The Impact of Gender on the Evaluation of Teaching: What We Know and What We Can Do

Heather Laube, Kelley Massoni, Joey Sprague, Abby L. Ferber

NWSA Journal, Volume 19, Number 3, Fall 2007, pp. 87-104 (Article)

Published by Johns Hopkins University Press

For additional information about this article
https://muse.jhu.edu/article/224753

Access provided by Dickinson College (29 Jul 2016 16:36 GMT)

The Impact of Gender on the Evaluation of Teaching: What We Know and What We Can Do

HEATHER LAUBE, KELLEY MASSONI, JOEY SPRAGUE, and ABBY L. FERBER

The importance of teaching evaluations to the tenure and promotion of women faculty cannot be underestimated. Administrators routinely consider classroom teaching in hiring, tenure, promotion, and salary decisions and increasingly rely most heavily on quantitative student ratings. Scholars who have attempted to determine whether/how gender enters into students' evaluations of their teachers generally fall into two camps: those who find gender to have no (or very little) influence on evaluations, and those who find gender to affect evaluations significantly. Drawing on insights developed from sociological scholarship on gender and evaluation, we argue that the apparent inconsistency on the question of whether student evaluations are gendered is itself an artifact of the way that quantitative measures can mask underlying gender bias. We offer concrete strategies that faculty, researchers, and administrators can employ to improve the efficacy of the system of evaluation.

Keywords: teaching evaluations / gender / faculty assessment / tenure and promotion

Not long ago, an Associated Press news story with the headline, "Women a Minority of Tenured Faculty and Administrators" began: "Women comprise 58 percent of the nation's 13 million college undergraduates and, in 2002, earned more doctorates than men. They're a dominant force on campuses—until they receive a degree" (*Giegerich* 2004). Various facts support this statement: Women's professional preparation has not translated into a higher number of tenured women faculty, and women assistant professors are 23 percent less likely than men to become associate professors (Cook 2004, 16). In addition, while men and women make roughly equal starting salaries, over time a wage gap appears across universities and disciplines. Moreover, there is greater movement of women than men into part-time teaching positions, women's advancement is slower, they receive fewer awards and prizes, and they are especially underrepresented at top-tier institutions (Valian 2005).

In large part due to the feminist movement, large numbers of women have moved into higher education, and feminist scholars have contributed to significant changes in their disciplines and institutions. For example, in 1969 only 16 women's studies course syllabi existed in the United States, while by 1982 there were some 20,000 courses and 450 certificate

or degree-granting programs. However, although there were well over 600 academic women's studies programs in 2004 (Stimpson 1982; CreativeFolk.com 2007), feminist academics have not succeeded in truly transforming the masculine structure of academe. One area that continues to require feminist attention is the evaluation of teaching. Because research points to the importance of students' gender- and race-based expectations of their professors and student evaluation of teaching plays a role in most tenure and promotion decisions, the significant consequences for women faculty must be addressed.

Gender and the Evaluation of Teaching

Teaching evaluations are increasingly important to the careers of faculty. In fact, reliance upon student evaluations has been steadily increasing. For example, Peter Seldin (1999) surveyed the academic deans of all 740 non-university-related undergraduate liberal arts colleges listed in the *Higher Education Directory* in 1997 on their practices for evaluating faculty. Of the 598 who responded, 97 percent identified classroom teaching as a major factor in evaluating overall faculty performance at their institutions. The next most frequently cited major factor was student advising, named by 64.2 percent of these academic deans. Increasingly, administrators are assessing classroom performance through the use of quantitative student ratings: 88.1 percent reported that they "always used" systematic student ratings of teaching in evaluating teachers, up from 80.3 percent in 1988 and only 54.8 percent in 1978. In comparison, only 38.6 percent of the deans reported using course syllabi and exams and 40.3 percent reported using classroom visits in evaluating teaching. The increasing reliance upon teaching evaluations in personnel decisions prompts the following question: *(How) does gender enter into students' evaluations of their teachers?*

Scholars who have attempted to answer this question are divided in their findings. Reviews of the research literature on the student evaluation of teaching could easily give the impression that gender is not a major factor in the process (e.g., Aleamoni 1999; Feldman 1993, 1992; Fernandez and Mateo 1997; Freeman 1994; Wheeless and Potorti 1989). In fact, a respected guide for administrators on how to evaluate teaching specifically recommends they *not* take gender into account in interpreting student ratings, concluding that research revealed that, if anything, students rated women teachers higher than men (Cashin 1999). The only citations in the guide to support this claim are two papers by Kenneth Feldman (1993, 1992) reporting separate meta-analyses of experimental studies and studies of students' evaluations of actual professors. Feldman concluded that direct effects of gender on evaluation were minor, trivial

in size, and in the case of evaluations of actual professors, favored women rather than men.

We argue that a more careful reading of the research literature reveals that the evidence is more mixed. For example, JoAnn Miller and Marilyn Chamberlin (2000) argue that Feldman's methodology may have depressed findings of gender effects because it combines information from qualitatively different sources. For example, Feldman averages correlations between gender and student rating of the teacher that are quite disparate and aggregates ratings from studies that vary by discipline, type of institution, and unit of analysis.

One review reports that "a majority of studies" find no relationship between teacher gender or student gender and student ratings (Aleamoni 1999, 156). But this claim of no gender difference is supported only by the referencing to seven articles cited in a single article from 1971. Lawrence Aleamoni also reports that while there is very little research on whether overall items are consistent with specific rating scales (e.g., "teacher is available to students outside of class," vs. "teacher is available at least x hours a week outside of class"), the five studies that examine this question do find that global assessments are not strongly related to more specific measures and are more highly correlated with gender, status, and other contextual factors than are specific measures. Given that personnel decisions tend to rely heavily on these *overall* assessment scores (Seldin 1999), Aleamoni is implicitly calling attention to gender biases in common practices.

However, as many other studies show, gender affects student evaluations in several ways. Some researchers argue it is not the sex of the professor but instead the degree to which the professor's personality meets or escapes traditional notions of gender that makes a difference in the kinds of ratings students give. Harvey Freeman (1994) found that students expected a hypothetical teacher described in androgynous terms (i.e., as having both feminine and masculine attributes) to be more effective than one described in either masculine or feminine terms.[1] Virginia Wheeless and Paul Potorti (1989) reported that students' descriptions of their teachers as androgynous corresponded to their having more positive feelings about a course, believing they were learning more in it, and intending to pursue similar courses in the future.

The impact of the androgynous personality effect on students' evaluations, however, sometimes depends on whether the teacher is a woman or a man. James Bray and George Howard (1980) gave 100 instructors the Bem Sex Role Inventory (BSRI) questionnaire and selected 12 teachers (6 men and 6 women) to represent the respective categories of masculine, feminine, and androgynous (36 teachers total). Their students were then given a survey to measure satisfaction with the teacher and perceived progress in the class. Overall, Bray and Howard found that students gave

androgynous teachers higher ratings than either masculine or feminine teachers. However, androgyny was more rewarded in women teachers than in men teachers. Students gave higher evaluations to women teachers who were androgynous than to those who were masculine or feminine; however, they did not make this kind of distinction with their men teachers. This difference may occur because women teachers are held to a higher standard of teaching by their students and are asked, in effect, to exhibit personalities that incorporate both masculine and feminine attributes in the classroom.

We can see several sources of confusion in this research literature. For one, it seems reasonable to expect that contextual factors influence students' responses to their teachers, yet few studies take contextual factors into account. One contextual factor highly salient to most students is grades. Clayton Tatro (1995) found that students' expected grades were significantly related to their rating of the professor, and there is reason to suspect that this relationship is gendered. Lisa Sinclair and Ziva Kunda (2000) report that students who received better grades gave their teachers higher evaluations, while low grades disproportionately reduced the ratings of women instructors in comparison to male instructors.

Another source of varying findings seems to be whether evaluations of teachers are assessed in quantitative or qualitative terms. The survey findings of Christine Bachen, Moira McLoughlin, and Sara Garcia (1999) are particularly striking in this regard. When male students were asked to rate their experience with a male or female instructor based on the use of quantitative scales, male students' responses were not affected by the instructor's gender. However, when later given the opportunity to answer an open-ended question about gender differences in teaching, half of the male students volunteered that women were not as professional or challenging as men. These kinds of qualitative investigations can often identify gendered relational dynamics that quantitative assessments are unable to detect.

Frankly, as sociologists specializing in gender, we are puzzled by conclusions that gender has no impact on teaching evaluations. Three decades of scholarship have found gender to be a significant factor in shaping interactions, practices, and outcomes in every major realm of human social life: family, work, science, medicine, religion, sports, and popular culture—to mention just a few (see, for example, the review of research in Chafetz 1999, and Ferree, Lorber, and Hess 1999). Why would the classroom be any different? As Lana Rakow notes, "We cannot set aside the social relationships of the larger world—a world in which classifications of gender, race, and class are among the most paramount—as we take up the more temporary relationship of professor and student" (1991, 10). We believe that taking into account what researchers have learned about how gender operates in other contexts provides a methodological explanation for why

many studies on gender and the evaluation of teaching conclude there is no gender difference.

Social Science Scholarship on Gender and Evaluation

Candace West and Don Zimmerman (1987) first coined the term "doing gender" to emphasize that gender is less an attribute than a performance, something we do according to culturally defined scripts for masculinity and femininity. Since then, many scholars have demonstrated that we hold one another accountable for the gender-appropriateness of our performance and we do gender because we *know* we will be held accountable to these standards (Biernat and Kobrynowicz 1997; Connell 1995; Kobrynowicz and Biernat 1997; West and Fenstermaker 1995).

One particularly relevant area in which these gendered evaluation processes can be observed is in the review of work and workers. Comparable worth studies have revealed that perceptions of job demands and the assessment of the performance and promise of a worker are highly gendered (e.g., England et al. 1994; Ferree and McQuillan 1998; Martin 1996; Steinberg and Haignere 1987). Linda Carli (2001) provides an extensive review of the literature on gender differences in influence, demonstrating that gender does affect the degree of social influence individuals are perceived to have. The literature supports the contention that women's degree of social influence is more conditional than men's and is based on the degree to which their communication and leadership styles correspond to accepted feminine styles. Studies of decision-making groups show that except in gender-stereotyped feminine tasks, people (men especially) discount the contributions of women and are less willing to be influenced by them, particularly women who do not conform to traditional gender expectations (Carli 1990; Ridgeway 1987). Carli notes that the broad and consistent message across all the studies in this area is that all people tend to discount women's skills and effort, are not comfortable with women in positions of power, and respond poorly to women who overstep their culturally assigned bounds.

Extrapolating from this literature to the classroom, men professors need to live up to race- and class-specific expectations of "man" and "professor," whereas women professors need to live up to similarly specific expectations of "woman" and "professor." However, the overlap between expectations for the categories of "man" and "professor" is considerably greater than for "woman" and "professor" (Martin 1984; Rakow 1991). When Alyson Burns-Glover and Dale Veith (1995) asked college students to rate the desirable traits for a hypothetical professor, referred to as either "Sam," "Sarah," or "Dr. Lawson," students selected similar traits for "Sam," ostensibly a man professor, and the gender-neutral "Dr.,"

but diverged from this pattern when they selected desirable traits for "Sarah," the woman professor. For these students, "man" and "Dr." were synonymous. Similarly, Bachen, McLoughlin, and Garcia (1999) found that college students most often contrasted women professors to men professors—using men professors as their referent or standard from which women deviate (Martin 1984). In the classroom, as in the social world, men are still the norm.

Similarly, many college students in another study perceived their men instructors (even graduate student teaching assistants) to be "professors" and their women instructors (even full professors) to be "teachers" (Miller and Chamberlin 2000). As sociologist Michael Messner observes, "White male professors are likely beginning from a position of assumed and automatically accepted authority and respect" (2000, 460). If "man" does equal "professor" for most students, then it is not surprising that both men and women students often rate their men instructors similarly, while diverging in their ratings of women instructors (Centra and Gaubatz 2000; Basow 1995; Basow and Silberg 1987). Philomena Essed (2000) finds the increased questioning of authority (by students and colleagues) is even worse for faculty who are women of color. White students may assume the right to put women of color professors "in their place," and these professors may feel the need to adjust their teaching styles in order to protect themselves from negative evaluations and student complaints (Brown 2002; Luthra 2002).[2] Women faculty of color cannot take their authority as a given simply because of their position. As Messner frames it, the issue is not only that women and people of color are often "'graded down' because of gender and/or racial bias students hold; it is also that white males are actually being 'graded up'" (2000, 458).

This sociological research is corroborated by the psychological literature on social cognition that explores gender-specific evaluation processes under the rubric of "shifting standards." Whenever we are called on to make a judgment, we do so in relation to some point of reference. When an evaluation concerns a behavior or attribute that resonates with race or gender stereotypes, these stereotypes influence the standard or context used to judge a particular member of the group (Biernat 1995; Biernat and Kobrynowicz 1997; Biernat and Manis 1994; Biernat, Manis, and Nelson 1991; Kobrynowicz and Biernat 1997; Kobrynowicz and Biernat 1998). A study by Sheila Bennett (1982) points to a gendered shift in the standard being applied to evaluate teachers. She surveyed undergraduate students on how much personal attention they both expected and received from their instructors, as well as how they would rate these teachers on availability outside of class. Students expected and reported getting more personal time from women than from men teachers and yet were more likely to rate women teachers as not available enough. These students' reference point for "enough availability" clearly shifted to a higher order for women teachers (Burns-Glover and Veith 1995).

Further, both the "doing gender" and the "shifting standards" models suggest that shifts in standards of evaluation may be qualitative as well as quantitative. That is, the stereotypes that we hold for a group can influence our understanding of the meaning of a trait in members of that group. In the case of teacher evaluations, students' gender stereotypes are apt to "shift" not only their baseline expectations for their instructors' traits, but also their perceptions of what those traits entail. Thus, students may expect a female teacher to engage in a different set of behaviors to satisfy a particular standard than they would expect of her male counterpart. For example, students may expect a female teacher to spend office hours walking them through a task when they might expect a male teacher to give only brief directions in class. Additionally, in the name of "racial solidarity," students of color may have different expectations of women faculty of color, including additional time demands and, occasionally, expectations of lenient treatment (Essed 2000), and when an ethnic teacher challenges an ethnic student, she may be viewed as unsupportive of others "like her" (TuSmith 2002).

In summary, the wide range of research on how gender shapes other assessment processes should lead us to expect that the evaluation of teaching is also gendered. We suspect that the frequent failure to find gender differences in student evaluation of teachers is an artifact of the method used to look for them. In the typical approach, students are asked to rate a real or hypothetical teacher on some range of attributes using a quantitative scale, often ranging from one to five or one to six. Then the researchers compare the mean ratings for men teachers to those for women teachers. It is not surprising that most research has taken this approach. It parallels the way we usually ask students to evaluate teachers in the real world—we give them a list of traits and behaviors and ask them to rate their teacher on each using a numerical scale. We then take the mean scores of students on each item as an indication of teacher performance.

This approach to evaluating teaching or to studying whether and how gender enters into the evaluation process is based on two assumptions that the research literature suggests are untenable. First, it assumes a universal metric, that a "three" is a "three" and a "five" is a "five," no matter who the teacher is. Second, it assumes that a specific rating corresponds to equivalent behaviors or abilities across teachers. But if, as this research suggests, students are using different baselines for men and women, or they are drawing on totally different behaviors to evaluate a trait, quantitative studies will not be able to detect these gender differences. This would also explain why the research that does find gender differences tends to be qualitative (e.g., Basow 2000; Bachen, McLoughlin, and Garcia 1999; Siskind and Kearns 1997; Sprague and Massoni 2005).

Before we can reliably address the question of whether student rating scales are producing gendered outcomes for teachers, we must first resolve the question of whether the process of student ratings might itself be

reproducing gender. We must ask: *Are students seeing teachers in gendered ways, holding them accountable to gendered criteria?* Our hypothesis is that the frameworks students use to perceive and evaluate teachers vary to some extent with the gender of the teacher.

During the spring and fall semesters of 1998, Joey Sprague and Kelley Massoni (2005) surveyed students in lower division sociology and psychology courses at two public universities. Among other questions, students were asked to print up to four adjectives to describe their best- and worst-ever teacher. While Sprague and Massoni found considerable overlap in the ways students talk about their male and female teachers, they also found some signs of gender divergences. Students remembered their best men teachers as especially funny and personable, while they lauded their best women teachers for having been fun and nurturing. Others' research findings suggest that for women professors, signaling "fun" to students necessitated lots of smiling and eye contact, resulting in higher student evaluations. However, when a man professor smiled similarly, student evaluations were lowered (Kierstead, D'Agostino, and Dill 1988; Martin 1984).

Sprague and Massoni (2005) report that students' memories of their worst-ever teachers appeared to be more emotionally charged than their memories of their best-ever teachers. In remembering their worst men teachers, students used words like arrogant, disengaged, and pretentious. They saved their harshest critiques, however, for their worst women teachers, who were chastised for having been unfair, cold, mean, and either too intelligent or not intelligent enough. Perhaps most disturbingly, worst women teachers were sometimes seemingly indicted for being bad women, through the use of female-specific derogatory phrases such as "bitch" and "witch."

One potential explanation for students recalling men and women teachers differently is that there might actually be significant gender differences in teaching styles. Could it be that women teachers are just more nurturing and men teachers are just more entertaining? Certainly many people believe that men and women have innately different personality traits.[3] However, Sprague and Massoni (2005) find both poles of a dimension represented within each gender, suggesting that this is unlikely. According to the students, the best women teachers were caring and nurturing; the worst-ever women were mean—that is, not nurturing. The best men teachers were intelligent, funny, and energetic; the worst-ever men were boring—that is, not amusing. Sprague and Massoni argue that the polar relationships between the terms used to talk about the best and worst of a gender are evidence that students are, at least to some extent, using gender-specific expectations in evaluating their teachers.

Students' gendered expectations can have consequences for faculty. When teachers do not live up to the gendered standards their students expect and value, and when race/ethnicity and gender intersect in ways

with which students are uncomfortable, faculty seem to receive negative evaluations. When women teachers do gender in a way that students disapprove, they are most at risk for angry, punitive evaluations by their students (Sandler 1991; Sprague and Massoni 2005). Furthermore, anecdotal evidence suggests women teachers are more likely than men to receive comments that refer to their bodies, most often their physical appearance, and that objectify them sexually. Additionally, White students may expect professors of color (especially immigrants) to behave according to preconceived White American constructions of their racial or ethnic group as well as view any of their criticisms of U.S. culture as "bashing" (Vargas 2002). Messner notes, "Students tend to judge women professors first by their 'gender performance' and second by their teaching performance" (2000, 458).

Evidence also suggests the content of a course may further exacerbate this inequity. For example, women faculty are *assumed* to be biased, angry, and less objective than men faculty when teaching women's studies and gender courses (Messner 2000; Rakow 1991). When women of color faculty address racism in the classroom, students may feel they are being personally attacked and assess the professor as unfair and incompetent (Chavez 2002). As Messner, a heterosexual White man, observes, "I appear to be able to get away with making more overtly anti-patriarchal statements in the classroom without receiving a negative judgment from students as to my 'bias' or 'anti-male' attitudes" (2000, 459).

Studies that show teachers being held to gendered standards are consistent with our reading of the implications of the literatures on sociology of gender, social cognition, and the student evaluation of teaching. Together they raise concern that underlying apparently equivalent evaluation procedures there is covert gender bias that has negative implications for women professors seeking tenure and promotion.

Recommendations for Minimizing the Effects of Gender on Student Evaluations of Teaching

Given the evidence suggesting gender *does* influence students' evaluations of their teachers, and given the weight accorded these evaluations by academic institutions, how can women faculty protect themselves in the tenure and promotion process (Ditts, Haber, and Bialik 1994; Seldin 1999)? How can administrators and universities mitigate this impact? There are both structural and individual actions that can be taken to minimize the effects of gender on student evaluations of teaching, to raise awareness of this issue, and to ensure greater accuracy.

Campuses and administrators should acknowledge the effect of gender on student evaluation of teachers as a legitimate issue, work to raise

awareness, and take action to remedy the situation. One method to increase awareness of these issues is to develop workshops (actively supported by the administration) that include activities designed to promote consciousness of how gender influences behaviors and interaction both inside and outside the classroom, highlighting connections to student evaluation of teaching. For instance, women teachers spend more time in after-class/office contact with their students than do men teachers (Bennett 1982). Within the classroom, both men and women students are more likely to exhibit dominant behavior and to interrupt each other in a classroom run by a woman teacher (Brooks 1982; Siskind and Kearns 1997). Additionally, women and men teachers may have different teaching styles. For example, Centra and Gaubatz (2000) found that men college instructors reported lecturing twice as often as their women peers, while women reported using significantly more discussion. Students may inaccurately interpret these gendered differences as women's lack of knowledge or preparedness and men's vast knowledge and preparation (the "sage on the stage") and rate their instructors accordingly. These gendered differences have potential to impact women faculty negatively.

A second way to increase gender awareness is for individual instructors to raise concerns about gender and evaluation in the classroom, encouraging students to think critically about these practices and to consider how they may be evaluating faculty unfairly due to preconceived ideas about members of certain social categories (Messner 2000). Third, campuses also should take steps to reduce the impact of gender and promote more accurate student evaluations of teaching. Universities should at the very least offer, and better yet require, training in the interpretation of teaching evaluations. We urge those involved in evaluating teachers and teaching to find ways of assessing teaching effectiveness that are focused on the goals and outcomes of the course and not vulnerable to students' gendered expectations of the teacher. Because teaching portfolios, peer evaluations, and outcome-based measures are more reliable and valid measures of teaching effectiveness than student ratings, they should be weighed heavily by administrators and evaluation committees.

We echo Aleamoni's (1999) call that academics cease using global measures of the form "overall, s/he is an effective teacher," given the existing evidence that such measures invite (and obscure) perceptual bias. These global measures can be replaced with items that measure specific practices related to teaching effectiveness that students would be more competent to assess. Teaching evaluation procedures can be redesigned to include qualitative measures (e.g., ask students to provide examples that illustrate teaching strengths and weaknesses) that may reveal nuances that cannot be tapped by quantitative indicators that assume raters are using the same criteria and metric across genders and races (Bachen, McLoughlin, and Garcia 1999; Baker and Copp 1997; Messner 2000; Siskind and Kearns 1997; Sprague and Massoni 2005).

Additionally, any summary of student evaluations of teachers should include a statement that acknowledges the potential impact of gender and race on the evaluation. For example, the University of Mississippi includes a disclaimer with all evaluation results, acknowledging that gender does influence evaluations. Further, campuses should institute a policy providing for the exclusion of student evaluations of teachers that are physically or sexually explicit, offensive, abusive, or unrelated to teaching (the University of Colorado has recently adopted this recommendation).

At an individual level, instructors can contribute to the accuracy of student evaluations of teaching by building their own case for teaching effectiveness. Collecting additional measures of teaching success—even when they are not required for salary, retention, promotion, or tenure—can be a useful strategy. One approach is to assemble a teaching portfolio that contains multiple measures of teaching effectiveness like a teaching philosophy, innovative assignments and activities, unsolicited student comments, examples of student work (including instructor comments and suggestions), and the like. Materials on gender and evaluation, such as this article, can be included in the portfolio, and a copy can be provided to the department chair and dean.

While campuses, administrators, and instructors can do much to increase awareness of the pitfalls of current standards of student evaluation of teaching and the possible inaccuracy of these evaluations, researchers also can improve the situation by giving careful thought to the politics of the contemporary college classroom as they design their studies and interpret their data. Specifically, researchers need to look at variances in ratings as well as means. A few hostile students can distort mean ratings, rendering the results inaccurate and potentially harmful to instructors and these instructors are more likely to be women. As noted above, qualitative measures are likely to reveal findings that quantitative measures alone do not or cannot. Thus, researchers should supplement quantitative data with qualitative information—or use primarily qualitative measures. Meta-analyses must be conducted carefully, with particular attention paid to variations in the kinds of data sources and methodologies being combined. Gender differences in ratings can disappear within individual studies when data are collapsed into large categories or factors (Basow 2000), or when individual studies are collapsed for meta-analysis (Miller and Chamberlin 2000; Sprague and Massoni 2005), again rendering the findings inaccurate.

Finally, researchers should collect data on the race/ethnicity, sexual orientation, and/or social class backgrounds of teachers and/or students to provide a context for the interpretation of their findings. In a review of the literature on the scholarship of teaching and learning, Betsy Lucal and her colleagues note the dearth of information about the influence of race/ethnicity, social class, and sexual orientation on teaching assessments (2003). Additional research on the interaction of these factors is needed.

The handful of research studies that take these factors into consideration are most often based on small sample sizes of students and/or anecdotal experiences of individual teachers.[4] However, this anecdotal evidence *does* suggest that these other factors, and race/ethnicity in particular, very likely *do* impact students' evaluations of their teachers, but it is impossible to identify broad patterns of findings at this time. Nonetheless, just as race and gender intersect in shaping most aspects of our lives, we suspect that race and other identity categories are likely to influence the ways in which students evaluate their teachers' performances of gender.

Administrators' increasing reliance on student ratings of teachers in making personnel decisions has generated concern about the fairness and the wisdom of this practice on the part of teachers and those who advocate for them (Trout 2000). Our research not only contributes to this concern but also offers at least a partial corrective to the "common wisdom" of a singular quantitative assessment strategy. If teachers are being held accountable to gendered standards and responding to the standards to which they are being held accountable, women and men may be exerting very different levels of effort to achieve comparable results. If it takes more for a woman to get a "five" and she works above and beyond to do it, we won't see that difference in effort by applying these traditional scales. Sprague and Massoni (2005) call this the "Ginger Rogers effect," borrowing from the observation of Ann Richards, the former governor of Texas, that in the classic dance team of 1930s movies, Ginger Rogers had to do everything her dancing partner, Fred Astaire, did but she had to do it backwards and in high heels. They note that even though Ginger was exerting more effort, Fred still got most of the credit for the performance.

As members of Sociologists for Women in Society (SWS), "an international organization of social scientists—students, faculty, practitioners, and researchers—working together to improve the position of women within sociology and society in general," we work to promote and disseminate research about women (SWS 2007a). Accordingly, we have summarized the research on gender and the evaluation of teaching and posted our recommendations on the SWS website. Further, based on this research, SWS has adopted the following position statement: *The Sociologists for Women in Society recognizes that gender affects students' evaluations of teaching and urges universities and colleges to take steps to minimize the potential negative impact this may have on instructors* (SWS 2007b).

We encourage other academic organizations to begin discussing this issue and to consider adopting similar positions. We hope this research and set of recommendations will be useful to women faculty, and we welcome further input and suggestions from scholars beyond the field of sociology. While progress for women has been made across institutions of higher education, much work remains to be done. Researchers and administrators working together can and should set as a goal the development

of approaches to the evaluation of teaching in which effectiveness in promoting students' learning is the only determinant of outcomes.

Heather Laube *is Assistant Professor of Sociology and Women's and Gender Studies at the University of Michigan–Flint. She studies feminist academics and works on issues of social justice and civic engagement. Send correspondence to Department of Sociology/Anthropology/Criminal Justice, The University of Michigan–Flint, 303 E Kearsley St., Flint, MI 48502-1950; hlaube@umflint.edu.*

Kelley Massoni *is a doctoral candidate in the Department of Sociology at the University of Kansas. She is currently writing her dissertation, an historical examination of the origins of* Seventeen *magazine. Send correspondence to Department of Sociology, University of Kansas, 716 Fraser Hall, Lawrence, KS 66045; massoni@ku.edu.*

Joey Sprague *is Professor of Sociology at the University of Kansas, specializing in the social construction of knowledge. She is author of* Feminist Methodologies for Critical Researchers: Bridging Differences *(2005) and is currently studying the academy as a gendered institution. Send correspondence* to *Department of Sociology, University of Kansas, 716 Fraser Hall, Lawrence, KS 66045; jsprague@ku.edu.*

Abby L. Ferber *is Associate Professor of Sociology and Director of Women's Studies at the University of Colorado at Colorado Springs. She is the author of* White Man Falling: Race, Gender and White Supremacy *(1998); editor of Home Grown Hate: Gender and Organized Racism* (2005); *co-author of* Making a Difference: University Students of Color Speak Out *(2002); and edited an anthology with Michael Kimmel, Privilege: A Reader (2003). Send correspondence to Department of Sociology, University of Colorado at Colorado Springs, 1420 Austin Bluffs Parkway, Columbine Hall, Colorado Springs, CO 80933-7150; aferber@uccs.edu.*

Notes

1. While we are aware of the debates around the concept of "androgyny," much of the research that we review uses this terminology and so we use it to describe their studies accurately.

2. While research on women faculty of color is limited and mostly narrative in form, given what other research says about raced and gendered expectations, we feel safe in employing this work as examples of disparate treatment.

3. In a Gallup Poll (21 February 2001) roughly contemporaneous with this data, interviewers listed personality characteristics and asked respondents to indicate whether they were more likely to characterize men or women. Gallup respondents were far more likely to describe women than men as "emotional," "affectionate," "talkative," "patient," and "creative." They were also far more likely to describe men than women as "aggressive" and "courageous." Women and men were equally likely to be described as "intelligent."

4. For discussion of the impact of race on teacher evaluation, see Galguera 1998; Hendrix 1998; Vargas 2002; Messner 2000; for studies on sexual orientation and teacher evaluation see Ewing, Stukas, and Sheehan 2003; Liddle 1997; Russ, Simonds, and Hunt 2002; for studies on age and teacher evaluation, see Arbuckle and Williams 2003.

References

Aleamoni, Lawrence. 1999. "Student Rating Myths Versus Research Facts from 1924 to 1998." *Journal of Personnel Evaluation in Education* 13(2):153–66.

Arbuckle, Julianne, and Benne Williams. 2003. "Students' Perceptions of Expressiveness: Age and Gender Effects on Teacher Evaluations." *Sex Roles* 49(9/10): 507–16.

Bachen, Christine, Moira McLoughlin, and Sara Garcia. 1999. "Assessing the Role of Gender in College Students' Evaluations of Faculty." *Communication Education* 48:193–210.

Baker, Phyllis, and Martha Copp. 1997. "Gender Matters Most: The Interaction of Gendered Expectations, Feminist Course Content, and Pregnancy in Student Course Evaluations." *Teaching Sociology* 25(1):29–43.

Basow, Susan A. 2000. "Best and Worst Professors: Gender Patterns in Students' Choices." *Sex Roles* 43(5/6):401–17.

———. 1995. "Student Evaluations of College Professors: When Gender Matters." *Journal of Educational Psychology* 87:656–65.

Basow, Susan A., and Nancy T. Silberg. 1987. "Student Evaluations of College Professors: Are Female and Male Professors Rated Differently?" *Journal of Educational Psychology* 79:308–14.

Bennett, Sheila K. 1982. "Student Perceptions of and Expectations for Male and Female Instructors: Evidence Relating to the Question of Gender Bias in Teaching Evaluation." *Journal of Educational Psychology* 74(2):170–9.

Biernat, Monica. 1995. "The Shifting Standards Model: Implications of Stereotype Accuracy for Social Judgment." In *Stereotypes: Perspectives on Accuracy and Inaccuracy,* eds. Yueh-Ting Lee, Lee J. Jussim, and Clark R. McCauley, 87–114. Washington, D.C.: APA.

Biernat, Monica, and Diane Kobrynowicz. 1997. "Gender- and Race-based Standards of Competency: Lower Minimum Standards but Higher Ability Standards for Devalued Groups." *Journal of Personality and Social Psychology* 72:544–57.

Biernat, Monica, and Melvin Manis. 1994. "Shifting Standards and Stereotype-based Judgments." *Journal of Personality and Social Psychology* 66:5–20.

Biernat, Monica, Melvin Manis, and Thomas E. Nelson. 1991. "Stereotypes and Standards of Judgment." *Journal of Personality and Social Psychology* 60: 485–99.

Bray, James, and George Howard. 1980. "Interaction of Teacher and Student Sex and Sex Role Orientations and Student Evaluations of College Instruction." *Contemporary Educational Psychology* 5:241–8.

Brooks, Virginia. 1982. "Sex Differences in Student Dominance Behavior in Female and Male Professors' Classrooms." *Sex Roles* 8(7):683–90.

Brown, Kimberly Nichele. 2002. "Useful Anger: Confrontation and Challenge in the Teaching of Gender, Race, and Violence. In *Women Faculty of Color in the White Classroom*, ed. Lucila Vargas, 89–108. New York: Peter Lang.

Burns-Glover, Alyson, and Dale Veith. 1995. "Revisiting Gender and Teaching Evaluations: Sex *Still* Makes a Difference." *Journal of Social Behavior and Personality* 10(6):69–80.

Carli, Linda L. 2001. "Gender and Social Influence." *Journal of Social Issues* 57(4):725–41.

———. 1990. "Gender, Language, and Influence." *Journal of Personality and Social Psychology* 59:941–51.

Cashin, William E. 1999. "Student Ratings of Teaching: Uses and Misuses." In *Changing Practices in Evaluating Teaching: A Practical Guide to Improved Faculty Performance and Promotion/Tenure Decisions,* ed. Peter Seldin, 25–44. Boston: Anker Publishing.

Centra, John, and Noreen Gaubatz. 2000. "Is There Gender Bias in Student Evaluations of Teaching?" *Journal of Higher Education* 71:17–33.

Chafetz, Janet S., ed. 1999. *Handbook of the Sociology of Gender.* New York: Plenum.

Chavez, Lisa D. 2002. "Reading the Body Indian: A Chicana Mestiza's Experience Teaching Literature. In *Women Faculty of Color in the White Classroom*, ed. Lucila Vargas, 73–88. New York: Peter Lang.

Connell, Robert W. 1995. *Masculinities*. Berkeley: University of California Press.

Cook, Sarah Gibbard. 2004. "Mothers in the Faculty Pipeline." *Women in Higher Education*. August 13(8):16.

CreativeFolk.com. 2007. "Women's Studies Programs Mailing List." Retrieved from http://creativefolk.com/ws.html.

Ditts, David A., Lawrence J. Haber, and Donna Bialik. 1994. *Assessing What Professors Do*. Westport, CT: Greenwood.

England, Paula, Melissa S. Herbert, Barbara Stanek Kilbourne, Lori L. Reid, and Lori McCreary Megdal. 1994. "The Gendered Valuation of Occupations and Skills: Earnings in the 1980 Census Occupations." *Social Forces* 73(1):65–99.

Essed, Philomena. 2000. "Dilemmas in Leadership: Women of Colour in the Academy." *Ethnic and Racial Studies* 23(5):888–904.

Ewing, Vanessa L., Arthur A. Stukas, Jr., and Eugene P. Sheehan. 2003. "Student Prejudice against Gay Male and Lesbian Lecturers." *The Journal of Social Psychology* 143(5):569–80.

Feldman, Kenneth. 1993. "College Students' Views of Male and Female College Teachers: Part II—Evidence from Students' Evaluations of their Classroom Teachers." *Research in Higher Education* 34:151–211.

———. 1992. "College Students' Views of Male and Female College Teachers: Part I–Evidence from the Social Laboratory and Experiments." *Research in Higher Education* 33:317–73.

Fernandez, Juan, and Miguel Mateo. 1997. "Student and Faculty Gender in Ratings of University Teaching Quality." *Sex Roles* 37(11/12):997–1003.

Ferree, Myra Marx, Judith Lorber, and Beth B. Hess, eds. 1999. *Revisioning Gender*. Walnut Creek, CA: AltaMira Press.

Ferree, Myra Marx, and Julia McQuillan. 1998. "Methodological And Policy Issues In University Salary Studies." *Gender & Society 12*(1):7–39.

Freeman, Harvey R. 1994. "Student Evaluations of College Instructors: Effects of Type of Course Taught, Instructor Gender and Gender Role, and Student Gender." *Journal of Educational Psychology* 86(4):627–30.

Galguera, Tomas. 1998. "Students' Attitudes towards Teachers' Ethnicity, Bilinguality, and Gender." *Hispanic Journal of Behavioral Sciences* 20(4):411–29.

Giegerich, Steve. 2004. February 17. "Women Represent Majority of College Students, Minority of Tenured Faculty and Administrators." *The Associated Press*, 17 February:1.

Hendrix, Katherine Grace. 1998. "Student Perception of the Influence of Race on Professor Credibility." *Journal of Black Studies* 28(6):738–63.

Higher Publications, Inc. 1997. *Higher Education Directory*. Falls Church, VA: Higher Education Publications, Inc.

Kierstead, Diane, Patti D'Agostino, and Heidi Dill. 1988. "Sex Role Stereotyping of College Professors: Bias in Students' Ratings of Instructors." *Journal of Educational Psychology* 80(3):342–4.

Kobrynowicz, Diane, and Monica Biernat. 1998. "Considering Correctness, Contrast, and Categorization in Stereotyping Phenomena." In *Stereotype Activation and Inhibition: Advances in Social Cognition, Vol. 11*, ed. Robert S. Wyer, 109–26. Florence, KY: Lawrence Erlbaum Associates.

———. 1997. "Decoding Subjective Evaluations: How Stereotypes Provide Shifting Standards." *Journal of Experimental Social Psychology* 33:579–601.

Liddle, Becky J. 1997. "Coming Out in Class: Disclosure of Sexual Orientation and Teaching Evaluations." *Teaching of Psychology* 24(1):32–5.

Lucal, Betsy, and Cheryl Albers, with Jeanne Ballantine, Jodi Burmeister-May, Jeffrey Chin, Sharon Dettmer, and Sharon Larson. 2003. "Faculty Assessment and the Scholarship of Teaching and Learning: Knowledge Available/Knowledge Needed." *Teaching Sociology* 31(2):146–61.

Luthra, Rashmi. 2002. Negotiating the Minefield: Practicing Transformative Pedagogy as a Teacher of Color in a Classroom of Suspicion. In *Women Faculty of Color in the White Classroom*, ed. Lucila Vargas, 109–24. New York: Peter Lang.

Martin, Elaine. 1984. "Power and Authority in the Classroom: Sexist Stereotypes in Teaching Evaluations." *Signs: Journal of Women in Culture and Society* 9(3):482–92.

Martin, Patricia Yancy. 1996. "Gendering and Evaluating Dynamics: Men, Masculinities, and Managements." In *Men as Managers, Managers As Men,* eds. David Collinson and Jeff Hearn, 186–209. Thousand Oaks, CA: Sage.

Messner, Michael. 2000. "White Guy Habitus in the Classroom: Challenging the Reproduction of Privilege." *Men and Masculinities* 2(4):457–69.

Miller, JoAnn, and Marilyn Chamberlin. 2000. "Women Are Teachers, Men Are Professors: A Study of Student Perceptions." *Teaching Sociology* 28:283–98.
Rakow, Lana. F. 1991. "Gender And Race in the Classroom: Teaching Way Out of Line." *Feminist Teacher* 6(1):10–3.
Ridgeway, Cecilia L. 1987. "Nonverbal Behavior, Dominance, and the Basis of Status in Task Groups." *American Sociological Review* 52:683–94.
Russ, Travis L., Cheri J. Simonds, and Stephen K. Hunt. 2002. "Coming Out in the Classroom . . . an Occupational Hazard? The Influence of Sexual Orientation on Teacher Credibility and Perceived Student Learning." *Communication Education* 51(3):311–25.
Sandler, Bernice R. 1991. "Women Faculty at Work in the Classroom." *Communication Education* 40(1):6–15.
Seldin, Peter. 1999. "Current Practices—Good and Bad—Nationally." In *Changing Practices in Evaluating Teaching: A Practical Guide to Improved Faculty Performance and Promotion/Tenure Decisions*, ed, Peter Seldin, 1–24. Boston, MA: Anker Publishing.
Sinclair, Lisa, and Ziva Kunda. 2000. "Motivated Stereotyping of Women: She's Fine If She Praised Me, But Incompetent If She Criticized Me." *Personality and Social Psychology Bulletin* 26(11):1329–42.
Siskind, Theresa, and Sharon Kearns. 1997. "Gender Bias in the Evaluation of Female Faculty at the Citadel: A Qualitative Analysis." *Sex Roles* 37(7/8): 495–525.
Sociologists for Women in Society (SWS). 2007a. "Homepage." Retrieved from http://www.socwomen.org.
———. 2007b. "SWS Position Statement on Gender and the Evaluation of Teaching." Retrieved from http://www.socwomen.org/page.php?ssss=722.
Sprague, Joey, and Kelley Massoni. 2005. "Student Evaluations and Gendered Expectations: What We Can't Count Can Hurt Us." *Sex Roles* 53(11/12):779–93.
Steinberg, Ronnie, and Louise Haignere. 1987. "Equitable Compensation: Methodological Criteria for Comparable Worth." In *Ingredients for Women's Employment Policy,* eds. Christine Bose and Glenna Spitze. Albany: State University of New York Press.
Stimpson, Catherine. 1982. "Our Search and Research: The Study of Women Since 1969." *Comment* 14 (May).
Tatro, Clayton. 1995. "Gender Effects on Student Evaluations of Faculty." *Journal of Research and Development in Education* 28(3):169–73.
Trout, Paul. 2000. "Flunking the Test: The Dismal Record of Student Evaluations." *Academe* (July/August):58–61.
TuSmith, Bonnie. 2002. "Out on a Limb." In *Race in the College Classroom*, eds. Bonnie TuSmith and Maureen T. Reddy. New Brunswick, NJ: Rutgers University Press.
Valian, Virginia. 2005. "Equity Resources." *Gender Equity Project*. Retrieved from http://www.hunter.cuny.edu/genderequity/equitymaterials.html.
Vargas, Lucila. 2002. "My Classroom in Its Context: The Struggle for Multiculturalism." In *Women Faculty of Color in the White Classroom*, ed. Lucila Vargas, 23–34. New York: Peter Lang.
West, Candace and Sarah Fenstermaker. 1995. "Doing Difference." *Gender and Society* 9(1):8–37.

West, Candace, and Don Zimmerman. 1987. "Doing Gender." *Gender & Society* 1(2):125–51.

Wheeless, Virginia, and Paul Potorti. 1989. "Student Assessment of Teacher Masculinity and Femininity: A Test of the Sex Role Congruency Hypothesis on Student Attitudes Toward Learning." *Journal of Educational Psychology* 81(2):259–62.