall 2019) is on the left while the comparable question during online learning (Fall 2020) is on the right.

Table 3: Possible Responses to Course Evaluation Questions

Numerical Value	5	4	3	2	1
Availability (response on left is from 2019; response on right is from 2020)	Excellent/Strongly Agree	Good/Agree	Okay/Neutral	Poor/Disagree	Very Poor/Strongly Disagree
Effectiveness	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Amount Learned	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Time Spent	10 or more hours	7-9 hours	4-6 hours	1-3 hours	Less than 1 hour

Note:
There were four comparable questions between the Fall 2019 and Fall 2020 course evaluations for TAs. Their themes were a TA's availability, effectiveness, the amount learned from the TA, and the time spent studying for the discussion section outside of class time.
A numerical value was assigned to each possible response in preparation for analysis through Wilcoxon Rank-Sum tests.
The response options were the same across years for the questions regarding effectiveness, amount learned, and time spent. The response options for the question regarding availability changed between in-person learning (Fall 2019) and online learning (Fall 2020).

Table 4: OLS and FGLS Regression Results with Race Broken Down

	Dependent variable:	
	Weighted Average Grade	
	OLS	FGLS
	(1)	(2)
GPA	13.706***	17.828***
	(0.522)	(0.597)
SAT	0.052***	0.040***
	(0.003)	(0.004)
Online Semester	0.300	0.059
	(0.407)	(0.353)
Virginia Resident	0.641	-0.096
	(0.402)	(0.343)
Sophomore	-0.571	-0.458
	(0.424)	(0.399)
Junior	-1.209	-2.323**
	(1.041)	(1.163)
Senior	-1.523	-1.599
	(1.389)	(1.430)
Graduate TA	-0.102	0.056
	(0.426)	(0.365)
Female	-1.109***	-1.237***
	(0.396)	(0.341)
American Indian or Alaska Native	11.451*	10.803***
	(6.158)	(1.336)
Asian	-0.894*	-0.693
	(0.529)	(0.456)
Black or African American	-1.626*	-1.229
	(0.957)	(1.083)
Hispanic	0.158	1.014
	(0.809)	(0.800)
Multi-Race	0.826	0.625
	(0.836)	(0.628)
Female TA	-0.096	0.105
	(0.428)	(0.372)
Asian TA	-1.529***	-1.101***
	(0.473)	(0.418)
Black or African American TA	-1.849	-1.509
	(1.272)	(1.513)
Hispanic TA	-1.870*	-1.408
	(0.963)	(0.940)
Multi-Race TA	0.187	0.429
	(0.810)	(0.674)
Constant	-7.568***	-13.545***
	(2.721)	(2.925)
Observations	2,067	2,067
R^2	0.441	0.435
Note:	*p<0.1; **p<0.05; ***p<0.01	

'Online Semester' is equal to 0 for the Fall 2019 semester and 1 for the Fall 2020 semester. 'Virginia Resident' is equal to 0 if a student is an out-of-state student and 1 if the student is an in-state student. The omitted category for the student's academic year is 'Freshman'. 'Graduate TA' is equal to 0 if the TA is an undergraduate student and 1 if the TA is a graduate student. The omitted category for student and TA race is 'White'. 'Female TA' is equal to 0 if the TA is male and 1 if the TA is female. Standard errors are in parentheses.

Table 5: OLS and FGLS Regression Results with URM Indicator

	Dependent variable:					
	Weighted Average Grade					
	OLS	FGLS	OLS	FGLS	OLS	FGLS
	(1)	(2)	(3)	(4)	(5)	(6)
GPA	13.890***	18.274***	13.904***	18.053***	13.957***	18.118***
	(0.519)	(0.611)	(0.519)	(0.608)	(0.520)	(0.611)
SAT	0.051***	0.039***	0.051***	0.040***	0.051***	0.039***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Online Semester	0.050	-0.118	0.040	-0.040	1.214	1.010
	(0.396)	(0.345)	(0.396)	(0.347)	(0.788)	(0.688)
Virginia Resident	0.538	-0.137	0.534	-0.101	0.549	-0.161
	(0.396)	(0.339)	(0.396)	(0.342)	(0.397)	(0.343)
Sophomore	-0.573	-0.474	-0.593	-0.463	-0.555	-0.376
	(0.425)	(0.405)	(0.425)	(0.405)	(0.426)	(0.405)
Junior	-1.020	-1.383	-1.055	-1.418	-0.987	-1.391
	(1.038)	(1.495)	(1.038)	(1.462)	(1.042)	(1.485)
Senior	-1.540	-1.029	-1.636	-1.139	-1.550	-1.401
	(1.384)	(1.260)	(1.386)	(1.198)	(1.386)	(1.220)
Graduate TA	-0.103	0.081	-0.094	0.087	0.233	0.285
	(0.422)	(0.366)	(0.422)	(0.370)	(0.642)	(0.523)
Female	-1.162***	-1.241***	-1.614***	-1.508***	-1.003	-1.111*
	(0.395)	(0.343)	(0.495)	(0.431)	(0.726)	(0.609)
Female TA	-0.444	-0.083	-1.085*	-0.514	0.249	0.317
	(0.400)	(0.345)	(0.581)	(0.535)	(0.900)	(0.744)
URM	0.179	0.727	0.138	0.812	0.974	1.731**
	(0.537)	(0.500)	(0.627)	(0.613)	(0.875)	(0.821)
URM TA	-1.052**	-0.724*	-1.096**	-0.732	-1.096	-0.266
	(0.455)	(0.409)	(0.496)	(0.450)	(0.838)	(0.768)
Online Semester * Graduate TA					-0.705	-0.278
					(0.859)	(0.743)
Female * Female TA			1.196	0.721	-0.834	-0.159
			(0.790)	(0.701)	(1.204)	(0.992)
Online Semester * Female					-1.041	-0.746
					(0.973)	(0.856)
Online Semester * Female TA					-2.384**	-1.445
					(1.184)	(1.063)
URM * URM TA			0.234	-0.280	0.136	-0.841
			(1.143)	(1.026)	(1.974)	(1.932)
Online Semester * URM					-1.720	-2.054*
					(1.207)	(1.187)
Online Semester * URM TA					0.140	-0.687
					(1.053)	(0.967)
Online Semester * Female * Female TA					3.545**	1.609
					(1.595)	(1.393)
Online Semester * URM * URM TA					0.777	1.645
					(2.442)	(2.365)
Constant	-7.572***	-14.782***	-7.275***	-14.088***	-8.152***	-14.072***
	(2.628)	(2.891)	(2.637)	(2.878)	(2.663)	(2.902)
Observations	2,067	2,067	2,067	2,067	2,067	2,067
R^2	0.436	0.424	0.437	0.424	0.439	0.423
Note:					*p<0.1; **p<0.05; ***p<0.01	

'Online Semester' is equal to 0 for the Fall 2019 semester and 1 for the Fall 2020 semester. 'Virginia Resident' is equal to 0 if a student is an out-of-state student and 1 if the student is an in-state student. The omitted category for the student's academic year is 'Freshman'. 'Graduate TA' is equal to 0 if the TA is an undergraduate student and 1 if the TA is a graduate student. The omitted category for student and TA race is 'White'. 'Female TA' is equal to 0 if the TA is male and 1 if the TA is female. Standard errors are in parentheses.

Table 6: Wilcoxon Rank-Sum Test Results and Median Course Evaluations

	Year		URM Status		Gender	
Question	In-Person	Online	URM TA	White/Asian TA	Male TA	Female TA
Amount Learned	4 *	4 *	4	4	4 *	4 *
Effectiveness	4 *	5 *	4 *	5 *	5 *	4 *
Availability	5 *	5 *	5 *	5 *	5 *	5 *
Time Spent	5 *	5 *	4	4	5	5

Note:
* indicates significant Wilcoxon Rank-Sum test (p = 0.05).
The distributions being tested against each other are grouped under the highest heading. For example, under the heading 'Year', the distribution of course evaluations for TAs who taught in-person (Fall 2019) is compared to that of TAs who taught online (Fall 2020).
The questions listed correspond to those in Tables 2, 3, and 7.

Table 7: Wilcoxon Rank-Sum Test Results and Median Course Evaluations Categorized by Year

	URM Status in 2019		URM Status in 2020		Gender in 2019		Gender in 2020	
Question	URM TA in 2019	White/Asian TA in 2019	URM TA in 2020	White/Asian TA in 2020	Male TA in 2019	Female TA in 2019	Male TA in 2020	Female TA in 2020
Amount Learned	4	4	4	4	4 *	4 *	4 *	5 *
Effectiveness	4	5	5 *	5 *	4 *	5 *	4 *	5 *
Availability	5 *	5 *	5 *	5 *	5	5	5 *	5 *
Time Spent	5 *	5 *	5 *	5 *	5	5	5 *	5 *

Note:

* indicates significant Wilcoxon Rank-Sum test ($p = 0.05$).

The distributions being tested against each other are grouped under the highest heading. For example, under the heading ‘URM Status in 2019’, the distribution of course evaluations for URM TAs in 2019 is compared to that of white and Asian TAs in 2019.

The questions listed correspond to those in Tables 2, 3, and 6.

References

Andersen, Kristi, and Elizabeth D. Miller. 1997. "Gender and Student Evaluations of Teaching." *PS: Political Science & Politics* 30 (02): 216–19. https://doi.org/10.1017/S1049096500043407.

Anderson, Kristin J., and Gabriel Smith. 2005. "Students' Preconceptions of Professors: Benefits and Barriers According to Ethnicity and Gender." *Hispanic Journal of Behavioral Sciences* 27 (2): 184–201. https://doi.org/10.1177/0739986304273707.

Ary, Eddie J, and Christopher W Brune. 2011. "A Comparison of Student Learning Outcomes in Traditional and Online Personal Finance Courses." *MERLOT Journal of Online Learning and Teaching* 7 (4): 465–74.

Atchley, Wayne, Gary Wingenbach, and Cindy Akers. 2013. "Comparison of Course Completion and Student Performance through Online and Traditional Courses." *International Review of Research in Open and Distributed Learning* 14 (4): 104–16.

Bachen, Christine M., Moira M. McLoughlin, and Sara S. Garcia. 1999. "Assessing the Role of Gender in College Students' Evaluations of Faculty." *Communication Education* 48 (3): 193–210. https://doi.org/10.1080/03634529909379169.

Ballard, Charles L., and Marianne F. Johnson. 2004. "Basic Math Skills and Performance in an Introductory Economics Class." *The Journal of Economic Education* 35 (1): 3–23. https://doi.org/10.3200/JECE.35.1.3-23.

Basow, Susan, Stephanie Codos, and Julie Martin. 2013. "The Effects of Professors' Race and Gender on Student Evaluations And Performance." *College Student Journal* 47 (2): 352–63. https://www.ingentaconnect.com/content/prin/csj/2013/00000047/00000002/art00011.

Boring, Anne. 2017. "Gender Biases in Student Evaluations of Teaching." *Journal of Public Economics* 145 (January): 27–41. https://doi.org/10.1016/j.jpubeco.2016.11.006.

Carr, Sarah. 2000. "As Distance Education Comes of Age, the Challenge Is Keeping the Students." *The Chronicle of Higher Education*, February. https://www.chronicle.com/article/as-distance-education-comes-of-age-the-challenge-is-keeping-the-students/?bc_nonce=54matuynypd6rwpqsz9itb&cid=reg_wall_signup.

Chisadza, Carolyn, Nicky Nicholls, and Eleni Yitbarek. 2019. "Race and Gender Biases in Student Evaluations of Teachers." *Economics Letters* 179 (June): 66–71. https://doi.org/10.1016/j.econlet.2019.03.022.

Coates, Dennis, Brad R Humphreys, John Kane, and Michelle A Vachris. 2004. "'No Significant Distance' between Face-to-Face and Online Instruction: Evidence from Principles of Economics." *Economics of Education Review* 23 (5): 533–46.

Daymont, Thomas, and Gary Blau. 2008. "Student Performance in Online and Traditional Sections of an Undergraduate Management Course." *Journal of Behavioral and Applied Management* 9 (3): 275.

Dee, Thomas D. 2007. "Teachers and the Gender Gaps in Student Achievement." *Journal of Human Resources* 42 (3): 528–54.

Dunwoody, Philip T., and Michael L. Frank. 1995. "Why Students Withdraw from Classes." *The Journal of Psychology* 129 (5): 553–58. https://doi.org/10.1080/00223980.1995.9914927.

Gershenson, Seth, Cassandra M. Hart, Joshua Hyman, Constance Lindsay, and Nicholas Papageorge. 2018. "The Long-Run Impacts of Same-Race Teachers." Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w25254.

Gratton-Lavoie, Chiara, and Denise Stanley. 2009. "Teaching and Learning Principles of

Microeconomics Online: An Empirical Assessment." *The Journal of Economic Education* 40 (1): 3–25.

Grove, Wayne A., and Tim Wasserman. 2004. "The Life-Cycle Pattern of Collegiate GPA: Longitudinal Cohort Analysis and Grade Inflation." *The Journal of Economic Education* 35 (2): 162–74. https://doi.org/10.3200/JECE.35.2.162-174.

Hastedt, Sarah. 2008. "Group Differences in Performance: The Effects of Teaching Assistants on Collegiate Grades." http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.214.472&rep=rep1&type=pdf.

Hiltz, Starr Roxanne. 1997. "Impacts of College-Level Courses via Asynchronous Learning Networks: Some Preliminary Results." *Journal of Asynchronous Learning Networks* 1 (2): 1–19.

Hoffman, Heather J, and Angelo F Elmi. 2020. "Comparing Student Performance in a Graduate-Level Introductory Biostatistics Course Using an Online Versus a Traditional In-Person Learning Environment." *Journal of Statistics Education*, 1–10.

Liu, Yuliang. 2005. "Effects of Online Instruction vs. Traditional Instruction on Students' Learning." *International Journal of Instructional Technology and Distance Learning* 2 (3): 57–64.

Lockwood, Penelope. 2006. "'Someone Like Me Can Be Successful': Do College Students Need Same-Gender Role Models?" *Psychology of Women Quarterly* 30 (1): 36–46. https://doi.org/10.1111/j.1471-6402.2006.00260.x.

Lusher, Lester, Doug Campbell, and Scott Carrell. 2018. "TAs like Me: Racial Interactions between Graduate Teaching Assistants and Undergraduates." *Journal of Public Economics* 159 (March): 203–24. https://doi.org/10.1016/j.jpubeco.2018.02.005.

Marcham, Cheryl Ann, Ann Marie Ade, Patti Clark, and James Marion. 2020. "Bias and Trends in Student Evaluations in Online Higher Education Settings." *Collegiate Aviation Review* 38 (2). https://ojs.library.okstate.edu/osu/index.php/CARI/article/view/8036.

Mengel, Friederike, Jan Sauermann, and Ulf Zölitz. 2019. "Gender Bias in Teaching Evaluations." *Journal of the European Economic Association* 17 (2): 535–66. https://doi.org/10.1093/jeea/jvx057.

Morgenroth, Thekla, Michelle K. Ryan, and Kim Peters. 2015. "The Motivational Theory of Role Modeling: How Role Models Influence Role Aspirants' Goals." *Review of General Psychology* 19 (4): 465–83. https://doi.org/10.1037/gpr0000059.

Neumark, David, and Rosella Gardecki. 1998. "Women Helping Women? Role Model and Mentoring Effects on Female Ph.d. Students in Economics." *The Journal of Human Resources* 33 (1): 220. https://doi.org/10.2307/146320.

Nisbett, Richard E., and Timothy D. Wilson. 1977. "Telling More than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84 (3): 231–59. https://doi.org/10.1037/0033-295X.84.3.231.

Olitsky, Neal H., and Sarah B. Cosgrove. 2014. "The Effect of Blended Courses on Student Learning: Evidence from Introductory Economics Courses." *International Review of Economics Education* 15 (January): 17–31. https://doi.org/10.1016/j.iree.2013.10.009.

Orlov, George, Douglas McKee, James Berry, Austin Boyle, Thomas DiCiccio, Tyler Ransom, Alex Rees-Jones, and Jörg Stoye. 2020. "Learning during the COVID-19 Pandemic: It Is Not Who You Teach, but How You Teach." National Bureau of Economic Research.

Patterson, Belinda, and Cheryl McFadden. 2009. "Attrition in Online and Campus Degree Programs." *Online Journal of Distance Learning Administration* 12 (2).

Paul, Jasmine, and Felicia Jefferson. 2019. "A Comparative Analysis of Student Performance in an Online vs. Face-to-Face Environmental Science Course From 2009 to 2016." *Frontiers of Computer Science* 1 (November). https://doi.org/10.3389/fcomp.2019.00007.

Porter, Catherine, and Danila Serra. 2020. "Gender Differences in the Choice of Major: The Importance of Female Role Models." *American Economic Journal. Applied Economics* 12 (3): 226–54.

Rask, Kevin N., and Elizabeth M. Bailey. 2002. "Are Faculty Role Models? Evidence from Major Choice in an Undergraduate Institution." *The Journal of Economic Education* 33 (2): 99–124. https://doi.org/10.1080/00220480209596461.

Reid, Landon D. 2010. "The Role of Perceived Race and Gender in the Evaluation of College Teaching on RateMyProfessors.Com." *Journal of Diversity in Higher Education* 3 (3): 137–52. https://doi.org/10.1037/a0019865.

Robb, Roberta Edgecombe, and A. Leslie Robb. 1999. "Gender and the Study of Economics: The Role of Gender of the Instructor." *The Journal of Economic Education* 30 (1): 3. https://doi.org/10.2307/1183028.

Sidanius, Jim, and Marie Crane. 1989. "Job Evaluation and Gender: The Case of University Faculty." *Journal of Applied Social Psychology* 19 (2): 174–97. https://doi.org/10.1111/j.1559-1816.1989.tb00051.x.

Summers, Jessica J, Alexander Waigandt, and Tiffany A Whittaker. 2005. "A Comparison of Student Achievement and Satisfaction in an Online versus a Traditional Face-to-Face Statistics Class." *Innovative Higher Education* 29 (3): 233–50.

Waschull, Stefanie B. 2001. "The Online Delivery of Psychology Courses: Attrition, Performance, and Evaluation." *Teaching of Psychology* 28 (2): 143–47.

Werhner, Matthew J. 2010. "A Comparison of the Performance of Online versus Traditional On-Campus Earth Science Students on Identical Exams." *Journal of Geoscience Education* 58 (5): 310–12.

Wooldridge, Jeffrey M. 2013. "Heteroskedasticity." In *Introductory Econometrics: A Modern Approach*, 5th ed. Mason, OH: South-Western.

Xu, Di, and Shanna Jaggars. 2013. "Adaptability to Online Learning: Differences across Types of Students and Academic Subject Areas."

Zirkel, Sabrina. 2002. "Is There A Place for Me? Role Models and Academic Identity among White Students and Students of Color." *Teachers College Record* 104 (2): 357–76. https://doi.org/10.1111/1467-9620.00166.