Mixing and Matching: The Effect on Student Performance of Teaching Assistants of the Same Gender

Daniel M. Butler, Stanford University
Ray Christensen, Brigham Young University

Introduction

In Failing at Fairness, Sadker and Sadker (1994) provide extensive evidence that America’s school system has systematically discriminated against girls and women. In areas that boys are encouraged to excel, girls are denied opportunities. They are taught to speak quietly, to defer to boys, to avoid math and science, and to value neatness over innovation and appearance over intelligence (Sadker and Sadker 1994, 13). Girls consistently outperform boys in elementary school but lag behind them by the time they graduate. Sadker and Sadker show that discrimination does not end with high school graduation, but follows female students through their college careers.

How are political science departments doing? Though roughly half of political science majors are women, are they systematically discriminated against? Two comprehensive reports on women in political science have documented gender discrimination in graduate schools and in the hiring and promotion of women faculty (Committee on the Status of Women in the Profession 2001; Committee on the Status of Women 1992). However, these reports only briefly address issues of gender discrimination in the undergraduate classroom.

We are surprised by the relative lack of attention paid to issues of gender discrimination in undergraduate teaching because our jobs, our departments, and our universities exist to teach undergraduates. Potential gender discrimination in undergraduate teaching is an especially important issue at our university where our political science major attracts more men than women, raising the question of whether gender discrimination contributes to this pattern. Thus, we hope to contribute to a discussion of gender discrimination in undergraduate teaching by identifying obstacles women face in political science today and testing for the presence of those obstacles in an undergraduate classroom setting.

We find evidence to support claims that the content of political science courses, specifically statistics, may create barriers for women. We also find support for the claim that women teaching assistants effectively motivate women students. Surprisingly, however, we find that the gender match between a teaching assistant and a student appears largely irrelevant for student performance. Women seem to be more motivated to complete a course when their teaching assistant is a woman, but we find no consistent evidence that women perform better on course assignments when they are taught by a woman.

Hypotheses

It would be difficult to make a comprehensive list of all the potential obstacles female political science students face. We have condensed the extensive literature on gender discrimination to three distinct hypotheses that were testable and seemed appropriate for a study of gender discrimination in undergraduate political science courses.

Our first hypothesis concerns role model effects, the primary concern of the report on gender and the profession. Female instructors may inspire female students to perform better. Ehrenberg (1997) and his associates carried out the most prominent research project in this area. Using data from the National Educational Longitudinal Study of 1988, they concluded that student performance had little association with the match between the students’ race, gender, and ethnicity, and that of the teacher's. Although the Ehrenberg study concluded that a gender match of the instructor and student did not affect student performance, we have decided to test this hypothesis for several reasons. First, there has not been any follow-up research on their findings. Second, the Ehrenberg study looked at students in high school, not college. Third, we can obtain information about students taking courses that was not available to Ehrenberg and his colleagues, allowing us to control for different factors. Finally, we have the advantage of looking at students’ performance on a variety of course assignments. We expect that the role model effect would be greatest on assignments that require a close interaction between a student and a teaching assistant and weakest on assignments in which the interaction with the teaching assistant was less important. Thus, a multiple choice exam in a large lecture class should be less influenced by the role model effect of the gender of the teaching assistant while an intense writing and research assignment that required multiple interactions with a teaching assistant should exhibit a greater role model effect.

A second hypothesis comes out of studies about the relationship between gender and course content. Recent studies suggest that in math young girls perform as well as boys do; however, over time fewer girls continue in advanced math courses because of a lack of encouragement and tracking practices (Leahy and Guo 2001; Weber 2001, 147–48). By the time students enter college, women have developed adverse attitudes toward math, affecting their performance in math-related areas (Nosék et al. 2002). In contrast, Fox and Ronkowski (1997) show a gender advantage for women in tasks that are spelled out with clarity. We expect that political science classes that use extensive math or statistics might exhibit a gender disadvantage for women students.

A final hypothesis is derived from descriptions of a “chilly learning environment.” Men volunteer comments more than women; men are called on more than women; instructors learn the names of men faster than they learn the names of women (Banks 1988). Similarly,

Daniel M. Butler graduated from Brigham Young University in 2003 with an undergraduate degree in political science. He will begin Stanford University’s political science Ph.D. program in the fall.

Ray Christensen is associate professor of political science at Brigham Young University.
topics typically chosen by men and styles of argumentation used by men are favored in political science and other disciplines (Andersen and Miller 1997; Taber et al. 1988; Banks 1988, 138). Though one female teaching assistant could not correct for all of these biases, we expect that a female teaching assistant is less likely to contribute to this type of a biased learning environment, somewhat counteracting the effects of the discipline, the content of the course, and the male gender of the professors teaching the course.

Data

We saw an excellent opportunity to examine these hypotheses by collecting data on the performance of more than 600 students who took the course “Political Inquiry” over a three-year period. This course is a required, introductory course that emphasizes writing, library research, research designs, and statistics. The course meets three times a week; two of those meetings are lectures by the professor and the third is an hour-long lab taught by an undergraduate teaching assistant. The lecture class has from 120 to 160 students; the labs have from 10 to 20 students with each teaching assistant teaching two labs during a semester. Students write weekly papers including an essay on a theory and a review of an article; several research designs; and exercises on grammar, style, library research, statistics, and citation style. The teaching assistants’ main responsibility during labs is to explain the writing assignments. They also grade these assignments, administer quizzes, review the week’s lectures, answer questions, and handle all complaints about grading.

The unusually prominent role of teaching assistants in this course provided an excellent test case to assess gender discrimination. The best course in which to test the effect of the gender of an instructor would be a nearly identical course taught by a variety of instructors. For example, different sections of a calculus course that used the same textbook and same exam would create an excellent test of the effect of the gender of the instructor because differences in the course structure and evaluations are minimized. Unfortunately, there are very few, if any, political science courses that are taught in such a nearly identical manner. Typically, political science courses with the same course title will vary in the amount and type of reading, assignments, and exam formats.

Thus, our study of teaching assistants seems appropriate for an analysis of the gender effects of an instructor. All students take the same identical course with identical readings, exams, and writing assignments. Only the gender of the teaching assistant varies from one lab to the next. Furthermore, the role of teaching assistants in our course is more like that of an instructor than a typical teaching assistant. The teaching assistants explain the writing assignments, work closely with students on those assignments, grade the assignments, and handle student complaints about grading, making the teaching assistants nearly equivalent to instructors in the course.

A wide range of factors have been linked to discrimination, oppression, and barriers in society and the classroom. Weber (2001) examines factors thought to contribute to discrimination based on race, class, gender, and sexuality. In our course, societal norms, university practices, assigned texts, the discipline of political science, and the white, male, upper middle class characteristics of the two professors of the course could all potentially contribute to the presence of discriminatory practices and attitudes. Nevertheless, in our study we attempt to hold these factors constant by selecting a course in which these factors do not vary. In our results we recognize the constant effect of these other factors while simultaneously drawing conclusions about the impact of gender matching between students and teaching assistants.

In addition, the nature of the class controls for self-selection. “Political Inquiry” is a required course for the political science major; very few students choose to take the course as an elective. More importantly, students can not choose their teaching assistant. Because they register for a lab section without knowing who their teaching assistant will be, student lab assignments are made regardless of any preference the students might have concerning the gender of their instructors. We checked for the random nature of the gender of the teaching assistant by running a cross-tabulation on the gender of the student against that of the teaching assistant. The chi-square value was 0.3889 with a probability significance level of .53. This is strong evidence that gender did not play a role in deciding which lab students registered for.

There are three other advantages to using this course for our analysis. First, in most of the semesters studied in the course teaching assistants used a blind-grading system. The anonymity of students was protected through the use of code names, preventing overt bias in grading for or against individual students or groups of students. Second, the course has a reputation among students as being the most rigorous of the courses in the major. If women are discouraged from being political science majors, a likely place for that discouragement to occur is in this course. Finally, our course data has both objective and subjective measures of student performance. The course grade is based primarily on the writing assignments (subjective) and a multiple-choice final exam (objective). In our analysis we regress both of these measures on the
same set of independent variables. The different nature of these two measures allows us to test between several different hypotheses of gender discrimination.

The role model effect could be manifest as either a general or a specific influence. Generally, women students with female teaching assistants would be more enthusiastic about the overall course. They would be more likely to complete the course and more likely to do better on all course assignments. Specifically, the role model effect might appear more strongly on writing assignments for which students work closely with their teaching assistants. In contrast, having a female teaching assistant should be less relevant for performance on the final exam which covers material from class lectures and is only tangentially related to the teaching assistant’s lab sections.

The course content suggests that women will do worse in the class, regardless of the gender of the professor or teaching assistant, but the exam is more heavily weighted toward statistics and the assignments are more heavily weighted toward writing and specific task-oriented activities, leading to the conclusion that women should do better (regardless of the gender of their TA) on the assignments and worse on the final.

Unfortunately, hypotheses derived from observations about a “chilly” learning environment mirror those of the role model hypothesis. Women students should do better on all course work if their teaching assistant is female because there will be less of a “chilly” learning environment for those students. This advantage should be greatest for the writing assignments where the interaction with the teaching assistants is extensive, and least for the exam where the interaction with the teaching assistants is much less important.

Model

Our first analysis is of students who fail to complete the course. We define this term broadly as any student who failed to take the final exam or received less than a total of 40% on their combined writing assignments. We analyze in a logistic regression the dependent variable “complete the course” which is coded as 1 for all students who met the minimal requirements for course completion and 0 for those who did not.

Our most important independent variable was the gender match between a student and his or her teaching assistant. We divided students into four categories: male students with a male teaching assistant, male students with a female teaching assistant, female students with a male teaching assistant, and female students with a female teaching assistant. To avoid problems of multicolinearity, we excluded the category of male students with male teaching assistants from this and subsequent regressions. We also included several other control variables, including the student’s prior GPA, year in school, age, credit hours currently enrolled in, marital status, and U.S. citizenship.

We included GPA since performance in past classes is a measure of both innate ability and work ethic. Students who have done well in previous classes should perform well in their current classes. We included the number of credit hours a student took because the course requires many time consuming writing assignments. Students with fewer credits have more time to study, extremely important for rewriting and revising their writing assignments. Similarly, we included a variable for marital status because we believed that married students, on average, would be more serious about their course work than unmarried students. We also expected year in school to be correlated with class performance; more advanced students should do better in the class, with the possible exception of seniors who may have delayed taking this introductory class until their final year of school for reasons that might be correlated with a poorer performance in the class. Finally, we included a dummy variable to capture whether or not the student was a citizen of a country where English is the native language. Citizenship is a surrogate measure for being a native speaker of English. We expect native speakers to do better in the course, especially on writing assignments.

During the five semesters that the data set covers, two different professors took turns teaching the course and 15 teaching assistants taught various lab sections. With both of the professors and each of the teaching assistants there are certain unobserved characteristics associated with him or her that affect the performance of his or her students. To control for this unobserved heterogeneity, we included dummy variables for one of the professors and for 14 of the 15 teaching assistants. The dummy variable for the professors proved to be significant, as were several of the dummy variables for the teaching assistants. To minimize visual confusion and to keep the tables from becoming too cluttered we do not include the dummy variables for the professor and teaching assistants in our regression results.

Initially there were 669 observations; however, two students never signed up for a lab section, so the working data set has 667 observations. Of these 667 students, 82 or 12.3% failed to complete the course.

### Descriptive Statistics

Of the 669 students, 64.9% were male, and 73% took the class from a male teaching assistant. Most of the students were single (81.1%) and were U.S. citizens (96.7%). While students from each stage of their college careers were represented in the sample, juniors were the largest group (38.7%). Prior to taking the course, the average GPA for the students was 3.16. On average, a student in the course was enrolled in 14 credits during the semester. Finally, the average score on the final exam was 60.9%, and the average score on their combined written assignments was 67.7%.

### Regression Results

In a logistic regression of 30 independent variables on the dependent variable of whether or not a student completed the course, only two variables
more motivated to study.

We continued our analysis of the effect of instructor gender by assessing student performance for the 585 students who completed the course. We analyzed separately two dependent variables: the students’ final exam scores and their combined scores on all of their written assignments. The final exam score is a percent correct on a 100-question, multiple choice exam. The maximum possible score on all of the writing assignments combined was 650 points. To facilitate interpretation of the regression results and comparison between them, we divided the point total on the assignments for each student by 6.5, measuring both variables as a score out of 100 points possible. We included the same battery of independent variables.

After controlling for fixed effects by including dummy variables for the teaching assistants and the professor, most of the results came out as expected. The number of credit hours was the only variable with a different coefficient value than we had initially expected. Students taking more credits actually did better on course assignments. Perhaps credit hours operate as a surrogate measure for the number of hours employed. Students who take fewer credits may work more hours, and those taking more credits may work fewer hours or not at all. In addition, students with higher GPAs do better, as do married students.

U.S. citizens also have the strongest advantage on writing assignments as we predicted. First year students consistently earn the lowest grades, and sophomores and juniors outperform seniors on written assignments with all three grades earning similar exam scores. Students age 20–21 generally outperform students in other age cohorts.

The values of the coefficients regarding the gender match are interesting. On both assignments and finals men do better than women and the gap between men and women does not narrow for writing assignments, contrary to the course content hypothesis. In addition, the gender of the teaching assistant does not produce a clear effect on student performance. For example, men do 2.3% better than women on the final exam for students that have male teaching assistants, but for students who have female teaching assistants, the male advantage increases to 3.7%. We found similar results when we analyzed assignments. In labs with male teaching assistants, men on average earn grades that are 2.5% better than women in those same labs, but in labs with female teaching assistants men on average earn scores 3.7% better than women. These findings do not show an advantage for women with female teaching assistants, and this lack of advantage is nearly identical for both exams and writing assignments.

In both regressions, men with male teaching assistants served as the baseline. On assignments, it was true that women with female teaching assistants earned higher grades (1.4%) than women with male teaching assistants. However, men with female teaching as-

---

Table 2
Regressing Final Exam Scores on Demographic Variables

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>coefficients</th>
<th>standard errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>.3</td>
<td>4.8</td>
</tr>
<tr>
<td>Women in male labs</td>
<td>−2.3</td>
<td>1.2</td>
</tr>
<tr>
<td>Women in female labs</td>
<td>−5.3</td>
<td>3.5</td>
</tr>
<tr>
<td>Men in female labs</td>
<td>−1.6</td>
<td>3.3</td>
</tr>
<tr>
<td>U.S. citizen</td>
<td>2.8</td>
<td>2.3</td>
</tr>
<tr>
<td>Married</td>
<td>2.1*</td>
<td>1.0</td>
</tr>
<tr>
<td>GPA</td>
<td>14.8*</td>
<td>.8</td>
</tr>
<tr>
<td>Credits</td>
<td>−.6*</td>
<td>.2</td>
</tr>
<tr>
<td>Sophomores</td>
<td>−.04</td>
<td>1.2</td>
</tr>
<tr>
<td>Juniors</td>
<td>−0.2</td>
<td>1.0</td>
</tr>
<tr>
<td>First year students</td>
<td>−0.8</td>
<td>2.3</td>
</tr>
<tr>
<td>Age 17–18</td>
<td>−1.5</td>
<td>2.3</td>
</tr>
<tr>
<td>Age 19</td>
<td>−3.0*</td>
<td>1.5</td>
</tr>
<tr>
<td>Age 22–25</td>
<td>−3.7*</td>
<td>1.1</td>
</tr>
<tr>
<td>Age 26 and older</td>
<td>−8.2*</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Adjusted R Square .43
*statistically significant at the .05 level
sists also earned higher grades (2.6%) than men in labs taught by male teaching assistants. On final exams women on average did 3% better if they had a male teaching assistant, but men also did 1.6% better if they had a male teaching assistant. Students with female teaching assistants did better on assignments, regardless of their gender, and students with male teaching assistants did better on the final, regardless of their gender. In all cases, however, average scores for men were consistently higher than the scores for women in the same categories of labs.

These results, however, rarely reached levels of statistical significance for matching the gender of the student with the teaching assistant, raising questions about the robustness of our findings.9 Further analysis of separate regressions run for each individual teaching assistant underscores this concern. In these 15 separate regressions there are no consistent differences between the gender of the teaching assistant and the performance of students of the same gender. For example, six of the nine male teaching assistants had men outperforming women on writing assignments, and three of six women teaching assistants showed a similar male advantage. On the final, men did better for five of the nine male teaching assistants and for four of the six female teaching assistants. Similarly, women did relatively better on assignments than on the final (in comparison to the average male performance) for two of the six female teaching assistants but also for four of the nine male teaching assistants.

Finally, for most of the teaching assistants there was not a large differential between the average performance of women relative to men on the objectively graded final compared with the subjectively graded writing assignments. The four largest differentials occurred in two male and two female teaching assistants’ results. However, one male and one female both had women doing better than men on assignments but men doing better than women on the exam. The other male teaching assistant had men doing better on assignments with women doing better on exams. The remaining female teaching assistant had men doing better on both assignments and exams with a much larger male advantage on assignments. These inconsistent results suggest that the gender of the instructor may not be a significant factor in student performance.

Though our results support the more general claims of the role model hypothesis (women with female teaching assistants are more likely to finish the course), the specific claims of the role model hypothesis and the “chilly learning environ-

### Conclusion

We have articulated some of the obstacles that women political science students still face. Barriers to women in courses that rely heavily on math or statistics can not easily be eliminated by changes in only one college course. Though the main work needs to be done when students are younger, it is still important that college instructors encourage students, especially women, to pursue statistics and mathematics. Along with encouraging women to study statistics, instructors can try to follow Fox and Ronkowski’s (1999) advice to use methods that will better reach female students.

While our findings seem to suggest that women with female teaching assistants (and by implication instructors) do not necessarily learn the material better, they also suggest a strong motivational component of female instructors acting as role models. Female students with female teaching assistants were more likely to finish the course. As women continue to play an increasingly important role in politics, political science faculty should redouble their efforts to eliminate teaching practices that advantage one gender over the other. Faculties with increased numbers of female faculty will be more effective at motivating their female students to finish their programs and play a meaningful role in the world.
Notes

1. BYU is part of a category of the “Protestant Schools” that in general have more men expressing interest in majoring in political science than women (Mann 1996). At BYU approximately 38% of political science majors are women, in contrast to the national average of about 50% women.

2. A cursory analysis of teaching evaluations suggests the overt discriminatory impact of professors is negligible. The overall evaluation of the professor by women students was slightly higher than the evaluation given by male students (6.0 vs. 5.9 on a 7.0 point scale). In addition, student responses to the question “Does the instructor respond respectfully to student questions and viewpoints?” were identical for both male and female students (6.1).

3. A reflection of BYU’s religious nature and the reputation of the course is a recent student evaluation suggesting that the course be renumbered as Political Science 666.

4. Age and year in school are both controlled for because of the unique situation at BYU where most male students take two years out of their university education for church service at the age of 19.

5. We also had a measure of ACT scores, but it performed worse than GPA, suggesting that for this course, GPA rather than ACT scores is a better predictor of class performance. In addition, using ACT scores did not change the results for class performance compared to the gender match between student and teaching assistant.

6. The average scores for the final and the assignments include the students who failed to complete the course. For students who took the course during their first semester at BYU, prior GPA is measured as their GPA during that first semester.

7. The 30 independent variables include 14 dummy variables for specific teaching assistants, three dummy variables for years in school, and four for age groupings. There are three dummy variables for student gender-teaching assistant gender combinations, and single dummy variables for the professor, marital status, country of origin, and teaching assistant experience. There are also two variables for student GPA and credit hours taken.

8. For dummy variables we used the mean value rather than the median value. However, using the mean value produces nearly identical results. Females in male labs have a predicted completion rate of 86.2% and females in female labs have a predicted rate of 95.5%.

9. For the category of women in labs with male teaching assistants, the level of statistical significance was between a probability of .05 and .10 for both regressions.

References


Mann, Sheilah. 1996. “Political Science Departments Report Declines in Enrollments and Majors in Recent Years.” PS: Political Science and Politics 29 (September): 527.


