GENDER DIFFERENCES IN RATING THE TEACHING OF ECONOMICS

Kathryn H. Anderson
Vanderbilt University

and

John J. Siegfried
Vanderbilt University

INTRODUCTION

In evaluating the performance of faculty, universities usually focus on research productivity, teaching effectiveness and administrative or service contributions. While it is argued that research output can be measured adequately by the quantity and quality of publications and administrative contributions can be documented by committee assignments and other administrative appointments, the evaluation of teaching presents unique problems. In most cases, a teacher's effectiveness is measured with students' ratings. While we would like to believe that these ratings accurately reflect the amount of learning achieved in a class, it is likely that student perceptions of teaching effectiveness are affected by their expectations prior to enrolling in the class about the instructor's effectiveness, as well as by the difficulty of the course and the grading standards. If students' ratings of teaching are systematically related to their preconceptions about instructors, many teachers may be wrongly valued and promotion decisions distorted.

Various instructor characteristics may affect students' ratings of teaching. This paper focuses on the instructor's gender. Gender can affect the way an instructor teaches and the way students react to the instructor. Research by sociologists suggests that gender role socialization or biological factors may account for some of these differences. For example, research by Lever [1976] on the role play of children finds that girls are socialized to be more nurturing towards others. In sociobiological research [Delozier and Engel, 1989, for example], higher levels of male hormones such as testosterone are associated with less nurturing behavior. If these inherent gender differences affect teaching style and interaction with students, then students would likely evaluate male and female instructors differently.

The link between gender and ratings of teaching effectiveness has been examined extensively although, to our knowledge, there is only one study [Saunders, 1994] of gender differences in the rating of teaching effectiveness in economics, and it examines the ratings of teaching assistants rather than instructors. Three recent reviews of this diverse literature outside of economics [Feldman, 1992; 1993; Goodwin and Stevens, 1993] reveal little evidence of gender bias. The empirical evidence has been generated from both controlled experiments and specific evaluations of male and fe-

Eastern Economic Journal, Vol. 23, No. 3, Summer 1997

male teachers. In the laboratory experiments, instructors are described to samples of college students with descriptions differing only by the instructor's gender. These studies find no differences in the perceived quality of male and female instructors. For example, only two of 13 studies that examined overall ratings of teachers, controlling for students' gender, among other factors, find instructor gender effects [Feldman, 1992]. The same pattern emerges from 28 studies of the ratings of actual teachers, in which the average correlation between instructors' gender and students' ratings was negligible.

In spite of little difference in the overall evaluation of male and female instructors, differences in the rating of specific dimensions of teaching — enthusiasm, preparation, difficulty, for example — have been linked to gender of instructor. On 21 dimensions of teaching performance examined in the literature, Feldman [1993] reports a significant gender difference on eight, most favoring women instructors. Male instructors tend to score higher on student ratings of knowledge of subject, clarity of presentation, and personality; females tend to score higher on sensitivity, value of material, encouragement of discussion, and accessibility. In almost all cases, the average correlation between rating and instructors' gender is small and diminishes further when controls for other determinants of teaching effectiveness are included. Overall, the literature linking instructors' gender to the students' rating of teaching shows no consistent pattern favoring men or women, although students appear to believe there is a difference in the way male and female instructors teach.

In this paper, we explore the linkages between student ratings of instructors and the instructors' gender in introductory economics at 53 different colleges and universities in the United States. We find, controlling for other characteristics of the instructors and students, no difference between the ratings of male and female instructors of introductory macroeconomics, but, on all instructor dimensions, women receive higher ratings than men in introductory microeconomics.

MODEL

Our model of ratings assumes that the rating of economics instruction depends on the utility (U) a student derives from a class. If a student enjoys the class, learns from the experience, and has little difficulty absorbing the material, utility is positive and the rating of the instructor increases, *ceteris paribus*. If the classroom experience is influenced by both the characteristics of teachers that affect the learning of students (T) and by the motivation and background of students (S) that affect the ability to learn, then ratings of the instructor and the class are determined by T and S.

Assume that Y_{ijk} is rating measure k for student i in class j and that Y_{ijk} is affected by a vector of characteristics X_{ij} through the following regression:

$$Y_{ijk} = X_{ij}\beta_k + \epsilon_{ijk}, \text{ and} \tag{1a}$$

$$\epsilon_{ijk} = c_{jk} + u_{ijk}. \tag{1b}$$

c_{jk} is the systematic effect associated with the j^{th} class, and u_{ijk} is the individual random error. We estimate this model by assuming that c_{jk} is a fixed effect for each class

j and that this fixed effect measures the learning atmosphere created by the instructor and the institution; c_{jk} measures the effect of teacher or other classroom characteristics such as location and time of class (T_j) on ratings that is not captured by other exogenous variables in the model. Fixed effects are estimated jointly with the β's by including a series of dummy variables (D_j) for each of the J classes in the sample.[1] [Davidson and MacKinnon, 1993].

Because many of the explanatory variables in which we are interested are class-level or instructor variables, we first estimate the fixed effects model above excluding all class variables; ratings are regressed on the student characteristics (S_i) and the class dummy variables. Then we regress the estimated fixed effect for each class (c_{jk}) on the class variables of interest (T_j). The coefficients on these class variables measure the marginal effect of each class characteristic on the average rating of that class.

The estimated regression model is given below. Equation (2) is estimated over the sample of individual students; equation (3) is estimated over the sample of classes [Mundlak, 1978].

$$Y_{ijk} = S_i \boldsymbol{B}_k + \sum_{j=1}^{J-1} D_j \Delta_{jk} + u_{ijk} \tag{2}$$

where Δ_{jk} is the estimated fixed class effect of class j for rating measure k, and

$$\Delta_{jk} = T_{jk}\gamma_k + \eta_{jk} \tag{3}$$

where γ_k is a vector of the estimated effects of instructor and classroom characteristics on the classroom rating and η_{jk} is a random error.

DATA AND VARIABLES

We used data collected during the norming of the third edition of the Test of Understanding College Economics (TUCE III) (Saunders, 1994) to determine whether the instructor's gender affects teaching ratings. The TUCE III data include students from 189 introductory economics classes at 53 different U.S. institutions in 1990. The sample is not random, but rather "opportunistic."[2] Information on characteristics of students, instructors, the class, students' TUCE scores, and students' ratings of their instructors on several dimensions is available.[3] Our data consist of 2185 introductory macroeconomics students in 80 different classes and 2408 introductory microeconomics students in 87 different classes who rated their instructors. Ratings data were missing for the remaining 22 classes. The dependent variable (Y_{ijk}) is the students' rating of their instruction. Rating of teaching was based on a scale of lowest rating of 10 to highest rating of 50 (a 1-5 scale multiplied by 10). Students were asked to rate the course on four dimensions: amount learned, interest of subject matter, importance of subject matter, and difficulty of subject matter. They rated instructors on five dimensions: enthusiasm, preparation for class, ability to speak English well, rigor of grading standards, and overall teaching effectiveness. We estimated the model for all nine ratings.

The student characteristics (S_{ij}) include demographic traits and measures of the potential difficulty a student might have with economics. The demographic measures include student gender and race (black, hispanic, Asian, other ethnic, relative to white). The student characteristics to measure the potential difficulty with the course are college grade point average prior to taking the sample economics class and the number of hours of calculus the student had completed prior to the sample class. The instructors' characteristics (T_j) which can affect student learning are gender and race (Asian or non-Asian), whether the instructor's native language is English, and the number of years of instructor teaching experience. We also include a measure of the number of students in the class.[4] We have no unambiguous predictions about the effects of the instructor or student demographic characteristics on ratings.

RESULTS

The means and standard deviations of the variables in the model are reported in Table 1. Eleven percent of macroeconomics instructors, 18 percent of microeconomics instructors, and 43 percent of students are female.[5] In addition, on average, ratings are high. The average instructor received a rating of about 40 (above midpoint) or slightly higher on every instructor dimension. The ratings of the classes are lower, on average, than the ratings of the instructors but are still above 30 on all rated dimensions. The students' interest in economics received the lowest average rating with a 33 in macroeconomics and a 34 in microeconomics.

Course rating results are reported in Table 2 (microeconomics) and Table 3 (macroeconomics). Instructor rating results are in Table 4 (microeconomics) and Table 5 (macroeconomics). The effects of student characteristics on ratings were obtained by estimating Equation (2). Effects of the class-specific characteristics were obtained by estimating Equation (3). The estimated fixed effects are available from the authors on request.

With the exception of instructors' gender, few of the instructor or student level controls has much effect on ratings of microeconomics. On the course ratings (difficulty, amount learned, interest in, and importance of the course content), instructor gender affects only the difficulty rating in microeconomics. The courses taught by women are rated as more difficult by almost two (out of a range of 40) points. Instructor gender is not an important factor in course ratings of macroeconomics classes. On the instructor ratings in microeconomics, women are rated higher than men on all dimensions. Women are perceived as more enthusiastic in the classroom (three point difference), better prepared (two point difference), speaking better English (two point difference), having higher standards (1.6 point difference), and as better instructors overall (2.4 point difference). In macroeconomics classes, no instructor gender differences are observed in the student ratings.[6]

We also found some effects of the students' gender on ratings, although these influences were smaller than the effects of the instructors' gender. In rating both subjects, women students are, on average, less interested in economics, find the material less important, and find economics more difficult than men. On ratings of the instructors, women students are more positive than men. Women tend to give instructors higher scores on preparation, grading standards, and English skills in both

TABLE 1
Descriptive Statistics

Variables	Macroeconomics[a]	Microeconomics[a]
Dependent[b]		
Amount learned	36.32 (8.05)	37.11 (7.64)
Interest	33.18 (9.83)	34.11 (9.09)
Importance	37.97 (7.90)	37.36 (7.94)
Difficulty	37.88 (8.05)	37.14 (8.20)
Instructor enthusiasm	41.68 (7.25)	42.24 (7.44)
Instructor preparation	42.78 (7.12)	43.72 (6.68)
Instructor English	44.96 (7.20)	45.63 (6.75)
Instructor standards	38.94 (7.66)	40.00 (7.41)
Instructor overall	39.48 (8.72)	40.92 (8.02)
Independent		
Instructor gender (1=female)	0.11 (0.31)	0.18 (0.38)
Instructor Asian (=1)	0.09 (0.29)	0.04 (0.20)
Native English speaker (=1)	0.91 (0.29)	0.94 (0.24)
Years teaching experience	12.85 (9.32)	12.81 (8.55)
Class size (number of students)	91.16 (77.45)	67.14 (44.74)
Student gender (1=female)	0.43 (0.50)	0.43 (0.50)
Student black (=1)	0.04 (0.18)	0.04 (0.19)
Student hispanic (=1)	0.02 (0.15)	0.02 (0.13)
Student asian (=1)	0.03 (0.18)	0.04 (0.19)
Student other ethnic (=1)	0.01 (0.08)	0.01 (0.08)
Grade point average (out of 4 points)	2.92 (0.55)	2.91 (0.58)
Hours of calculus	2.58 (3.00)	2.86 (3.11)

a. There were 2184 students in 80 classes in the macroeconomics sample and 2408 students in 87 classes in the microeconomics sample.
b. The statistics in the table are the mean and standard deviation (in parentheses).

macroeconomics and microeconomics. There are no student gender differences on enthusiasm of the instructor.[7]

The rating of economics instruction, at least at the introductory level, appears related to the gender of both instructors and students. After careful examination of several dimensions of the rating of teaching, controlling for many characteristics of students and instructors that could affect the learning experience in the classroom, we find that students, on average, rate women microeconomics instructors higher than men. These results are consistent with some of the literature on student ratings of teaching in other disciplines. The results may indicate that students think women are, on average, better microeconomics teachers than men. An alternative interpretation, however, is that men may be discriminated against by students and are actually better teachers than their ratings suggest.

We tried to determine which of these two interpretations of our results is more likely by examining the students' scores on the TUCE- III exams they all took. We used two measures of achievement: the score on the TUCE-III exam taken at the end

TABLE 2
Regression Models of Course Ratings: Microeconomics

Independent Variables	Amount Learned	Interest	Importance	Difficulty
Student characteristics [estimates of Equation (2)]:				
Constant	34.75[a]	32.86[a]	38.26[a]	41.91[a]
	(2.04)	(2.44)	(2.22)	(2.07)
Female (=1)	-1.04[a]	-2.81[a]	-0.82[a]	2.13[a]
	(0.23)	(0.38)	(0.34)	(0.32)
Black (=1)	-0.91	1.47	0.81	-0.42
	(0.86)	(1.02)	(0.91)	(0.88)
Hispanic (=1)	-1.60	-0.96	-0.23	-0.88
	(1.20)	(1.43)	(1.27)	(1.22)
Asian (=1)	-0.53	1.06	0.66	0.50
	(0.83)	(0.99)	(0.88)	(0.84)
Other ethnic (=1)	-0.70	1.22	-0.60	2.63
	(1.86)	(2.22)	(1.96)	(1.89)
Hours calculus	0.04	0.10	-0.12[b]	-0.54[a]
	(0.06)	(0.07)	(0.06)	(0.06)
Grade point average × 100	0.144[a]	0.007[b]	0.005[c]	-0.133[a]
	(0.003)	(0.003)	(0.003)	(0.003)
Instructor characteristics [estimates of Equation (3)]:				
Constant	-5.02[b]	-3.75	-3.29	-2.87
	(2.16)	(2.33)	(2.16)	(2.37)
Female (=1)	1.23	0.48	0.42	1.70[c]
	(0.86)	(0.92)	(0.86)	(0.94)
Asian (=1)	1.30	0.98	0.22	0.51
	(2.44)	(2.63)	(2.44)	(2.68)
Native English speaker (=1)	2.15	2.09	-0.17	0.35
	(2.20)	(2.38)	(2.20)	(2.42)
Years teaching experience	0.33[a]	0.32[b]	0.25[c]	0.03
	(0.13)	(0.14)	(0.13)	(0.15)
Experience squared	-0.01[a]	-0.01[b]	-0.01	-0.00
	(0.00)	(0.01)	(0.00)	(0.01)
Class size	-0.00	-0.00	-0.00	0.02[c]
	(0.01)	(0.01)	(0.01)	(0.01)

a. Significant at the 1% level.
b. Significant at the 5% level.
c. Significant at the 10% level.

of the course (post-TUCE) and the difference between the post-TUCE score and the score earned at the beginning of the course (value-added). Value-added controls for the students' initial knowledge of economics. We tested for mean differences in these two measures by gender of the instructor. The instructor differences, however, do not present us with a clear pattern. In micro, the average student's post-TUCE score in courses taught by men is 1.5 points higher than the post-TUCE score in courses taught by women, but there is no difference in value-added.[8] In macro, both the students' post-TUCE and value-added measures are higher in courses taught by men; students of male instructors score about one point higher on the post-TUCE exam.[9]

TABLE 3
Regression Models of Course Ratings: Macroeconomics

Independent Variables	Amount Learned	Interest	Importance	Difficulty
Student characteristics [estimates of Equation (2)]:				
Constant	27.61[a]	27.43[a]	35.03[a]	43.16[a]
	(1.08)	(1.32)	(1.10)	(1.05)
Female (=1)	-1.66	-2.86[a]	-0.72[b]	2.09[a]
	(0.34)	(0.42)	(0.35)	(0.34)
Black (=1)	-1.15	-0.99	0.03	-0.70
	(1.05)	(1.30)	(1.07)	(1.03)
Hispanic (=1)	0.72	1.35	1.88	0.29
	(1.17)	(1.43)	(1.19)	(1.14)
Asian (=1)	-0.12	1.24	0.02	0.04
	(0.99)	(1.22)	(1.01)	(0.97)
Other ethnic (=1)	-0.50	-0.19	-5.37[a]	0.01
	(2.02)	(2.46)	(2.05)	(1.97)
Hours of calculus	0.03	0.04	-0.16[a]	-0.58[a]
	(0.06)	(0.08)	(0.07)	(0.06)
Grade point average × 100	0.025[a]	0.017[b]	0.009[a]	-0.006[b]
	(0.003)	(0.004)	(0.003)	(0.003)
B				
Constant	4.29[a]	5.21[a]	2.04	-6.35[a]
	(1.68)	(1.95)	(1.60)	(1.59)
Female (=1)	0.75	0.89	0.70	-1.30
	(1.16)	(1.34)	(1.09)	(1.09)
Asian (=1)	-1.66	-2.50	0.16	-1.19
	(1.54)	(1.78)	(1.45)	(1.45)
Native English speaker (=1)	2.51[c]	2.19	1.28	1.97
	(1.35)	(1.56)	(1.27)	(1.27)
Years teaching	-0.51[a]	-0.52[b]	-0.35[b]	0.29[c]
	(0.18)	(0.21)	(0.17)	(0.17)
Experience squared	0.02[a]	0.02[b]	0.01[a]	-0.01[c]
	(0.01)	(0.01)	(0.00)	(0.00)
Class size	-0.02[b]	-0.03[a]	-0.02[c]	0.01
	(0.01)	(0.01)	(0.01)	(0.01)

a. Significant at the 1% level.
b. Significant at the 5% level.
c. Significant at the 10% level.

The pattern of exam scores implies that the higher ratings for women micro instructors is not due to students learning more in the classes taught by women. The higher overall ratings of female micro instructors could reflect students' perceptions that the learning environment in female instructors' classrooms is more pleasant. Or it could reflect students' beliefs that they are learning things in female instructors' classrooms that are not measured by the TUCE-III exam. In macro, students appear to learn more in classes taught by men, but students do not rate the male instructors higher than the female instructors. Men may be better teachers than their ratings

TABLE 4
Regression Models of Instructor Ratings: Microeconomics

Independent Variables	Enthusiasm	Preparation	English	Standards	Overall
Student characteristics [estimates of Equation (2)]:					
Constant	41.91[a]	40.10[a]	43.79[a]	35.73[a]	37.72[a]
	(1.82)	(1.73)	(1.58)	(2.01)	(2.02)
Female(=1)	0.18	1.00[a]	0.71[a]	0.71[b]	-0.09
	(0.28)	(0.27)	(0.25)	(0.31)	(0.31)
Black (=1)	0.99	-0.46	-0.36	0.39	0.66
	(0.77)	(0.73)	(0.67)	(0.85)	(0.86)
Hispanic	1.67	1.56	0.11	1.90	0.98
(=1)	(1.07)	(1.02)	(0.93)	(1.18)	(1.19)
Asian (=1)	-0.33	-0.22	-0.40	0.71	0.99
	(0.74)	(0.71)	(0.65)	(0.82)	(0.83)
Other ethnic	1.39	-1.59	0.10	1.14	-0.63
(=1)	(1.71)	(1.62)	(1.48)	(1.89)	(1.90)
Hours of	-0.03	-0.04	-0.07[c]	-0.09	0.02
calculus	(0.05)	(0.05)	(0.04)	(0.06)	(0.06)
Grade point	0.002	0.000	-0.002	0.000	0.002
average X 100	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
Instructor characteristics [estimates of Equation (3)]:					
Constant	-4.26	0.42	-10.24[a]	0.10	-1.48
	(2.71)	(2.05)	(1.73)	(1.72)	(2.89)
Female(=1)	3.00[a]	1.89[b]	2.01[a]	1.62[b]	2.36[b]
	(1.08)	(0.81)	(0.69)	(0.69)	(1.15)
Asian (=1)	-2.27	-3.43	-2.26	-0.38	-1.38
	(3.07)	(2.32)	(1.95)	(1.95)	(3.27)
Native English	1.52	0.72	11.44[a]	3.05[c]	2.90
speaker (=1)	(2.77)	(2.09)	(1.76)	(1.76)	(2.95)
Years teaching	0.48[a]	0.23[c]	0.28[a]	0.08	0.31[c]
experience	(0.17)	(0.13)	(0.11)	(0.11)	(0.18)
Experience	-0.02[a]	-0.01[b]	-0.01[a]	-0.00	-0.01[b]
squared	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Class size	-0.00	0.02	0.01	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)

a. Significant at the 1% level.
b. Significant at the 5% level.
c. Significant at the 10% level.

suggest, or the ratings may reflect student disutility from the less pleasant learning environment they perceive in male instructors' classes.

We also find that student interest in economics and difficulty with the material is related to the gender of students. Women students think economics is more difficult, report being less interested in economics, and find the material less important than do men. This negative reaction by women students to economics is reinforced by their poorer performance on the post-TUCE exam and their lower value-added than male students. The post-TUCE scores for women were 1.6 points lower in micro and 2.1 points lower in macro. Women students appear to learn less than men, and this may

TABLE 5
Regression Models of Instructor Ratings: Macroeconomics

Independent Variables	Enthusiasm	Preparation	English	Standards	Overall
Student characteristics [estimates of Equation (2)]:					
Constant	39.74[a]	38.08[a]	44.57[a]	37.71[a]	2.46[a]
	(0.96)	(0.94)	(0.85)	(1.08)	(1.14)
Female(=1)	-0.13	0.66[b]	0.74[a]	0.62[c]	-0.60[c]
	(0.30)	(0.30)	(0.27)	(0.34)	(0.36)
Black (=1)	-2.76[a]	-1.32	-1.58[c]	-0.31	-1.72
	(0.93)	(0.92)	(0.83)	(1.05)	(1.11)
Hispanic	1.48	0.32	0.40	-0.03	0.53
(=1)	(1.04)	(1.02)	(0.92)	(1.17)	(1.24)
Asian (=1)	-0.87	-1.69[b]	-0.83	-0.71	0.30
	(0.88)	(0.87)	(0.79)	(1.00)	(1.06)
Other ethnic	0.77	3.37[c]	1.03	2.31	1.41
(=1)	(1.79)	(1.76)	(1.59)	(2.02)	(2.14)
Hours of	-0.01	-0.06	-0.02	-0.08	0.02
calculus	(0.06)	(0.06)	(0.05)	(0.06)	(0.07)
Grade point	0.002	-0.002	0.005[b]	-0.005	0.002
average × 100	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Instructor characteristics [estimates of Equation (3)]:					
Constant	-0.58	3.94[a]	-10.26[b]	1.32	6.28[a]
	(1.67)	(1.54)	(0.90)	(1.16)	(2.04)
Female (=1)	1.09	-1.43	0.04	0.24	-0.07
	(1.15)	(1.06)	(0.62)	(0.80)	(1.41)
Asian (=1)	-0.88	-0.92	-2.45[a]	-1.07	-0.43
	(1.53)	(1.41)	(0.83)	(1.06)	(1.87)
Native English	1.96	2.82[b]	11.42[a]	1.41	3.01[c]
speaker (=1)	(1.34)	(1.24)	(0.72)	(0.93)	(1.64)
Years teaching	0.21	0.01	-0.13	0.05	-0.12
experience	(0.18)	(0.17)	(0.10)	(0.12)	(0.22)
Experience	-0.01	-0.00	0.00	-0.00	0.00
squared	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
Class size	-0.01	-0.01	-0.00	-0.01	-0.02[c]
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R-square	.12	.13	.84	.10	.09

a. Significant at the 1% level.
b. Significant at the 5% level.
c. Significant at the 10% level.

partly explain their lower ratings of economics classes (but not instructors). These negative reactions by women students to economics may be partly responsible for the fact that only three out of ten economics majors in the U.S. is female (Siegfried, 1995) while about 5.5 out of 10 college and university students are women.

When compared to student learning, the evidence we summarize from student course ratings of 167 different introductory economics classes reveals no evidence of student bias against female instructors. If anything, there is some evidence that in micro students rate female instructors higher than male instructors while learning

similar amounts from each and, in macro, students rate male and female instructors similarly in spite of learning less in the classes of women.

NOTES

The authors wish to thank Robin Bartlett, Rebecca Blank, Paul Pieper, Barbara Kilbourne and colleagues in the Department of Economics and Business Administration at Vanderbilt University for helpful comments on previous versions of this paper.

1. One class is omitted from the set of dummy variables; it is the base against which we compare all other class effects.
2. An "opportunistic sample" is a nonrandom sample that is not generated by any probability rules [Becker, 1995, 5].
3. The TUCE is a 30 (or 33 if international economics is included) question exam. It was administered at the beginning of the course and again at the end in order to obtain a measure of the amount of learning (in contrast to the level of understanding at the end).
4. Because students who performed poorly on the pretest dropped the course more frequently in larger classes, relatively more dissatisfied students are missing from the course evaluation sample in the larger classes. This will tend to obscure any negative relationship between learning and class size [Becker, Powers, and Saunders, 1996].
5. Among the 29 macroeconomics instructors with less than 7 years of experience, six were women (21 percent) while only four of the 51 more experienced instructors were women. In the microeconomics classes, 43 percent of inexperienced instructors and 14 percent of experienced instructors were women. This difference in the gender composition by experience is largely due to the increase in women economics instructors over time in general.
6. We were concerned that our gender effects were capturing the effects of differential experience by gender. To see if this was a significant problem, we estimated the regressions over classes with only experienced (more than six years of teaching experience) or inexperienced (six or fewer years of teaching experience) instructors. The results were almost identical to the results in Tables 2-5. Women received significantly higher ratings on all instructor dimensions, except the overall rating, in microeconomics classes, and instructor's gender had no effect on ratings in macroeconomics classes.
7. One hypothesis to explain the significantly higher instructor ratings received by women instructors is that women students, who give significantly higher average ratings to all instructors than do men students, may enroll disproportionately in classes taught by women instructors. To assess this possibility, we cross-tabulated instructor gender and student gender for both microeconomics and macroeconomics. There is no evidence that women students are more likely to sort into microeconomics classes taught by women (Chi-square=.0965, P-value=.756). For macroeconomics, however, there is evidence of sorting; women students are significantly more likely to enroll in macroeconomics classes taught by women (Chi-square=4.443, p-value=.035).

   To further explore the implications of same gender rating in enrollment decisions, we included an interaction term (the product of the proportion of female students in the class and the instructor's gender) in the second stage regression of fixed effects on class characteristics. This term was insignificant in both the micro and macro regressions, providing no evidence that female students gave female instructors higher ratings.
8. The standard deviation of the mean difference in post-test scores is 0.31; the standard deviation of the mean difference in value-added is 0.26.
9. The standard deviation of the mean difference in post-test scores is 0.38; the standard deviation of the mean difference in value-added is 0.37.

REFERENCES

Becker, W.E., Jr. *Statistics for Business and Economics*. Cincinnati, OH: South-Western College Publishing, 1995.

Becker, W.E., Jr., Powers, J.R., and Saunders, P. Problems of Missing Student Data in the TUCE III Data Set and the Importance of Class Size in Student Achievement. Unpublished manuscript. Indiana University, July 1996.

Davidson, R. and MacKinnon, J. G. *Estimation and Inference in Econometrics*. New York: Oxford University Press, 1993.

Delozier, C. D. and Engel, E. Sexual Differentiation as a Model for Genetic and Environmental Interaction Affecting Physical and Psychological Development, in *The Fundamental Connection Between Nature and Nurture,* edited by W. Gove and R. Carpenter. Lexington, MA: Lexington Books, 1989.

Feldman, K. A. College Students' Views of Male and Female College Teachers: Are Female and Male Professors Rated Differently? *Research in Higher Education*, June 1992, 317-75.

_____. College Students' Views of Male and Female College Teachers: Part II. Evidence From Students' Evaluations of Their Classroom Teachers. *Research in Higher Education*,April 1993,151-211.

Goodwin, L.D. and Stevens, E.A. The Influence of Gender on University Faculty Members' Perceptions of "Good" Teaching. *Journal of Higher Education*, March/April 1993, 166-85.

Lever, J. Sex Differences in the Games Children Play. *Social Problems*, April 1976, 478-87.

Mundlak, Y. On the Pooling of Time Series and Cross Section Data. *Econometrica,* January 1978, 69-86.

Saunders, P. *The TUCE III Data Set: Background Information and File Codes*. New York: National Council on Economic Education, 1994.

_____. The Influence of Gender, Nationality, Experience, and 'Warm-up' on Beginning Instructor Performance. Paper presented at the March 1994 meetings of the Midwest Economics Association.

Siegfried, J. Trends in Undergraduate Economics Degrees: a 1993-94 Update. *Journal of Economic Education*, Summer 1995, 282-87.