Psychology of Women Quarterly, **30** (2006), 25–35. Blackwell Publishing, Inc. Printed in the USA.

GENDER PATTERNS IN COLLEGE STUDENTS' CHOICES OF THEIR BEST AND WORST PROFESSORS

Susan A. Basow, Julie E. Phelan, and Laura Capotosto
Lafayette College

Students ($N = 175$) at a large public university described their best and worst professors. Male students chose a female professor as "best" less often than expected whereas female students chose a female professor as often as expected. There were no gender differences in choice of "worst" professors. Descriptions of best and worst professors varied by professor and student gender. Overall, the male student–female professor pairing seemed most affected by gender dynamics, with male students most likely to emphasize their best female professors' interpersonal skills, especially being approachable. Male students also were most likely to describe their worst female professor in terms of poor classroom interactions, especially closed-mindedness. Although the most frequently used descriptive qualities of best and worst professors showed similarities to those found in previous research, the hidden gendered patterns raise troubling questions for the use of student nominations and comments to choose superior teachers.

How to be an effective college teacher is the topic of numerous books, articles, and research studies. By studying award-winning professors (e.g., Baiocco & DeWaters, 1998; Beidler, 1997; Lough, 1997; Lowman, 1984) and students' descriptions of superior teachers (Buskist, Sikorski, Buckley, & Saville, 2002; Feldman, 1976, 1988; Halonen, 2002; Mowrer, Love, & Orem, 2004; Schaeffer, Epting, Zinn, & Buskist, 2003; Sheehan & DuPrey, 1999), writers have concluded that what distinguishes the best teachers are such qualities as enthusiasm, stimulation of interest, knowledge, dedication, caring, availability, fairness, interactivity, openness, clarity, and organization.

Superior teachers tend to combine both active-instrumental and expressive-nurturant traits (Basow, 2000; Freeman, 1994). Although these traits align with cultural stereotypes of masculinity and femininity, respectively (Bem, 1981), rarely do gender issues appear in the literature on effective teaching. Little is written regarding whether different qualities are important for female and male instructors, whether male and female students prefer different styles or qualities, or whether the gender of the teacher–student pairing affects the teaching dynamic. Social psychology research, however, abounds with evidence that gender matters in ratings of and reactions to professional women compared to professional men (Biernat, 2003; Carli, 2001; Eagly & Karau, 2002; Foschi, 2000). A double standard seems to exist whereby it is harder to demonstrate excellence or competence for lower status individuals (e.g., women); it is also harder to demonstrate incompetence for higher status individuals (e.g., men; Foschi, 2000). From this perspective, it might be more difficult for female professors than male professors to be considered excellent or outstanding. A similar prediction could be based on role congruity theory (Eagly & Karau, 2002), which posits that people show prejudice toward individuals (e.g., women) whose stereotypic characteristics (e.g., nurturant) do not match those of the social roles they inhabit (such as a competent and knowledgeable professional). Furthermore, people tend to use different standards in judging men and women (Biernat, 2003); thus, the qualities that are viewed as notable or superior for male and female professors may vary. Finally, male raters appear to be particularly likely to use gender stereotypes, especially when rating female professionals (Carli, 2001; Eagly & Karau, 2002). Consequently, evaluations by male students of female professors may be most likely to show the effects of gender stereotypes.

Susan A. Basow, Julie E. Phelan, and Laura Capotosto, Department of Psychology, Lafayette College.

The research on which this article is based was conducted while the first author was on sabbatical leave at the University of California, Santa Cruz. Much appreciation goes to the faculty there, especially Faye Crosby, for their assistance with this project. We thank student experimenters Rachael Lando, Kiran Mehta, and Lisa Ravelli for their hard work collecting the data. Thanks also go to several Lafayette College students for their coding and statistical assistance, especially Courtney DeThomas and Karen Mandel. Melissa Scerbo also assisted in the coding.

Address correspondence and reprint requests to: Susan A. Basow, Department of Psychology, Lafayette College, Easton, PA 18042. E-mail: basows@lafayette.edu

In contrast to the literature on superior teaching, the literature on student ratings of professors suggests that gender issues may play a role, although these findings tend to be complex, inconsistent, and small in size (Centra & Gaubatz, 2000; Feldman, 1993). Nonetheless, common findings are a gendered reaction to female professors such that female students often rate them higher and/or male students rate them lower as compared to ratings of male professors, especially when rating overall effectiveness (Basow, 1995; Basow & Silberg, 1987; Centra & Gaubatz, 2000; Feldman, 1993). In contrast, female professors generally receive higher ratings from both male and female students on questions relating to faculty–student interactions and quality (Bachen, McLoughlin, & Garcia, 1999; Basow & Montgomery, in press; Bennett, 1982). Other research suggests female and male students may prefer different teaching styles or teacher qualities (Belenky, Clinchy, Goldberger, & Tarule, 1986; Centra & Gaubatz, 2000; Donaldson, Flannery, & Ross-Gordon, 1993). In particular, female students may prefer "connected" interactive classrooms, a style that may be more typical of female professors (Statham, Richardson, & Cook, 1991). More research thus is needed to determine whether the qualities characterizing superior teaching vary by professor and student gender.

In one of the few studies examining gender aspects of best and worst professors, Basow (2000) found that female professors were chosen as best disproportionately more than expected by female students and disproportionately less than expected by male students. These results suggest that although reactions to male professors are similar from both male and female students, female professors are marked for gender, evoking a bifurcated response from their students: appreciation by female students and lack of appreciation by male students. However, no actual denigration was observed; there were no gender differences in choice of worst professors. The qualities used to describe best professors emphasized effective instructor–student and instructor–class interactions; however, they varied somewhat by both faculty and student gender. Although best professors were most often described as caring and knowledgeable, best female professors were more likely to be described as helpful (especially by female students) than their male counterparts, and best male professors were more likely to be described as interesting and organized (especially by female students) than their female counterparts. Female students were more likely than male students to praise best professors for their knowledge, whereas male students were more likely than female students to comment upon professor clarity. Worst professors were most frequently described as disorganized, unclear, and indifferent, with few variations by professor or student gender. The one exception was that more students described negative instructor–group interactions (such as sexism, playing favorites) for worst male professors than for worst female professors.

One limitation of the Basow (2000) study is that the population sampled was from a small private liberal arts college with a strong emphasis on teaching and on faculty–student interactions. This campus climate, where the average class size is 15, may be the reason that best teachers were described mainly in terms of interpersonal qualities both in and out of the classroom, and worst teachers were mainly described in terms of poor interactions with students on a personal level, along with lack of organization and clarity. The school, coed only for the previous 30 years, also has a politically conservative student body, which may contribute to gendered choices of best professors. Because characteristics of superior teaching may vary depending on student qualities and/or type of institution (Donaldson et al., 1993; Mowrer et al., 2004), further examination at a larger university with a more heterogeneous student body, larger class sizes, and a continuous history of coeducation might offer additional information regarding gendered aspects of superior and inferior teaching.

Another limitation of Basow's (2000) study was that the qualitative comments were coded to fit into Leventhal, Perry, and Abrami's (1977) description of five main teaching factors: scholarship (e.g., "discusses points of view other than his or hers"), organization/clarity (e.g., "explains clearly"), instructor–group interaction (e.g., "encourages class discussion"), instructor–individual student interaction (e.g., "has a genuine interest in students"), and dynamism/enthusiasm (e.g., "energetic and interesting presentation style"). Although useful for comparison purposes, the forced coding may have missed students' intuitive categories. Thus a more open-ended content analysis might allow for more subtle comparisons, especially with respect to the four gendered combinations (e.g., male students describing female faculty, etc.).

The current study attempts to overcome these limitations by examining students' choices of best and worst professors at a large public university using the students' own intuitive categories. Based on shifting standards theory—that it is more difficult for lower status individuals (e.g., women) to demonstrate excellence and for higher status individuals (e.g., men) to demonstrate incompetence (Biernat, 2003; Foschi, 2000), it was predicted that students would nominate female professors as best less often than expected, with no gender differences for worst professors. However, this pattern was predicted only for male students because male raters have been previously shown to be more likely than female students to use gender stereotypes when rating female professionals (Carli, 2001; Eagly & Karau, 2002). Based on role congruity theory (Eagly & Karau, 2002), it was expected that the qualities describing best and worst professors would vary by teacher gender; however, previous research with college professors (e.g., Basow, 2000) suggests that this relationship may be qualified by an interaction between student and teacher gender. In addition, students' ratings of the gender-typed characteristics of their best and worst professors was assessed using the Bem

Inventory (Bem, 1981). As previously found (Basow, 2000; Freeman, 1994), best professors were expected to be rated high in both active-instrumental and expressive-nurturant traits, while worst professors were expected to be rated low in both. It was unclear whether differences based on professor gender would appear.

METHOD

Participants

We initially contacted 220 undergraduates (83 males, 137 females) from a wide variety of locations at a public university in California to participate in a study about best and worst professors. Response rates were 80.1% for males and 83.2% for females, yielding a final sample of 64 males and 111 females. (Six participants did not indicate their sex and were eliminated from analyses.) The final sample (71% European American, 14% Asian, 6% Hispanic, and 9% Other/missing) had an age range from 18 to 39 years ($M = 21.1$, $SD = 2.6$). Male students ($M = 21.7$) were significantly older than female students ($M = 20.7$; $F(1, 179) = 7.35$, $p = .007$). Thirty-six percent were seniors, 29% juniors, 22% sophomores, and 13% first-year students. (Older students, who had experience with more professors, were deliberately oversampled.) In terms of majors, 40% were in the social sciences, 29% humanities, 23% natural science and engineering, and 8% undecided/missing. Female participants were slightly less likely than male participants to major in the natural sciences and engineering, but this difference did not reach significance.

Materials

The questionnaire consisted of two open-ended questions: (a) "Think of the best professor you've had in college" and (b) "Describe what made him or her the 'best' in your opinion." The same question was asked on a separate page with the term "worst" substituted for "best." The order of these questions was counterbalanced. At the bottom of each of these pages was a space for students to write in the faculty member's department, age, ethnicity, gender, approximate class size, and school (community college, present university, other college or university). Following each open-ended question was the short form of the Bem Sex-Role Inventory (Bem, 1981), which asked students to rate their professor on 30 traits using a 7-point Likert scale ranging from 1 (*never or almost never true*) to 7 (*always or almost always true*). The scale has two subscales composed of 10 adjectives each, one for stereotypical masculine (active-instrumental) traits and one for stereotypical feminine (expressive-nurturant) traits. The final page asked students for their year in school, gender, major, ethnicity (open-ended), an estimate of the number of faculty they have had so far, the number of those faculty who were women, and the number who were over 50 years old (a masking variable). It took approximately 15 to 20 minutes to complete the questionnaire.

Procedure

Over a 2-month period, three female undergraduate research assistants distributed questionnaires to potential participants in a variety of campus locations where students gathered: library, snack bars, dining halls, and outside classrooms. The majority (82.3% when those who didn't indicate their gender were included) agreed to participate, filling out the form anonymously and returning it to the researcher's envelope of completed questionnaires.

Several coders, blind to hypotheses, coded qualitative comments. Comments first were coded into separate descriptors of best professors and worst professors, staying close to respondents' own words. The two master lists of comments were stripped of gender identifiers and given to two new coders who grouped similar descriptors by main themes using content analysis. Initial inter-rater reliabilities ranged from 42% to 100%. Categories were collapsed and modified until all comments on both lists could be placed into mutually exclusive categories on which two other coders agreed. Final inter-rater reliabilities were 86% to 100%, with discrepancies discussed and resolved by the first author.

Twenty-nine themes were found in descriptions of best professors (see Table 1) and 21 in descriptions of worst professors (see Table 2). Examples of the best professor category "approachable/accessible/personable/helpful" included "approachable in and out of class," "always friendly," "likeable," and "nice," whereas the "impressive breadth of knowledge and/or experience" category included "wise," "super intelligent," and "incredibly well versed in the field." Examples of the worst professor category "poor teaching skills" included "teaching methods were awful," "knows a lot but can't teach worth hell," and "lectures are too fast or so muddled it's hard to understand; " whereas the category "unenthusiastic, uninspiring, boring" included comments such as "made fun material not fun," "mind-numbingly boring," and "put me to sleep."

These themes were further coded by two raters using Leventhal et al.'s (1977) description of five main teaching factors. All categories but two were utilized and inter-rater reliability was 85% to 100%, with discrepancies discussed and resolved between the raters. As indicated in Table 1, categories within factors were as follows: Best Factor 1, Scholarship/Knowledge, is composed of "knowledgeable," "challenging," "commendable course content," and "learned a lot." Best Factor 2, Organization/Clarity, is composed of "clear expectations," "organized," "clear presentation," and "talented lecturer." Best Factor 3, Instructor–Group Interaction, is composed of "open minded," "humble," "encouraging/inspiring," "quality feedback," "interactive," "fair," and "comfortable classroom." Best Factor 4, Instructor–Individual Student Interaction, is composed of "respected students," "approachable/accessible," "interested in student learning," "warm/caring," and "related to students." Best Factor 5, Dynamism/Enthusiasm, is composed of "made learning enjoyable," "stimulated

Table 1

Percentage of Students Who Described Their Best Professor Using 29 Themes Grouped Into Leventhal et al.'s (1977) Teaching Factors

	Total Percentages	
Factors and Descriptor Categories	*Category*	*Factor*
1. Scholarship/Knowledge		15.73
Impressive breadth of knowledge and/or experience	7.94	
Intellectually challenging	3.61	
Commendable course content, materials, and assignments	3.75	
Learned a lot	0.43	
2. Organization/Clarity		14.86
Clear expectations	2.02	
Organized, prepared for lectures	4.18	
Clear, understandable presentation	5.63	
Talented lecturer, good communication skills	3.03	
3. Instructor–Group Interaction		18.04
Open-minded, receptive to others' opinions, flexible	4.04	
Encouraging/Inspiring	4.62	
Encouraged class discussion, utilized an interactive teaching style	2.02	
Made students feel comfortable in classroom	1.59	
Fair	1.01	
Humble	1.30	
Provided high quality feedback to student work and questions	3.46	
4. Instructor–Individual Student Interaction		23.52
Treated students with respect	2.16	
Approachable/accessible/personable/helpful	8.08	
Warm/caring/understanding to students' needs	3.46	
Personal interest in/related to students, established rapport	4.33	
Interested in student learning, progress	5.48	
5. Dynamism/Enthusiasm		25.54
Utilized innovative or varied means of presenting material	6.20	
Made learning enjoyable	2.45	
Stimulated student interest	3.61	
Exhibited passion/enthusiasm/dedication for subject, teaching	7.36	
Assertive/strong	2.31	
Nontraditional style	1.44	
Humorous/lighthearted	2.16	
Physically Attractive	0.58	
Globally Positive	1.73	

student interest," "passionate," "assertive/strong," "humorous/lighthearted," and "nontraditional style." "Physically attractive" and "globally positive" were not included in any of the five teaching factors.

As indicated in Table 2, Worst Factor 1, Scholarship/Knowledge, is composed of "irrelevant/excessive work" and "lacked knowledge." Worst Factor 2, Organization/Clarity, is composed of "unclear," "poor teaching skills," and "disorganized." Worst Factor 3, Instructor–Group

Table 2

Percentage of Students Who Described Their Worst Professor Using 21 Themes Grouped Into Leventhal et al.'s (1977) Teaching Factors

	Total Percentages	
Factors and Descriptor Categories	*Category*	*Factor*
1. Scholarship/Knowledge		8.60
Work was irrelevant, boring, excessive, and/or unintegrated	5.29	
Lacked knowledge and/or experience	3.31	
2. Organization/Clarity		22.98
Unclear goals, assignments, and overall presentation style	3.47	
Poor teaching skills	12.73	
Disorganized, unprepared for lecture, lacked structure	6.78	
3. Instructor–Group Interaction		15.70
Failed to facilitate successful classroom discussion, teaching methods lacked interactive component	3.64	
Closed-minded, not receptive to others' opinions, overly structured	6.94	
Employed favoritism	1.49	
Annoying habits which detracted from overall learning experience	3.64	
4. Instructor–Individual Student Interaction		24.30
Did not care about students	5.62	
Unkind to students (mean, rude)	5.45	
Arrogant and condescending; made students feel incompetent	7.60	
Not helpful	3.47	
Intimidating, unapproachable	2.15	
5. Dynamism/Enthusiasm		22.31
Disliked teaching, uninterested in teaching	2.48	
Poor communication (language barrier/poor vocal qualities/monotone)	5.79	
Unenthusiastic, uninspiring, boring	11.90	
Unassertive	1.65	
Emotionally unstable	0.50	
Complaints about grading/tests	4.79	
Globally negative	1.32	

Interaction, is composed of "poor discussion/interaction," "closed-minded," "employed favoritism," and "annoying habits." Worst Factor 4, Instructor–Individual Student Interaction, is composed of "did not care about students," "rude/mean," "arrogant/condescending," "not helpful," and "unapproachable." Worst Factor 5, Dynamism/Enthusiasm, is composed of "disliked teaching," "poor communication," "unenthusiastic/boring," "unassertive," and "emotionally unstable." "Complaints about grading/tests" and "globally negative" were not included in any of the five teaching factors.

RESULTS

For the sample as a whole, students had taken classes with 4 to 55 professors ($M = 23.5$, $SD = 11.2$), 5% to 100% of whom were female ($M = 39.3\%$, $SD = 0.20$), 0% to 100% of whom were over age 50 ($M = 39.2\%$, $SD = 0.26$), 0% to 100% of whom were not European American ($M = 17.1\%$, $SD = 0.18$). Female students had a significantly higher percentage of courses with female professors than did male students ($M = 42\%$ and 34%, respectively; $F(1,178) = 6.83$, $p = .01$).

Best Professors

Students' choices of their best and worst professors were analyzed first by cross-tabulating participant gender and professor gender.[1] As Table 3 indicates, 71% of the nominated best professors were male, but male students were more likely (84.4%) than female students (63.1%) to nominate a male as best professor. Female students nominated female professors approximately in proportion to the percent of female professors they had had, whereas the percentage of male students who nominated female professors as best was only about half of the percentage of female professors that male students had had, $\chi^2(1) = 8.93$, $p = .003$. Thus, whereas female students chose female professors as best in direct proportion to their availability, male students were less likely to choose female professors and more likely to choose male professors than expected.

Table 3

Percent of Best Female and Male Professors Chosen by Female and Male Students Compared to Expected Value (Based on Actual Percent of Female Professors)

Faculty gender	*Female Students (N = 111)*		*Male Students (N = 64)*	
	Observed	*Expected*	*Observed*	*Expected*
Female[a]	36.9%	41.5%	15.6%	33.9%
Male[b]	63.1%	58.5%	84.4%	66.1%

[a]$N = 51$. [b]$N = 124$.

In terms of gender-typed characteristics, best professors were rated typically as strong in both expressive-nurturant traits ($M = 5.41$, $SD = 1.14$) and active-instrumental traits ($M = 5.10$, $SD = .83$). Gender patterns emerged, however, on ratings of the latter. A 2 × 2 (Participant Gender × Professor Gender) analysis of variance (ANOVA) on ratings of best professor instrumental traits revealed a significant interaction, $F(1, 167) = 6.62$, $p = .011$. Each gender gave highest ratings to same-gender professors, with the most variability occurring in ratings of female professors: Male students rated their best female professors lower ($M = 4.79$, $SD = 0.75$) and female students rated them higher ($M = 5.34$, $SD = 0.79$) on instrumental traits ($p < .05$). In contrast, male students rated their male professors significantly higher ($M = 5.23$, $SD = 0.88$) on these traits than did female students ($M = 4.95$, $SD = 0.79$).

As shown in Table 1, students most often described their best professors as approachable/accessible (8.08%), having impressive breadth of knowledge and/or experience (7.94%) and exhibiting passion/enthusiasm (7.36%). Best female and best male professors were described in significantly different ways, however, $\chi^2(26) = 43.15$, $p < .05$. These ratings further varied significantly as a function of student gender, $\chi^2(60) = 98.95$, $p < .01$.[2] Best female professors were most often described by their students as approachable/accessible (12.09%), followed by passionate (6.98%), and relating to students (6.51%); best male professors were most often described by their students as knowledgeable (8.79%), passionate (7.53%), and innovative (6.9%). Best female professors were significantly more likely to be described as approachable/accessible than their male counterparts, $\chi^2(1) = 6.23$, $p = .01$, but mostly by male students, $\chi^2(3) = 14.24$, $p = .003$. As shown in Figure 1, nearly one-quarter (23.68%) of male students' descriptions of their best female professors noted these qualities whereas

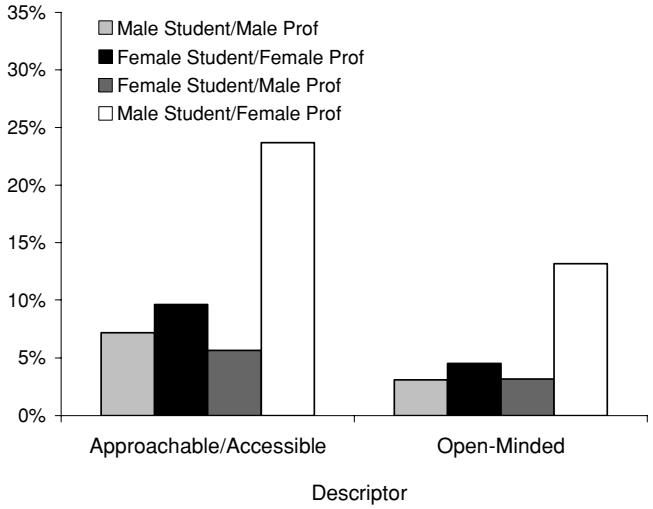

Fig. 1. Frequency of best professor descriptors that varied significantly by student gender and professor gender.

these qualities constituted 9.6% of the comments given by female students. Male professors, in contrast, were significantly less likely to be described in this manner. Best female professors also were more likely than best male professors to be described as open-minded/flexible by male students only (13.16%), $\chi^2(3) = 8.88$, $p = .03$ (see Figure 1).

After consolidating the categories into the five teaching factors, results indicated that best professors were most often described by their students using Factor 5, Dynamism/Enthusiasm (25.54%), and Factor 4, Instructor–Individual Student Interaction (23.52%); the least used factor was Factor 2, Organization/Clarity (14.86%). The top two categories were the same for male and female professors, although the rank order differed, as shown in Figure 2. Students most often described their best female professors using Factor 4, Instructor–Individual Student Interaction (30.23%), followed by Factor 5, Dynamism/Enthusiasm (26.51%), whereas best male professors were most often described by Factor 5, Dynamism/Enthusiasm (25.2%), followed by Factor 4, Instructor–Individual Student Interaction (20.5%). Chi-square analysis of the five teaching factors combined found a significant main effect for professor gender, $\chi^2(4) = 14.71$, $p = .005$, as well as an interaction between student gender and professor gender that approached significance, $\chi^2(12) = 20.72$, $p = .06$. Follow-up analyses revealed that best female professors were significantly more likely to be described by Factor 4, Instructor–Individual Student Interaction (30.23%), $\chi^2(1) = 6.01$, $p = .01$, by both their male and female students, but especially by their male students (39.47%), $\chi^2(3) = 9.17$, $p = .03$. Male professors were significantly less likely to be described this way (20.5%), due primarily to their female students (19.72%), $\chi^2(3) = 9.17$, $p < .05$.

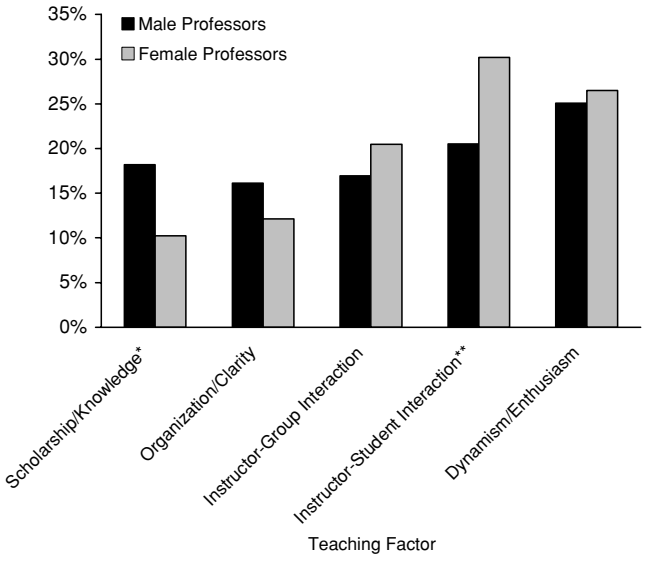

Fig. 2. Frequency of best professor factor descriptors by professor gender.
*$p < .05$. **$p < .01$.

Best male professors were also significantly more likely to be described using Factor 1, Scholarship/Knowledge (18.2%), $\chi^2(1) = 5.96$, $p = .02$, by both their male (19.07%) and female students (17.61%). Female professors were less likely to be described this way (10.23%), especially by their male students (5.26%), $\chi^2(3) = 6.65$, $p = .08$.

Worst Professors

For choice of worst professors, there was no student gender by faculty gender interaction. Twice as many male as female professors were nominated as worst (66.1% compared to 33.9%), a percent that did not vary by student gender. Students made their choice proportional to the number of male and female faculty they estimated they had had.

In terms of gender-typed characteristics, worst professors typically were rated as extremely low in expressive-nurturant traits ($M = 2.81$, $SD = 1.17$) as well as moderately low in active-instrumental traits ($M = 4.12$, $SD = 1.31$). No gender patterns emerged on ratings of instrumental traits, but there was a significant two-way (participant gender x professor gender) interaction on ratings of expressive-nurturant traits, $F(1, 164) = 4.86$, $p = .029$. Male students rated their worst female professors significantly higher ($M = 3.34$, $SD = 1.48$) on these traits than their worst male professors ($M = 2.54$, $SD = 1.01$), whereas female students rated their male and female professors exactly the same ($M = 2.82$, $SD = 1.16$ for both).

As shown in Table 2, worst professors were most frequently described as having poor teaching skills (12.73%) and as being unenthusiastic/boring (11.9%). There were significant differences in descriptors used by male and female students, $\chi^2(19) = 33.17$, $p < .05$, especially in interaction with professor gender, $\chi^2(48) = 94.99$, $p < .01$; however, differences based on professor gender approached significance, $\chi^2(19) = 28.27$, $p < .1$. See Figure 3 for the six most common descriptors by professor and student gender.

As shown in Figure 3, male students were more likely to describe their worst female professors as closed-minded/rigid (25%) than any other student–professor pairing, $\chi^2(3) = 25.34$, $p = .001$. Male students also were significantly more likely, and female students less likely, to describe their worst professors as arrogant/condescending (11.37% compared to 5.58%), $\chi^2(1) = 5.96$, $p < .02$, primarily regarding their male professors (13.77%), $\chi^2(3) = 11.77$, $p = .008$. Finally, male students described a worst female professor as having poor communication more than any other group (15.91%), $\chi^2(3) = 9.41$, $p = .02$.

Overall, female students were most likely to describe their worst professor as having poor teaching skills (13.45%) and as being unenthusiastic/boring (11.68%). Male students also described their worst professor using these descriptors, but in a different rank order (unenthusiastic/boring, 12.32%; poor teaching skills, 11.37%) as well as arrogant and condescending (11.37%).

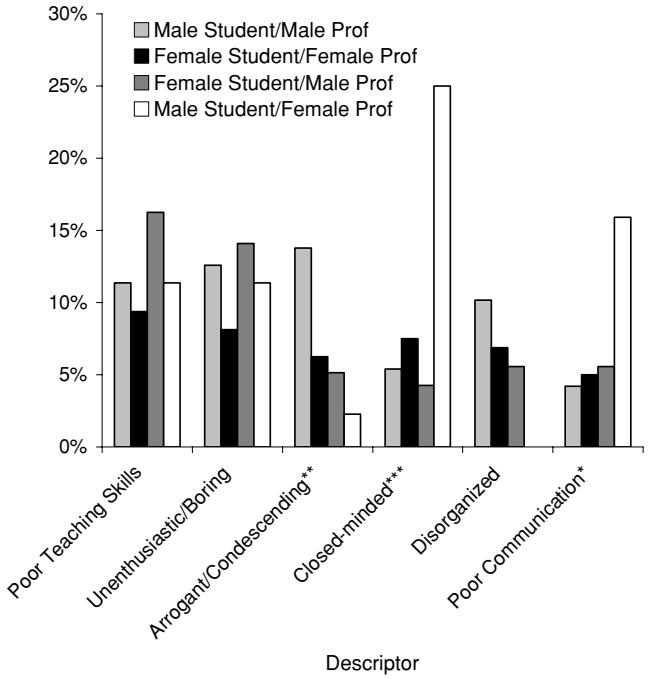

Fig. 3. The six most frequent categories used to describe worst professors as a function of professor gender and student gender. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

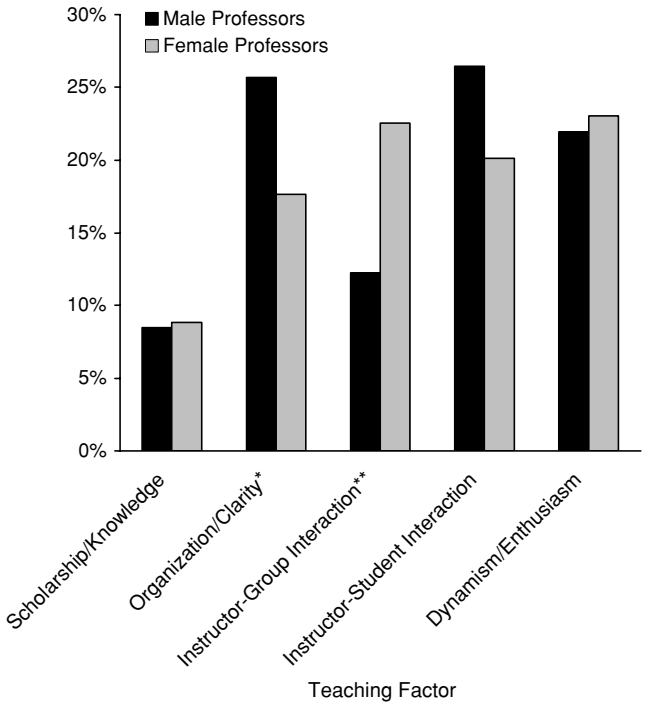

Fig. 4. Frequency of worst professor factors descriptors by professor gender.
$^{*}p < .05$. $^{**}p < .01$.

Examination of the five teaching factors revealed that Factor 4, Instructor–Individual Student Interaction (24.30%); Factor 5, Dynamism/Enthusiasm (22.31%); and Factor 2, Organization/Clarity (22.98%), were the three most important factors in describing worst professors. However, these patterns differed by professor gender, $\chi^2(4) = 15.36$, $p = .004$, especially in interaction with student gender, $\chi^2(12) = 27.67$, $p = .006$ (see Figure 4). Whereas Factor 4, Instructor–Individual Student Interaction (26.43%), and Factor 2, Organization/Clarity (25.69%), were most important for worst male professors, Factor 5, Dynamism/Enthusiasm (23.04%), and Factor 3, Instructor–Group Interaction (22.55%), were the most important for female professors. Worst male professors were described using Factor 3, Instructor–Group Interaction, much less than expected (12.22%), whereas female professors were described on this dimension much more than expected (22.55%), $\chi^2(1) = 8.8$, $p = .003$. This main effect of professor gender was qualified by an interaction with student gender, such that female and male students wrote comments pertaining to Instructor–Group Interaction more for their female professors, 20.0% and 31.82%, respectively, whereas fewer female students wrote comments about Instructor–Group Interaction for their male professors (10.26%). Male students wrote comments relating to Instructor–Group Interaction proportionate to their number of male professors (14.97%), $\chi^2(3) = 14.97$, $p = .002$. In addition, there was a significant main effect of professor gender for Organization/Clarity, with worst male professors receiving this comment proportionately more (25.69%) and female professors receiving this comment proportionately less relative to the number of professors of each gender students had had (17.65%), $\chi^2(1) = 4.06$, $p = .04$.

DISCUSSION

As expected, gender factors played a complex role in students' evaluations of their best and worst professors. As previous research has found, male students may have been particularly affected by gender stereotypes when considering female professors. Similar to male students at the small private liberal arts college Basow (2000) studied, male students in this study also nominated a female professor as best significantly less frequently than would be expected based on the percentage of female professors they had had in class. This finding, combined with other research suggesting that male students may hold more traditional attitudes toward women (Frieze et al., 2003) and rate female faculty lower than female students do (Basow, 1998; Basow & Silberg, 1987), has implications for interpreting student evaluations of professors. Although male students do not nominate female professors as worst more often than expected, they may hold female professors to a higher standard than they do male professors. This would be consistent with Biernat's (2003) shifting standards model of stereotyping and Foschi's (2000) findings that low-status individuals need more confirmatory information to be rated highly than do high-status individuals. The need for more confirmatory

information may be particularly apparent when male students are evaluating female professors because research on implicit prejudice finds that individuals who belong to socially advantaged groups (e.g., men) typically exhibit more bias against outgroups (e.g., women) than do members of socially disadvantaged groups (Dasgupta, 2004).

Female students, on the other hand, chose female professors both as best and as worst in direct proportion to their frequency as classroom professors. Thus, female students did not seem to use gender as a guiding factor in their choice, or they did not consider female professors to be low status. These results are slightly at variance with those of Basow (2000), who found that female students chose female professors as best more frequently than would be expected based on the number of female professors they had had. Perhaps the smaller campus size in the Basow (2000) study made female professors more noticeable as role models for female students (Basow & Howe, 1980) than at the larger public university in the present study.

As previous research has found, the most highly rated professors combined high levels of both active-instrumental and expressive-nurturant traits; that is, they were described as androgynous (Basow, 2000; Freeman, 1994). Conversely, worst professors had low levels of both sets of traits, especially expressive-nurturant ones. Contrary to Basow (2000), however, student gender and professor gender significantly interacted in ratings of active-instrumental traits for best professors and expressive-nurturant traits for worst professors. In both cases, the interaction was driven by male students rating professors in more gender-stereotypic ways than did female students. Male students rated best professors who were female particularly low on active-instrumental traits and rated worst professors who were female particularly high on expressive-nurturant traits. Again, male students may perceive female professors more through the lens of gender stereotypes than do female students, who rated their best female professors higher on active-instrumental traits than their best male professors. As Eagly and Karau (2002) found, role congruity may be particularly important in men's attitudes toward female leaders.

Role congruity theory may also explain the different descriptions given of both best and worst professors as a function of both professor and student gender. Consistent with previous research on effective teaching (Buskist et al., 2002; Donaldson et al., 1993; Feldman, 1976, 1988; Sheehan & DuPrey, 1999), best professors typically were described by such themes as approachable/accessible/personable/helpful, impressive breadth of knowledge and/or experience, and exhibited passion/enthusiasm/dedication. However, previous research did not examine descriptions based on professor or student gender. The current study shows the importance of doing so. Although the approachable/accessible/helpful theme was most often used to describe best professors in general, it was used significantly more for female professors than male professors, mainly by male students. In fact, nearly one out of four male students described their best female professors this way. For example, a male student described his best female professor as "willing to take time from her busy schedule and help us out." Male students were also more likely to describe their best female professors in terms of being open-minded/flexible (e.g., "open to student opinion, but willing to point out how they may be wrong if they are"). Best female professors were more likely than best male professors to be described as relating well to students (e.g., "understood needs of students") and providing a comfortable classroom (e.g., "made class a comfortable space so we could voice our opinions without being judged"). Male professors were more likely than female professors to be described as providing commendable course content (e.g., "worthwhile assignments") and in globally positive terms (e.g., "I love the guy").

Gendered patterns in describing best professors were even more apparent when themes were categorized into the five-factor structure used by Basow (2000). Both male and female students, but especially male students, most often described best female professors using qualities tapping Instructor–Individual Student Interactions. Perhaps because female professors were expected to be strong in interpersonal qualities (role congruity theory; Eagly & Karau, 2000) and indeed did seem to be rated highly on these interactions by everyone in this study and in others (Bachen et al., 1999; Basow, 1995; Basow & Montgomery, in press; Bennett, 1982; Feldman, 1993; Sears & Hennessey, 1996), male students focused on these gender-congruent behaviors in describing their best female professors. Best male professors, in contrast, were more likely than female professors to be described along the Scholarship/Knowledge dimension (more male-stereotypic qualities). These qualities were rarely attributed to female professors, especially by male students. Another reason male professors may be more likely to be described using the Scholarship/Knowledge factor may be that male faculty tend to lecture more than female faculty, who are more likely to use discussion (Statham et al., 1991). Although there was no measure of teaching style in the current study, male and female professors nominated as best did not differ in terms of discipline or size of class.

Male students were also more likely to describe their best female professors in terms of being open-minded and flexible and to criticize their worst female professors for being closed-minded and inflexible (e.g., "not much room for personal interpretation of subject matter"). Perhaps because male students challenge female professors more than male professors (Statham et al., 1991), male students may be especially appreciative when they are allowed to do so.

As the above indicates, descriptions of worst professors varied significantly by faculty and student gender, driven again by the distinctive ratings of male students. Although male and female students both described worst professors as using poor teaching skills, being unenthusiastic or boring, and being disorganized, only male students described their

worst male professors as arrogant or condescending (e.g., "talked down to students") and their worst female professors as closed-minded/inflexible and poor communicators (e.g., "unable to communicate in lecture"). Male students seemed particularly sensitive to language and communication style with both their male and female professors, perhaps as it conveyed something about their own status, which as males they expect to be high. With a male professor, they may have been particularly sensitive to being slighted; with a female professor, they may have been particularly sensitive to not having their voices heard and not liking the professor's voice ("poor communication" encompassed comments about vocal qualities, such as having an accent or talking in a monotone). This interpretation seems consistent with their unique praise of best female professors as open-minded/flexible and approachable/accessible.

For worst teaching factors as well, there were significant variations by faculty and student gender. Worst female professors were primarily described in terms of the dimensions of Dynamism/Enthusiasm and Instructor–Group Interactions, although male students mainly gave the latter ratings. In contrast, worst male professors were primarily described in terms of the dimensions of Organization/Clarity and poor Instructor–Individual Student Interactions. Basow (2000) also found these latter two factors to be the most frequently used to describe worst professors, but in the previous study, this did not vary by professor gender. In the current study, male and female professors were viewed as worst for different reasons, and these reasons were not exactly the opposite of what made teachers best. In particular, organization and clarity did not figure as an important strength in describing best professors, but lack of these skills could cause a professor, especially if male, to be considered worst. Similarly, interacting well with the class as a group did not figure strongly in descriptions of best professors, but lack of these qualities did figure strongly in descriptions of worst female professors. As with descriptions of best professors, gendered teaching styles may be involved (Statham et al., 1991): If male professors are more likely to lecture, lack of organization/clarity will be more deleterious; if female professors are more likely to use discussion, poor group interaction skills may be more salient for them. It is noteworthy that previous research has found that perceived lack of teacher organization and planning are the qualities most irritating to students (Basow, 2000; Perlman & McCann, 1998). Because male faculty tend to be rated higher on these qualities than do female faculty (Bachen et al., 1999; Basow, 1995; Basow & Silberg, 1987), it follows that male professors would be more criticized on this dimension as well. Although teaching styles were not examined, there were no significant differences in terms of discipline or class size for male and female worst professors.

Overall, the dimensions of Dynamism/Enthusiasm and Instructor–Individual Student Interactions appear to be important for choice of best professor because descriptors encompassing these traits were most frequently used in both Basow's (2000) previous research at a small private school and in the current study at a large state school. In addition, poor organization and lack of clarity appears to be a particularly important determinant in choice of worst professor at both institutions. These results vary slightly from those of Basow (2000) who found that in addition to Instructor–Individual Student Interactions, Instructor–Group Interaction was also highly rated for both male and female best professors. Perhaps because the school in the previous study had small classes and emphasized faculty–student interaction, those qualities were most often described. In the current study, perhaps because class sizes tended to be larger,[3] the dimension of Dynamism/Enthusiasm played more of a role and Instructor–Group Interaction less in descriptions of best professors. Large class sizes also may explain why negative descriptions on the dimensions of Dynamism/Enthusiasm and Instructor–Individual Student Interactions figured prominently for worst professors in the present research. These findings highlight the importance of variations in class characteristics in describing what qualities make for perception of teacher excellence (Donaldson et al., 1993; Mowrer et al., 2004).

One of the strengths of the current study is the extension of Basow's (2000) findings from a small private college to a large state research university, with a few variations that probably reflect the particular campus climate, especially class size and level of expected faculty–student interactions. Future research on student evaluations of professors needs to attend to campus factors as well. Future research should also utilize content analysis. Although time-consuming, it is a helpful approach in examining the natural categories that students use in describing their best and worst professors, descriptions that may not have been obtained in a fixed-format questionnaire. By categorizing these student-derived themes into the previously used five-factor structure, findings can be more easily compared.

Another strength of the current study is the use of a wide range of students, especially those with a variety of majors. Previous research examining the qualities of superior teachers has tended to use psychology students as participants (e.g., Halonen, 2002; Mowrer et al., 2004; Sheehan & DuPrey, 1999). Although informative, these findings may be somewhat limited, because in the current study, students disproportionately chose both their best and worst professor from the same division as their major, probably because they had more professors from that division.[4] It remains to be seen whether the qualities that make for superior teaching vary not only by gender but also by class type and/or division (Basow & Montgomery, in press). It is noteworthy that few natural science/engineering professors (18%) were nominated as best whereas 35% were nominated as worst. Conversely, 42% of best professors and only 27% of worst professors were from the humanities. More research on this dimension is needed.

A major limitation of this study is the lack of direct information about the teaching styles employed by male and

female professors and those preferred by male and female students. Although male and female professors chosen as best or worst came from similar disciplines, their teaching styles might have been different. As indicated above, men may employ the lecture format more than women. If this is the case (as suggested by Statham et al., 1991), then the fact that male professors, compared to female professors, are praised more for their dynamism and enthusiasm, and criticized more for their lack of organization and clarity, is understandable. Similarly, if women professors utilize class discussion more than men, then the finding that they are particularly likely to be praised for their interactions with students and criticized for their poor instructor–group interactions makes sense as well. Nevertheless, the significant interactions with student gender and the particular emphasis on the interpersonal qualities of female professors suggest that gender stereotyping, role congruity, and shifting standards are operating in the evaluation of college teaching as they operate in other professional settings.

There may also be gender-influenced stylistic preferences on the part of students with respect to particular teacher behaviors that should be further explored. There is some evidence that faculty and students sometimes cite somewhat different qualities as being most important in effective teaching (Feldman, 1976, 1988; Schaeffer et al., 2003), but little evidence regarding whether preferred qualities vary by professor and student gender pairings. Such preferences may not be conscious, given the implicit mechanisms by which stereotypes operate (Dasgupta, 2004).

A micro-level analysis of specific teacher behaviors and student perceptions and reactions would help to throw light on the degree to which male and female faculty teach differently, and the degree to which students perceive the same teacher behaviors differently depending on the gender of the professor and the student's gender. For example, do male students inaccurately perceive their male professors to be particularly arrogant and condescending, or do male professors actually behave in this manner, perhaps only with male students? A related caution is that students' descriptions of qualities that characterize their best or worst professors may not differentiate between professors who get the highest overall ratings and those who get the lowest (Feldman, 1976, 1988). Instead, students may form a general impression of a professor for reasons of which they are unaware and, when pressed, identify specific teaching-related descriptions to justify their impression. We thus need more detailed research into the perception and rating process involved in student evaluations, especially because there is some evidence that faculty rate themselves differently than students rate them (Basow & Montgomery, in press).

It also would be helpful to have direct measures of students' degree of gender traditionalism and implicit prejudice. Although it is likely that male students' greater use of gender stereotypes in describing their best and worst female professors is due to men's stronger adherence to traditional gender roles (Frieze et al., 2003) or men's socially advantaged ingroup status relative to women professors' outgroup status (Dasgupta, 2004), direct measures of these constructs would help to elucidate these dynamics.

In summary, this study demonstrates the importance of examining gender variables in students' descriptions of professors. Male and female professors, but especially the latter, are viewed against a background of gendered expectations and appear to be judged using gendered standards, as other social psychology researchers have found (Biernat, 2003; Foschi, 2000). Male students are particularly prone to making gendered judgments with respect to female faculty, as they are with respect to other female professionals (Basow & Silberg, 1987; Carli, 2001; Eagly & Karau, 2002). Accordingly, it is imperative that student evaluations of professors, especially recommendations regarding excellence, be evaluated critically to ensure that gender bias is not present. For example, if interpersonal skills such as accessibility and helpfulness are important for high teacher ratings and female professors tend to be rated higher on these qualities than male professors, why are female teachers not *more* likely to be viewed as superior teachers? Teachers also need to be aware that certain teaching styles and qualities may be differentially effective based on their own, as well as their students', gender. The current results suggest that if professors want to receive positive student ratings, it may be particularly important for female professors to appear strong in Instructor–Individual Student Interaction, especially to be viewed as accessible, and for male professors to appear knowledgeable, dynamic, clear, and organized. Thus, although best professors combine both active-instrumental and expressive-nurturant traits, the latter may be more important for female professors (especially with male students), and the former for male professors. Overall, it is important that we use student ratings and recommendations of teachers with greater awareness of the subtle role gender variables play.

Initial submission: August 25, 2004
Initial acceptance: April 24, 2005
Final acceptance: August 12, 2005

NOTES

1. Most of the nominated professors (83.7% of the best and 79.3% of the worst) were from the campus studied, 7.6% of the best and 9.8% of the worst were from community colleges, and the remainder (< 10%) were from other colleges or universities.
2. Note that descriptor items with more than one cell with $N < 5$ were omitted from the total chi-square computation. Complete tables of results are available upon request.
3. The most common class size taught by a best professor was 21 to 45 students (28.6%), followed by 91 to 200 (22.3%), with only 15.4% coming from classes with more than 200 students. The mean class size was 118.2 ($SD = 181.9$). The most common class size taught by a worst professor was 91–200 students (25.7%) followed closely by 21–45 students (24.6%); $M = 110.8$, $SD = 108.6$.

4. Students were most likely to choose as best and as worst a professor in his or her own major area, $\chi^2(4) = 98.53$, $p < .001$ and $\chi^2(4) = 84.41$, $p < .001$, respectively). Best professors were most likely to teach in the humanities (41.8%) and social sciences (40.1%), followed by natural sciences/engineering (18.1%). Worst professors were most likely to teach in the social sciences (37.7%) and natural sciences/engineering (35.1%), followed by the humanities (27.3%).

REFERENCES

Bachen, C. M., McLoughlin, M. M., & Garcia, S. S. (1999). Assessing the role of gender in college students' evaluations of faculty. *Communication Education, 48,* 193–210.

Baiocco, S. A., & DeWaters, J. N. (1998). Studies in excellence. In S. A. Baiocco & J. N. DeWaters, *Successful college teaching* (pp. 53–92). Needham Heights, MA: Allyn & Bacon.

Basow, S. A. (1995). Student evaluations of college professors: When gender matters. *Journal of Educational Psychology, 87,* 656–665.

Basow, S. A. (1998). Student evaluations: The role of gender bias and teaching styles. In L. H. Collins, J. C. Chrisler, & K. Quina (Eds.), *Career strategies for women in academe: Arming Athena* (pp. 135–156). Thousand Oaks, CA: Sage.

Basow, S. A. (2000). Best and worst professors: Gender patterns in students' choices. *Sex Roles, 43,* 407–417.

Basow, S. A., & Howe, K. G. (1980). Role model influence: Effects of sex and sex-role attitudes in college students. *Psychology of Women Quarterly, 4,* 558–572.

Basow, S. A., & Montgomery, K. (in press). Student ratings and professor self-ratings of college teaching: Effects of gender and divisional affiliation. *Journal of Personnel Evaluation in Education*.

Basow, S. A., & Silberg, N. (1987). Student evaluations of college professors: Are female and male professors rated differently? *Journal of Educational Psychology, 79,* 308–314.

Beidler, P. G. (1997). What makes a good teacher? In J. K. Roth (Ed.), *Inspiring teacher: Carnegie professors of the year speak* (pp. 2–12). Bolton, MA: Anker.

Belenky, M. F., Clinchy, B. M., Goldberger, N. R., & Tarule, J. M. (1986). *Women's ways of knowing*. New York: Basic Books.

Bem, S. L. (1981). *Bem Sex-Role Inventory: Professional manual*. Palo Alto, CA: Consulting Psychologists Press.

Bennett, S. 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,* 170–179.

Biernat, M. (2003). Toward a broader view of social stereotyping. *American Psychologist, 58,* 1019–1027.

Buskist, W., Sikorski, J., Buckley, T., & Saville, B. K. (2002). Elements of master teaching. In S. F. Davis & W. Buskist (Eds.), *The teaching of psychology: Essays in honor of Wilbert J. McKeachie and Charles L. Brewer* (pp. 27–39). Mahwah, NJ: Erlbaum.

Carli, L. L. (2001). Gender and social influence. *Journal of Social Issues, 57,* 725–741.

Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *Journal of Higher Education, 71,* 17–33.

Dasgupta, N. (2004). Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research, 17*(2), 143–169.

Donaldson, J. F., Flannery, D., & Ross-Gordon, J. (1993). A triangulated study comparing adult college students' perceptions of effective teaching with those of traditional students. *Continuing Higher Education Review, 57*(3), 147–164.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Bulletin, 108,* 233–256.

Feldman, K. A. (1976). The superior college teacher from the students' view. *Research in Higher Education, 5,* 243–288.

Feldman, K. A. (1988). Effective college teaching from the students' and faculty's view: Matched or mismatched priorities. *Research in Higher Education, 28,* 291–344.

Feldman, K. (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,* 51–211.

Foschi, M. (2000). Double standards for competence: Theory and research. *Annual Review of Sociology, 26,* 21–42.

Freeman, H. 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,* 627–630.

Frieze, I. H., Ferligoj, A., Kogovsek, T., Rener, T., Horvat, J., & Sarlija, N. (2003). Gender-role attitudes in university students in the United States, Slovenia, and Croatia. *Psychology of Women Quarterly, 27*(3), 256–261.

Halonen, J. S. (2002). Classroom presence. In S. F. Davis & W. Buskist (Eds.), *The teaching of psychology: Essays in honor of Wilbert J. McKeachie and Charles L. Brewer* (pp. 41–55). Mahwah, NJ: Erlbaum.

Leventhal, L., Perry, R., & Abrami, P. (1977). Effects of lecturer quality and student perception of lecturer's experience on teacher ratings and student achievement. *Journal of Educational Psychology, 69,* 360–374.

Lough, J. R. (1997). The Carnegie professors of the year: Models of teaching success. In J. K. Roth (Ed.), *Inspiring teaching: Carnegie professors of the year speak* (pp. 212–225). Bolton, MA: Anker.

Lowman, J. (1984). What constitutes masterful teaching. In J. Lowman, *Mastering the techniques of teaching* (pp. 1–22). San Francisco: Jossey-Bass.

Mowrer, R. R., Love, S. S., & Orem, D. B. (2004). Desirable teaching qualities transcend the nature of the student. *Teaching of Psychology, 31,* 106–108.

Perlman, B., & McCann, L. I. (1998). Students' pet peeves about teaching. *Teaching of Psychology, 25,* 201–203.

Schaeffer, G., Epting, K., Zinn, T., & Buskist, W. (2003). Student and faculty perceptions of effective teaching: A successful replication. *Teaching of Psychology, 30,* 133–136.

Sears, S. R., & Hennessey, A. C. (1996). Students' perceived closeness to professors: The effects of school, professor gender, and student gender. *Sex Roles, 35,* 651–658.

Sheehan, E. P., & DuPrey, T. (1999). Student evaluations of university teaching. *Journal of Instructional Psychology, 26,* 188–193.

Statham, A., Richardson, L., & Cook, J. A. (1991). *Gender and university teaching*. Albany, NY: State University of New York Press.