

Summaries of Articles of Research on Biases in Teaching

If you are interested in better understanding the concerns of bias with courses evaluations, we encourage you to review the archive of classic and recent research on this issue below.

[Google Drive Folder of Archive of Research on Biases in Evaluations](#)
[Summary and List of Articles of Research on Biases in Evaluations](#)

You can [submit an article or correction](#) to the list here.

Bibliography of Research on Gender and Racial Bias with Teaching and Evaluations

1. Abel, M. H. and A. L. Meltzer, Students ratings of a male and female professors' lecture on sex discrimination in the workforce, *Sex Roles*, 57, 173-180.
 - a. In an experiment, 41 male and 46 female undergraduate students taking psychology courses at a regional comprehensive southern university read a lecture written by what they perceive to be a male and female teacher that focuses on pay disparities between men and women. Students were then asked to rate the lecture. The male professor received significantly higher overall rating, higher rating on accuracy of the lecture, and a low rating on sexist rating. There was "no significant interaction between sex of student and sex of professor; male and female student ratings were similar for both professors." Students with more liberal attitudes were correlated with lower ratings of sexism for the female professor. Students with more traditional views rated female professors to be more sexist than male professors; no difference between students with liberal attitudes.
2. Anderson, K. J. Students' stereotypes of professors: An exploration of the double violations of ethnicity and gender. *Social Psychology of Education: AN International Journal*. (2010) 459-472
 - a. Students read a syllabus and then answer questions about the hypothetical professor and class. There were 2 different syllabi, one for each hypothetical course used, and they varied dependent on the gender, ethnicity, and teaching style of the professor. Gender and ethnicity are indicated by gender pronouns and first and last names, while teaching style was indicated through language in the syllabus to convey either a strict or lenient teaching style. 594 students rated professors on characteristic dimensions such as warmth, professional competence and difficulty. "Women professors were perceived to be warmer than men professors. Even though their syllabuses were identical, the simple presence of a gender identifiable name influenced students' assumptions about the professor." (469) "Latinas were viewed as more warm than Anglo women and Latino men. Latinas who teach composition were also viewed as more warm than Latinas who teach pre-calculus. These patterns suggest that students find Latina professors teaching the gender-consistent composition course as particularly warm compared to other combinations." (469) "Women professors of composition (a female-stereotyped course) with lenient teaching styles were viewed as more competent than men who taught the same course with the same style. This pattern suggests that students use stereotype consistency to rate a

Suggested Citation: Holman, Mirya, Ellen Key and Rebecca Kreitzer. 2019. "Evidence of Bias in Standard Evaluations of Teaching." <http://www.rebeccakreitzer.com/bias/>

professor's competence.” (469) “Women's and Latina/os' student ratings were contingent on teaching style—they are rewarded or penalized for their style—while Anglo men's ratings were not.” (470)

3. Anderson, K. J. and M. Kanner. Inventing a gay agenda: Students' perceptions of lesbian and gay professors. *Journal of Applied Psychology*, (2011) 41-1538-1564
 - a. A sample of ethnically diverse undergraduates (622 students in Study 1; 545 students in Study 2) read a syllabus for a proposed Psychology of Human Sexuality course and responded to it. Manipulated sources of information in the syllabi were whether or not the course was taught by (a) a woman or a man; (b) a lesbian/gay man or a heterosexual; (c) a politically conservative or politically liberal professor; and (d) a professor with a syllabus containing typographical errors or no errors. Study 1 utilizes a course syllabus paradigm to examine students' perceptions of lesbian and gay professors teaching a course on human sexuality. Study 2 added students' self-reported attitudes toward lesbians and gay men as another potential predictor of perceptions of professors. It used students' responses to the Modern Homonegativity and Homonegativity scales to “categorize the students according to old-fashioned and modern views about homosexuality.” Study 1 found lesbian and gay professors who teach a course on human sexuality were viewed as more biased than were heterosexual professors teaching the same course with the same syllabus. In study 2, they find that “regardless of who taught the course, and regardless of the political perspective of the professor, homonegatives found the course to be problematic. Specifically, homonegatives (relative to non-homonegatives) tended to believe that the professors teaching the course were politically biased.” “Regardless of political ideology, that modern homonegatives viewed lesbian/gay professors to be more politically biased than heterosexual professors with the same syllabus. Similarly, modern homonegatives viewed lesbian/gay professors as more biased than did non-homonegatives, but less biased than homonegatives.”
4. Anderson, K. and E. Miller. Gender and student evaluations of teaching. *Political Science & Politics*, 30(2):216-219, June 1997.
 - a. Authors briefly review the literature on SETs and broadly synthesize major findings, making recommendations for evaluation based on the findings. They draw three conclusions. First, issues in measurement on SETs should speak to the need for multiple means of evaluating teaching effectiveness; the authors suggest a strong reliance on SETs should be avoided. Second, variation in teaching styles are a factor not often considered in evaluation and the authors find that student-centered contact is often undervalued. Finally, students have gendered expectations of teaching styles that can impact how they evaluate faculty.
5. Anderson, Kristin J. and Gabriel Smith. 2005. Students' Preconceptions of Professors: Benefits and Barriers According to Ethnicity and Gender. *Hispanic Journal of Behavioral Sciences*
6. Arbuckle, J. and B. D. Williams, Students' perceptions of expressiveness: Age and gender effects on teacher evaluations, *Sex Roles*, 49 (9/10), 507-516.
 - a. All students watched the same 35-min picture slide-audio taped presentation. The slide was a picture of a computer-generated genderless stick figure; the audiotape was read by a 45-year-old woman whose voice did not suggest her age or gender. 198 female and 154

Suggested Citation: Holman, Mirya, Ellen Key and Rebecca Kreitzer. 2019. “Evidence of Bias in Standard Evaluations of Teaching.” <http://www.rebeccakreitzer.com/bias/>

- male college students from a medium-sized western university (N = 352). The participants were given a 12-item evaluation form to rate the effectiveness of the professor on a 1-6 scale, as well as determining the age and gender of the professor speaking. Students rated the ““young” male professor description higher than they did the “young” female, “old” male, and “old” female professors on speaking enthusiastically and using a meaningful voice tone during the class lecture regardless of the identical manner.” Participants rated professors similar for objective, content information.
7. Baldwin, T. and N. Blattner, Guarding against potential bias in student evaluations. What every faculty member needs to know, *College Teaching*, 51(1) (2003), 27-32.
 - a. This article describes the influences of course evaluations on promotion, tenure, and merit decisions. It describes how faculty can combat biases to ensure that their teaching effectiveness is accurately portrayed. It reviews alternative evaluation methods, including portfolios, peer feedback sessions, and informal student surveys
 8. Baker, P & M Copp, Gender performance matters most: The interaction of gendered expectations, feminist course content and pregnancy in students' course evaluations. *Teaching Sociology* 25(1) (1997), 29-43.
 - a. Used qualitative and quantitative teaching evaluation data from a female professor who taught a feminist course over three semesters while pregnant and carrying the child to term. They analyzed the shifting student reaction to the professor depending on her fulfillment of gendered expectations and the impact on teaching evaluations. The student evaluations contained 2 parts: respond to 10 statements on a scale from 1 to 7 and voluntarily write comments about the course and the teacher. The number of students varied by term: 99 students in the spring, 22 in the summer, and 124 in the fall filled out the questionnaire. Of the 99 in the spring 52 wrote comments, of the 124 in the fall 57 wrote comments, and of the 23 in the summer 20 wrote comments. Gendered expectations of the professor were evident in the written evaluations. “When gendered expectations of their professor were met, students did not find fault with Dr. Baker for being pregnant or teaching from a feminist perspective. However, when students’ expectations for Dr. Baker as a woman were unmet, students’ written comments became negative and hostile and the content of their comments often touched on professor bias, male-bashing, and rudeness, which students attributed to the feminist course content and to Dr. Baker’s pregnancy” (page 40). Their findings demonstrate that student expectations shift based on how professors presents themselves and their situation, which presents challenges for anyone who isn’t a white, able-bodied man.
 9. Basow, S. A. Best and worst professors: Gender patterns in students’ choices, *Sex Roles*, 43(%) (2000), 407-417.
 - a. The sample consisted of 61 female and 47 male students at a small private liberal arts college in the northeast, predominantly White and almost entirely sophomores to seniors (by design) The researchers used a 25-item questionnaire to evaluate 16 male and female professors that were matched on experience, tenure, and division of courses. Questions investigated scholarship, organization/clarity, instructor-group interaction, instructor-individual interaction, and dynamism/enthusiasm. Used a questionnaire consisting of two questions: “Think of the best teacher you’ve had in college” and

Suggested Citation: Holman, Mirya, Ellen Key and Rebecca Kreitzer. 2019. “Evidence of Bias in Standard Evaluations of Teaching.” <http://www.rebeccakreitzer.com/bias/>

“describe what made him or her the best in your opinion.” They found a significant student gender by professor gender interaction on all five factors such that male students gave female professors significantly poorer ratings than they gave male professors. The significant findings on the two interaction factors were attenuated when student-perceived ratings of their professors’ instrumental–active (stereotypically masculine) and expressive–nurturant (stereotypically feminine) personality traits were used as covariates. Both sets of traits were significantly and positively associated with student ratings, and the “best” professors typically combine both; i.e., they are androgynous.”

10. Basow, S. A. Gender dynamics in the classroom. In J. C. Chrisler, C. Golden, & P. D. Rozee, *Lectures on the psychology of women (2nd ed.)*. New York: McGraw-Hill (2000).
11. Basow, S. A. and N. T. Silberg, Student evaluation of college professors: Are female and male professors rated differently?, *Journal of Educational Psychology* 79 no. 3 308-314 (1987).
 - a. Data was collected from analyzing student evaluations of college professors. The study found that male students gave lower ratings to female professor than male professors. The study also found that male students’ ratings of female professors were poorer than female students’ ratings of female professors.
12. Basow, S. A. Student evaluations of college professors: When gender matters, *Journal of Educational Psychology*, 87, 656–665 (1995).
 - a. 553 male and 527 female college students evaluated 16 pairs of male and female professors (32 professors total). Using a multivariate analysis, these evaluations asked students to answer 26 questions, rating the professors on various criteria using a 5-point scale (with 1 being well below average and 5 being well above average). The questions were equally divided across 5 factors: scholarship, organization, instructor-group interaction, instructor-individual student interaction, and enthusiasm. The first question in the evaluation asked about overall teaching quality. The evaluation was administered in the 5th or 6th week of a 14-week semester. Of the 1,080 students who participated, 48% majored in the social sciences, 19% majored in natural sciences, 14% majored in engineering, 5% majored in the humanities, and 9% were undecided. 21% of the students were freshmen, 22% were sophomores, 30% were juniors, and 26% were seniors. The fields of study for the professors varied with 28% in the humanities, 47% in the social sciences, and 25% in the natural sciences. There was not a professor in the engineering department, leaving 1/3 of the class without representation. The study found that both male and female students rated female professors less favorably than male professors (male students rated female professors significantly lower). According to the results of the evaluations, scholarship, organization, and enthusiasm contributed the most to the discrimination of teacher sex. In other words, male students gave significantly different ratings for female professors on the basis on scholarship, organization and clarity while female students rated male and female professors more similarly.
13. Basow, S. A. and S. Montgomery, Student rating and professor self-ratings of college teaching: Effects of gender and divisional affiliation, *Journal of Personnel Evaluation in Education*, 18, 91-106, (2005).

- a. This study analyzed 803 undergraduate student evaluations of 20 female and 23 male professors at a liberal arts college. Students rated their professors in an evaluation of 26 questions, each focusing on one of 5 teaching factors meant to evaluate the effectiveness of the instructor. Professors rated themselves using the same questions and 9 extra exploratory questions. The questionnaire that sought to specifically find the effect of professor and student gender, and divisional affiliation on student ratings of professors and professor self-ratings. Based on the results of the study, Basow and Montgomery concluded that professor gender and divisional affiliation (department/field of study) contributed to the results of student evaluations. Female professors were rated higher than male professors on two interpersonal factors and on scholarship, and natural science courses were rated the lowest for most factors. Humanities professors received the highest overall ratings however, male professors in the humanities received lower ratings than female professors. However, professor self-ratings varied in regards to divisional affiliation. Overall, the researchers found that there was not a significant correlation between professor self-ratings and students' ratings.
14. Bennett, S. K. Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluations, *Journal of Educational Psychology* 74 no. 2 (1982), 170-179.
 - a. In this study, 253 students in non-science introductory courses at a liberal arts college completed a course evaluation questionnaire. Women instructors were perceived as warmer and more potent individuals. This largely accounted for their higher formal student ratings in specific areas of teaching performance. However, students required women instructors to offer greater interpersonal support and judged them more closely than male instructors in providing it.
 15. Boring, A., K. Ottoboni, and P. B. Stark, Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness. *ScienceOpen Research*, 2016. (doi:10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1)
 - a. Data from a randomized, controlled, blind experiment from the United States was combined with data for a natural experiment at a French university. The French natural experiment analyzed data from student evaluations dated from 2008 to 2013. Students in different sections of the same course, all with different professors, were instructed to take the same final exam. Data was collected from 4,423 students in 1,177 sections at a French university, taught by 379 instructors. The American experiment collected data from four online sections of the same class. One class was taught by a male professor, one was taught by a male pretending to be female, another was taught by a female professor, and the last class was taught by a female pretending to be male. Data was collected using 43 out of 47 student evaluations. The combined data found that student evaluations of teaching (SETs) are influenced by the gender of the instructor. Female instructors received lower scores than males. With regards to the US study, the instructors who were female or thought to be female (male pretending to be female) were given lower scores. The study also found that students reported "female" instructors to be less prompt, regardless of the fact that assignments in all class sections were released simultaneously. The data from the French study found that bias in SETs varies based on the gender of the

Suggested Citation: Holman, Mirya, Ellen Key and Rebecca Kreitzer. 2019. "Evidence of Bias in Standard Evaluations of Teaching." <http://www.rebeccakreitzer.com/bias/>

student as well as other factors. However, it is impossible to adjust for potential biases due to the nature of these factors. The evaluations studied were more indicative of gender bias than of teaching effectiveness therefore, gender bias can skew a SET so severely that more effective teachers could receive worse evaluations.

16. Boring, A. Gender Biases in Student Evaluations of Teachers and their Impact on Teacher Incentives, Working Paper, OFCE-PRESAGE-SCIENCES PO and LEDa-DIAL, April 22, 2015.
 - a. This study collected data from student evaluations of teaching (SETs) to test for potential relationships between teacher incentives of gender biases and student evaluations of professors. The study analyzed 22,665 SET observations collected from 4,423 students evaluating 372 undergraduate instructors at a French university. According to the data, male students were more likely to favor male instructors-- a male student is 30% more likely to give an excellent overall satisfaction rating to male instructors than female instructors. There was also an evident influence of gender stereotypes in the evaluation of teaching dimensions. Students valued organization and class preparation in female instructors and class leadership in male instructors. However, the students learned at a fairly equal level with both male and female instructors. From this data, the authors concluded that improved teaching methods are not necessarily able to be measured by SETs.
17. Boring, A., Gender biases in student evaluations of teaching, *J. of Public Economics*, 145 (January 2017), 27-41.
 - a. Data collected from a French university was analyzed to identify the potential of gender bias in student evaluations of teaching (SETs). The data was compiled from a database of 20,197 SET scores that were collected over the course of 5 years and varying student, professor, and course characteristics. 11,522 of these evaluations were completed by female students and 8,675 were completed by male students; the evaluations were for 359 different professors with 33% identifying as female and 67% identifying as male. The study found that male students give significantly higher scores to male professors on overall satisfaction and teaching dimensions such as the professor's perceived knowledge. Female students also give higher scores to male professors on ability to lead in the class, to relate to current issues, and to contribute to intellectual development. Female students give higher scores to female professors on the dimensions of teaching related to course content and assignments. The results from the study suggest that gender stereotypes may drive students' evaluations of professors.
18. Bray, J. H. and G. S. Howard, Interaction of Teacher and Student Sex and Sex Role Orientations and Student Evaluations of College Instruction, *Contemporary Educational Psychology* 5 (1980), 241-248.
 - a. 36 instructors in the College of Social Sciences at the University of Houston selected one Undergraduate class to participate in this study. 20 students were randomly selected from each class to contribute to a total sample size of 497 students. The study analyzed data from questionnaires to assess the potential relationship between gender role orientations, teacher sex, student sex, and student evaluations of course progress and satisfaction with instructors. The questions used were measured using two rating Bem Sex Role Inventory (BSRI) and the Instructional Development and Assessment System (IDEA). BSRI gave

Suggested Citation: Holman, Mirya, Ellen Key and Rebecca Kreitzer. 2019. "Evidence of Bias in Standard Evaluations of Teaching." <http://www.rebeccakreitzer.com/bias/>

each instructor an androgyny score, classifying them as masculine, androgynous, or feminine while IDEA provided student evaluations. The study found that androgynous teachers received somewhat higher student evaluations. Female students reported higher rates of progress in classes taught by female professors as opposed to male professors.

19. Brooks, V. R. Sex differences in student dominance behavior in female and male professors' classrooms. *Sex Roles*, 8 (7) (1982), 683-690.
 - a. Study examines the role gender (sex) plays in the dominance behaviors exhibited in the classroom--both among students and in male and female professor's classrooms. "Dominance" was to mean both assertiveness (measured by frequency/duration of speech) and aggression (measured by interruptions). These behaviors were recorded from students (n=294; 72 male, 222 female) in six separate instructors' (3 male, 3 female) graduate social work courses. In the male professors' courses, no gender-based differences were found in assertiveness, however, in female professors' courses men both spoke more often and at greater length than their female counterparts. Male students were found to be more aggressive than female students in both male and female professors' classrooms, and interrupted female professors slightly more than they did male professors. Similarly, male students were more likely--in both male and female-led classrooms--to interrupt other students, with a higher frequency of interruptions in female professors' courses. Interestingly, while male students were more likely to interrupt female students than female students were to interrupt male students, a strong majority of interruptions by each gender (about 90% for both genders) were of the opposite gender.
20. Buck, S. and D. Tiene, The combined effects of physical attractiveness, gender, and teaching philosophy on evaluations, *The Journal of Educational Research*, 82, no. 3, 172-177 (1989) (published online January 20, 2015, 172-177).
 - a. An experiment was conducted at a Midwestern state university. The sample size was composed of 42 undergraduate seniors, mostly Caucasian females; 10 of the 42 students were male and 3 were black (all females). The students were given 1 of 4 different photographs- attractive or unattractive teachers who were either male or female. Each photograph also included a description of the teachers' teaching style/philosophy, divided by either an authoritarian or humanistic style. The results of the study showed that attractiveness did not have an effect on the ratings of instructor effectiveness. The study also found that authoritarianism was strongly associated with negative evaluations. However, contradictingly attractive authoritarian females were rated significantly more positively than the other 3 possibilities of authoritarian instructor characteristics.
21. Burns-Glover, A. L. & D. J. Veith, Revisiting gender and teaching evaluations: Sex still makes a difference. *Journal of Social Behavior and Personality*, 10, 69-80 (1995).
 - a. 78 student evaluations were analyzed to test the hypothesis of bias towards masculine personality traits. Participants in the study (35% male, majority white) were asked to rate the desirability of 52 trait adjectives and the importance of 25 behaviors for 3 hypothetical professors- "Sam", "Sarah", or "Dr." Lawson. The results found that male students were mostly responsible for gender distinctions in evaluations. 13 of 20 highly desirable traits were considered to be traditionally masculine. In an analysis for discrimination, "Sam Lawson" was comparable to "Dr. Lawson" suggesting that the label

- of professor or doctor is associated with a male schema. The characteristics of “self-confident”, “stable”, and “steady” were rated more important for “Sarah Lawson” than the other identities.
22. Chamberlain, M. S. and J. S. Hickey, Student evaluation of faculty performance: The role of gender expectations in differential evaluations, *Educational Research Quarterly*, 25(2) (2001) 3-14.
 - a. This study was designed to identify gendered expectations, holding occupation constant, based on evaluations of instructors’ performance in the classroom. Undergraduate students (N= 198) evaluated male and female instructors from the same disciplines, using the same questionnaire. Questions were selected based on gendered expectations. The researchers found that male and female instructors are evaluated differently on aspects of presentation and classroom structure.
 23. DiPietro, M. & A. Faye. (2005). Online student-ratings-of-instruction (SRI) mechanisms for maximal feedback to instructors. Paper presented at the 30th Annual Meeting of the Professional and Organizational Development Network; Milwaukee, WI.
 24. Drake, S., A. Auletto, and J. Cowen. Grading Teachers: Race and Gender Differences in Low Evaluation Ratings and Teacher Employment Outcomes. *American Educational Research Journal* (2019).
 - a. In 2011, Michigan adopted high-stakes teaching evaluations, which required sanctions for low performing teachers. Using evaluations for 97,446 teachers, they find that Latinx and African American teachers were 2 and 3 times more likely to receive a low rating than White teachers. Male teachers were 50% more likely to receive a low rating compared to female teachers. Low effectiveness ratings were also associated with working in a charter, high poverty- or high-minority school.
 25. Dukes, R. and G. Victoria. The Effects of Gender, Status and Effective Teaching on the Evaluation of College Instruction. *Teaching Sociology* Vol. 17, No. 4 (Oct., 1989), pp. 447-457
 26. El Alayi, Amani, Ashley A. Hansen-Brown, Michelle Ceynar. Dancing Backwards in High Heels: Female Professors Experience More Work Demands and Special Favor Requests, Particularly from Academically Entitled Students. *Sex Roles* 79(3-4), 136-150. (2018)
 - a. Using both a survey of professors (n=88) and an experiment asking students about special favor requests (n=121), the researchers found that academically gifted students held stronger expectations that female (versus male) professors would grant their special favor request, and help stronger negative emotional and behavioral responses when those requests were not granted.
 27. Elmore, P. B. and K. A. LaPointe, Effects of Teacher Sex and Student Sex on Evaluation of College Instructors, *Journal of Higher Education* 66 no. 3 (1974).
 - a. This study used a matched pairs sampling method to analyze student evaluation surveys from Southern Illinois University at Carbondale for 1,474 courses in 1971. 1,607 student evaluations were used to compile data from a questionnaire called the Instructional Improvement Questionnaire that included questions on instructor evaluation, evaluation of course, strengths and weakness, research data, and other optional questions. According to the results of the survey, there were few significant differences between male and

- female faculty in the results of the surveys. There was also no evident relationship between faculty sex and student sex-- no evidence of potential gender bias in evaluations.
28. Elmore, P. B. and K. A. LaPointe, Effects of teacher sex, student sex & teacher warmth on evaluation of college instructor, *Journal of Higher Ed Psychology* 67 no. 3 (1975).
 - a. This study used 22 pairs of courses evaluated by 838 college students to evaluate if there was a potential influence of teacher sex, student sex, and teacher warmth on student evaluations of teachers as well as teacher evaluations of themselves. The evaluations were compiled through a questionnaire to evaluate instructor performance through a 3 factor analysis. The data showed that students gave higher ratings to teachers that were perceived as warmer and primarily interested in students. When female professors evaluated themselves lower on either factor, they received higher effectiveness ratings from the students. Contrastingly, when male professors perceived themselves as high in either indicator, they received higher effectiveness ratings.
 29. Ewing, Vanessa Lynn, Arthur A. Stokes, and Eugene P. Sheehan. Prejudice against gay male and lesbian lecturers. *The Journal of Social Psychology*, 143(5), 569-578, 2003.
 - a. This study used an “attributional ambiguity paradigm” to compare student reactions to lectures when students did not know the sexual orientation of the speaker and when the speaker was identified as gay or lesbian. When the lecture delivered was “weak,” participants actually rated the gay and lesbian lecturers more highly, which the authors suggest may be an effort to avoid discrimination. However, when the lecture was “strong,” participants rated known gay and lesbian lecturers more negatively than the lecturers whose sexual orientations were not known. According to the authors, these findings suggest how biases and prejudices based on sexual orientation may affect perceptions of “merit, skill, or likeability,” which often weigh heavily in hiring decisions.
 30. Fan, Y., L.J. Shepherd, E. Slavich, D. Waters, M. Stone, R. Abel, and E.L. Johnston. “Gender and cultural bias in student evaluations: Why representation matters.” *PLOS One* 2019
<https://doi.org/10.1371/journal.pone.0209749>
 - a. The researchers used course satisfaction data from student evaluations from 2010-2016 at a national Australian university. There is statistically significant evidence of gender and cultural bias from students, which tends to have a negative impact on the SET scores of women and teachers of non-English-speaking backgrounds. Women who do not come from an English-speaking background were the most affected. Female instructors that have no cultural barriers have a harsher feedback score when they are teaching in the science field rather than the English field. There are no significant biased results for women with no cultural barriers in the Arts and Social Sciences department, but both men and women with cultural barriers are more likely to have harsh feedback. In the Engineering and Medicine department, the significant discrepancy was for harsh feedback was towards female instructors that have cultural barriers. Excluding Engineering, all other departments had results that displayed that male instructors are more likely to receive the highest possible grade of 6 than are female instructors. Male students are going to be more willing to give the higher grade to male instructors, rather than female instructors, specifically in the science department.

31. Ferber, M.A. and J. Huber. Sex of Student and Instructor: A Study of Student Bias. *American Journal of Sociology*, 80(4), 1975.
 - a. 1,281 students at the University of Illinois at Urbana were sent questionnaires to collect data to test the hypothesis that women were prejudiced against women instructors/professionals. The questionnaires were composed of three questions and asked about the number of male and female professors the students' had, the hypothetical preference of having a male or female professor in different classroom settings, and to determine if a student was more willing to agree to an opinion based on whether a male or female stated it. The results of the study were that men received higher ratings from students in sciences, social sciences, and agriculture and women received higher ratings from students in home economics. Male students were also more likely to prefer male professors than female professors. The study also found that men were more likely to agree with an opinion stated by a male while females were equally likely to agree with an opinion stated by a male or female. From the results, the researchers were able to conclude that contact with sophisticated women helps to break stereotypes among both men and women.
32. Feldman, K. A. College students' views of male and female college teachers: Evidence from the social laboratory and experiments – Part 1. *Research in Higher Education*, 33, 317–375 (1992).
 - a. Data was collected from students evaluations in a controlled laboratory environment. College students were asked about their perceptions of individuals identified to be college professors. Students were asked whether they approved of their professors based on characteristics that were not directly related to academics/learning based on photographs, descriptions, and simulations. The professors only differed in their gender and other characteristics that naturally vary such as teaching style and age. The study found that there are no significant differences in the global evaluations of male and female professors, only a small percentage of situations resulted in male professors receiving higher evaluations. The study also found that evaluations of professors can occasionally be influenced by gender-typical behaviors and characteristics.
33. Feldman, K. A. College students' views of male and female college teachers: Evidence from the social laboratory and experiments – Part 2. *Research in Higher Education*, 34, 151–211 (1993).
 - a. Data was collected from student evaluations that were created in a controlled laboratory environment. The students used photographs, descriptions, and simulations that were provided to them to create their evaluations. In this case, the researchers were solely focusing on the potential effect of gender on teaching evaluations. From the study, the researchers were able to conclude that the majority of students did not allow gender to influence their ratings of professors. However, the researchers did claim that gender bias plays a role in the initial process of choosing a professor.
34. Freeman, H. R. 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), (1994) 627-630.
 - a. In the first experiment, researchers examined the effects of course type, student gender, and instructor gender and gender role on student evaluations of instructor effectiveness. In the second experiment, researchers explored students' perceptions of the importance of

Suggested Citation: Holman, Mirya, Ellen Key and Rebecca Kreitzer. 2019. "Evidence of Bias in Standard Evaluations of Teaching." <http://www.rebeccakreitzer.com/bias/>

various gender role characteristics in instructors of different course types. Results suggested that instructor gender role is more important than instructor gender in affecting student evaluations. Female and male students preferred instructors (science instructors, in particular) who had both feminine and masculine characteristics, regardless of the instructor gender.

35. Freiderike, M., J. Sauermann, & U. Zolitz, Gender bias in teaching evaluations, Institute of Labor Economics, September, (2017)
 - a. The researchers analyzed 19,952 student evaluations from a university in the Netherlands. Students at the university were randomly assigned to female or male professors, making the conditions ideal as the potential of gender bias in the selection process could be avoided. 35% of the instructors and 38% of the students involved in the study were females. From analyzing the evaluations, the researchers found that women receive lower ratings in teaching evaluations than males. Specifically, the researchers found that “male students evaluate female instructors evaluate female instructors 20.7% of a standard deviation worse than male instructors.” However, the study also found that female students rate female professors “25.8% of a standard deviation higher than male professors.” Overall, women received lower teaching evaluations than men despite the fact that the instructors’ genders had no clear effect on students’ grades or effort in class.
36. Greenwald, A. G. & G. M. Gillmore, No pain, no gain? The importance of measuring course workload in student ratings of instruction, *J of Educational Psychology*, vol. 89, no. 4, 743-751 (1997).
 - a. Data was collected from 205 academic courses over the course of 3 academic semesters at the University of Washington. The researchers analyzed various aspects of student evaluations such as evaluation ratings, expected course grades, and course workloads. According to the results, the grading leniency of instructors influences student ratings, in other words students gave higher ratings to courses in which they received higher grades and lighter workloads.
37. Hammerness, Daniel and Amy Parker. Beauty in the classroom: instructors’ pulchritude and putative pedagogical productivity. *Economics of Education Review* 24 (2005) 369-376
38. Heilman, M. E., and T. G. Okimoto, Why Are Women Penalized for Success at Male Tasks? The Implied Communitarian Deficit, *Journal of Applied Psychology* 92(1) (2007), 81-92.
 - a. Three studies were analyzed to determine whether women were penalized for success at typically male-oriented gendered tasks. All three studies used questionnaires that manipulated the sex of the individual in question and the information provided about the individual in the scenario. The first study..... The study found that women who are successful in male gender-typed domains are penalized for their success. These women are perceived more negatively, they are disliked more than men who are described identically and are found to be less desirable as bosses.
39. Hessler, M, D. M. Popper, D. Hollstein, H. Ohlenburg, P. H. Amemann, C. Massoth, L. M. Seidel, Z. Zarbock, W. Wenk. Availability of cookies during an academic course session affects evaluation of teaching. *Med Edu* 52(10): 1064-1072 (2018)
 - a. This study implemented a randomized control experiment during an emergency medicine course. Participants were randomly assigned to 20 groups, 10 of which had free cookies

Suggested Citation: Holman, Mirya, Ellen Key and Rebecca Kreitzer. 2019. “Evidence of Bias in Standard Evaluations of Teaching.” <http://www.rebeccakreitzer.com/bias/>

during the session. All groups were taught by the same teachers and the educational content was the same for all groups. 112 students completed the evaluation forms, which showed that students rated the teachers better, the course materials better, and the overall course significantly higher in the cookie group.

40. Hogan, C. Review of the literature: The evaluation of teaching in higher education, Instructional Development Center, McMaster University (1978).

41. Huston, T. (2005) "Race and Gender Bias in Higher Education: Could Faculty Course Evaluations Impede Further Progress Toward Parity?," Seattle Journal for Social Justice: Vol. 4 : Iss. 2 , Article 34.

- a. The authors of this publication reviewed and referenced data from other studies to identify the potential for race and gender bias in higher education. The majority of the data comes from student evaluations. The data provided that female faculty, in general, are not rated significantly lower than male faculty. However, there are two exceptions where there female instructors are systematically rated lower than male instructors-- for large introductory courses and in male-dominating disciplines (science, economics, engineering).

42. Kaschak, E. Another look at sex bias in students evaluations of professors: Do winners get the recognition that they have been given? Psychology of Women Quarterly, Summer, 1981.

43. Kaschak, E. Sex bias in students' evaluations of professors' teaching methods. Psychology of Women Quarterly, 3 (3) (1978), 135-143.

- a. 100 first year and senior graduate students (50 male and 50 female) at San Jose State University were asked to evaluate the teaching methods of three male and three female professors using a scaling method. The professors were evaluated teaching in two "male", two "female", and two non-sex-linked academic fields. The mean ratings for the professors found that both male and female instructors were rated equally by female students, but male students rated male professors higher than female professors. In regards to perceived "power", both male and female students rated male professors higher however, the students also said that they would prefer to take the female professors' course over the male professors'. The researchers mentioned that the lower power rating for female professors may not be influenced by student bias and may be attributed to the gender dynamic in their respective academic fields. It was concluded that the variation in academic field had little to no effect on the ratings of the professors, male or female.

44. Kierstead, D., D'Agostino, P., & Dill, H. (1988). Sex role stereotyping of college professors: Bias in students' ratings of instructors. Journal of Educational Psychology, 80(3), 342-344.

- a. Two experiments were analyzed to attempt to identify the impact of social contact, facial expression, and instructor sex on student ratings of instruction (SRIs). The first experiment consisted of a survey that asked participants to read about and rate a hypothetical professor (male or female), who either did or did not have social contact with students out of class. The second experiment asked students to watch videos of male or female professors lecturing either with or without smiling and then rate the performance of the instructor. The results of the first experiment showed that male instructors were rated higher than female instructors and that female instructors who has social contact with students were rated higher (social contact did not influence male

- instructor ratings). According to the second experiment, more students said they would take a course with the male instructors and were described using adjectives that emphasized intelligence. Female professors received fewer adjectives related to intelligence. The unsmiling male was rated a little higher than the smiling woman but the smiling woman was rated significantly higher than the unsmiling female.
45. Laube, H., K. Massoni, J. Sprague, A. Ferber, The impact of gender on the evaluation of teaching: What we know and what we can do. *NWSA Journal*, 19(3), (2007) 87-104.
 - a. Some scholars have found gender to have no (or very little) influence on evaluations of teaching, whereas other scholars have found gender to affect evaluations significantly. The authors draw on insights from sociological scholarship on gender and evaluation and argue that this apparent inconsistency is itself an artifact of the way that quantitative measures can mask underlying gender bias. The authors offer concrete strategies for faculty, researchers, and administrators to improve the efficacy of the system of evaluation.
 46. Lombardo, J. P. and M. Tocci, Attribution of positive and negative characteristics of instructor as function of attractiveness and sex of instructor and sex of subject, *Perceptual & Motor Skills*, 48 (1979).
 - a. 60 male and 60 female undergraduate students were asked to rank a professor based on their perceived level of attractiveness. The students were also provided with a short description that was identical for each hypothetical professor excluding the name. The students were then asked to rate each professor on an 11-point scale based on specific positive and negative personality traits. The study found that the students rated the attractive professors higher, regardless of sex, than the unattractive professor. However, the researchers also noted that in regards to the characteristics that the professors were rated on, almost all male instructors were rated as more competent than their female counterparts, regardless of attractiveness.
 47. MacNell, L., A. Driscoll, and A. N. Hunt, What's in a name: Exposing Gender Bias in Student Ratings of Teaching, *Innovative Higher Education*, 40 (4), pp 291-303 (2015) (published online 05 December 2014).
 - a. 72 students from an online introductory course were asked to complete online evaluations of their instructors' performance. The instructors were rated on a 5 point scale for 15 questions on various factors such as overall quality and effectiveness. The survey measured aspects such as effectiveness, interpersonal traits, and communication skills. The study found that students rated the male instructor identity significantly higher than the female (regardless of the actual unknown gender of the instructor). The bias in this experiment is not however, a result of the gendered behavior of instructors but is instead a result of student bias. The online environment in which this experiment was conducted allowed the actual gender of the instructor to be concealed, allowing the biases to be effectively analyzed.
 48. McPherson, M. A., R. T. Jewell, & M. Kim, What determines student evaluation scores? A random effects analysis of undergraduate economics classes, *Eastern Economic Journal*, 35, 37-51 (2009).

- a. This paper uses data from 24 consecutive semesters over the course of 9 years, specifically focusing on economics courses taught at the large university. Data was collected from 618 introductory courses taught by 60 different professors and 379 upper-level courses taught by 22 instructors. Student evaluations were collected from these courses to analyze what factors could impact evaluations of instructors. The study found that instructors who inflated grade expectations could “buy” better evaluation scores and that class size and instructor experience play a role in the evaluation ratings of introductory courses but not in upper-level courses. According to the data, male instructors received higher ratings/scores than female instructors and younger instructors were more popular than older faculty members.
49. Martin, E. Power and Authority in the Classroom: Sexist Stereotypes in Teaching Evaluations. *Journal of Women in Culture and Society*, 9 (1984), 482-492.
50. Martin, L. Gender, Teaching Evaluations, and Professional Success in Political Science. Presented at the American Political Science Association Annual Meeting, 2013.
 - a. The authors of this study referenced and reviewed data from other studies, specifically publically available student evaluation data from a political science department at a large public university to determine a potential relationship between gender and leadership assessments. According to the data collected, there is a relationship between instructor gender and student assumptions about leadership roles. In particular, the study found that traditionally male-oriented leadership stereotypes and roles play a large role in influencing evaluations of large lecture courses. Interestingly, the perception of leadership roles did not heavily influence student evaluations of smaller courses.
 51. Mengel, F., J. Sauermann, & Ulf Zölitz, Gender Bias in Teaching Evaluations, *Journal of the European Economic Association*, September 2017.
 - a. Students were randomly assigned to a male or female professor at a university in the Netherlands and 19,952 student evaluations were analyzed to identify potential gender bias in evaluating 735 different instructors. The researchers found that female faculty received systematically lower evaluations even though the students’ grades were relatively the same with male and female professors. Male students evaluated female instructors more negatively, 21% of a standard deviation worse than male professors. Female students rated female instructors 8% of a standard deviation lower than male instructors. Both male and female students rated female instructors lower than their male counterparts.
 52. Miles, P. and D. House, The Tail Wagging the Dog; An Overdue Examination of Student Teaching Evaluations, *International Journal of Higher Education*, 40, no. 2, 2015.
 - a. This research study collected data from over 30,000 student evaluations of 225 professors over the course of 6 semesters to measure the effects of factors beyond the professors’ control. The researchers tested six hypotheses: SETs in elective classes will be higher than in required classes, SETs in large, medium, and small classes will be statistically different, SETs for male and female professors will be unequal, SETs and expected course grades will be positively correlated, SETs in large required classes will be lower than those in large elective classes, SETs in medium and large sections of required classes will be statistically equal for male and female professors. The results found that all of the

Suggested Citation: Holman, Mirya, Ellen Key and Rebecca Kreitzer. 2019. “Evidence of Bias in Standard Evaluations of Teaching.” <http://www.rebeccakreitzer.com/bias/>

hypotheses were supported except the third hypothesis as the data supported that the SETs for male and female professors were actually relatively equal. However, female instructors teaching larger classes had substantially lower scores than their male colleagues.

53. Miller, J. & M. Chamberlin, Women are teachers, men are professors: A study of student perceptions, *Teaching Sociology*, 28 (4), 283– 298 (2000).
 - a. 300 student volunteers in 5 sociology classes completed a self-administered survey. The survey listed every professor or instructor who actively taught courses in the department with two added fictitious names. One of the “fake instructors was male and the other was female. The students were told to indicate how familiar they were with each of the instructors on the survey and to indicate the highest degree they believed the professors they were most familiar with had. According to the data, the students believed that on average the male faculty had higher degrees than the female faculty. In reality, all of the instructors on the survey had Ph.D’s.
54. Mitchell, K. and J. Martin, Gender Bias in Student Evaluations *Political Science & Politics*, 2018, *fgender51(3)*: 648-652.
 - a. This research study analyzed data from student evaluations, specifically the evaluation comments and ordinal scoring of instructors. The data was collected from student evaluations from online courses from different universities as well as RateMyProfessor ratings. According to the results of the study, the language used to evaluate male and female professors is significantly different. The female instructors were evaluated differently in intelligence, perceived competence, and personality. The researchers found that women were evaluated more heavily based on personality and appearance and were more likely to be labored as a “teacher” as opposed to a “professor”. The data also provided that male professors received higher ordinal scores regardless of the fact that the courses taught by male and female professors were administered identically. Overall, the data suggested that evaluations were largely different between male and female instructors, making the use of evaluations when making employment decisions discriminatory.
55. Moore, M. Student resistance to course content: Reactions to the gender of the messenger. *Teaching Sociology* 25(2) (1997), 128-33.
 - a. This study experimented with students resistance or the students’ unwillingness to accept or think about ideas that contradict with their sense of social order or values. In this particular case, Moore (an instructor at a University) measured differences in student feedback and facial expressions when Moore was teaching the class in comparison to when a male colleague taught the class. The male instructor was told to state that men in families were the “winners” and women in families were the “losers”. The students stated that the statements were much more acceptable coming from the male instructor than they were when Moore repeated them. A student said the male instructor was more qualified to make such a statement because he was identified to have a Ph.D. However, Moore also had a Ph.D indicating that the students were displaying a form of resistance by being more accepting of a statement made by a man as opposed to a woman.

56. Peterson, D., L. Biederman, D. Anderson, T. Ditonto, K. Roe. 2019. Mitigating gender bias in student evaluations of teaching. *PLOSOne*.
- a. The authors conducted an experiment in four large introductory courses in Spring of 2018: two introduction to biology courses and two introduction to American politics courses. Within each pair of courses, one section was taught by a white, female instructor and one section was taught by a white, male instructor. Students were randomized into one of two conditions. In the control condition, students received the standard electronic SET survey for their department. In the treatment condition, the solicitation and the evaluation instrument used language that informed students about biases. For each of the evaluation questions, there were significantly higher scores for the female faculty in the treatment condition compared to the control. The evidence from our experiment with SET suggests that a simple intervention informing students of the potential for gender biases can have significant effects on the evaluation of female instructors. These effects were substantial in magnitude; as much as half a point on a five-point scale.
57. Piatak, Jaclyn and Zachary Mohr. 2019. More gender bias in academia? Examining the influence of gender and formalization on student worker rule following. *Journal of Behavioral Public Administration*. 2(2)
- a. The researchers ran a survey experiment in which students were given either an oral or written rule for grading paper headings and varying if the rule was directed from a male or female professor. For the written rule, the professor's name changed., For the verbal rule, the name was changed and standardized, computer generated male and female audio recording. The students were then asked to evaluate paper headings that violated several requirements. Male student workers with a female professor 55.8% of the time compared to male students with a male professor, 88.9%. Female students complied with the rule more frequently when the professor was female, however the difference was only 86% versus 78%.
58. Pittman, Chavella. "Race and Gender Oppression in the Classroom: The Experience of Women Faculty of Color with White Male Students." *Teaching Sociology*, 2010.
59. Punyanunt-Carter, N. and S. Carter. Students' Gender Bias in Teaching Evaluations. *Higher Learning Research Communications*, 5(3), 28-37. <http://dx.doi.org/10.18870/hlrc.v5i3.234>
- a. Data was collected from student evaluations filled out by 30 males and 28 females with 17 males and 15 females evaluating a female professor and 13 males and 13 females evaluating a male professor. The students stated that they did not intentionally choose professors of certain genders. Female students were less likely to rate professors highly than male students in regards to teaching style and ability. Likeability did not influence the evaluations. As a result, the study concluded that any evidence of gender bias is completely unintentional.
60. Reid, Landon. D. 2010. The role of perceived race and gender in the evaluation of college teaching on RateMyProfessors.Com. *Journal of Diversity in Higher Education*, 3(3), 137-152.
61. Rivera, L. and A. Tilcsik. "Scaling Down Inequality: Rating Scales, Gender Bias, and the Architecture of Evaluation." January 2019. <https://doi.org/10.31235/osf.io/j2tw9>

Suggested Citation: Holman, Mirya, Ellen Key and Rebecca Kreitzer. 2019. "Evidence of Bias in Standard Evaluations of Teaching." <http://www.rebeccakreitzer.com/bias/>

- a. This study used natural experiment at a large North American university before and after a change in student evaluations that used a 10 point to a 6-point scale. They use instructor-course fixed effects to control for course quality as well as a survey experiment that varied gender and points on the evaluation scale. The natural experiment: “include 105,034 student ratings of 369 instructors in 235 courses (and 625 instructor-course combinations)” a large North American University (page 9) and the survey experiment: 400 students. In the natural experiment, the gender gap in evaluations was smaller with the 6-point scale than under the 10-point scale, controlling for individual instructor. The effect was most prominent in male-dominated fields. The survey experiment replicates the natural experiment. Respondents were more willing to get the perceived female instructor the top rating on a 6-point scale compared to the 10-point scale. Open-ended and Likert questions yield some insight into the mechanisms. “First, even when raters received identical evidence of teaching performance, there was a gender gap in evaluations under the 10-point rating scale. Yet, that gap virtually disappeared when the raters used the 6-point scale. Second, raters saw the female instructor as less brilliant than her otherwise identical male counterpart, but their relatively lower expectations of brilliance for a 6/6 rating meant she was more likely to receive the highest rating on the 6-point than on the 10-point scale.” (page 20)
62. Rosen, A. S. Correlations, Trends, and Potential Biases among Publicly Accessible Web-Based Student Evaluations of Teaching: A large-scale study of RateMyProfessors.com data, *Assessment & Evaluation in Higher Education* January 2017: 1–14.
 - a. Data was collected from RateMyProfessor to find correlations between the different rating aspects such as overall instructor quality, gender, discipline, physical attractiveness and easiness. 7,882,80 ratings were evaluated, only professors with over 20 ratings were included in the data set. The researchers found statistically significant strong positive correlation between clarity, helpfulness, overall quality and easiness. Gender had a statistically significant impact (at least on a department level).
63. Sandler, B. R. Women faculty at work in the classroom, or why it still hurts to be a woman in labor. *Communication Education* (1991)(January), 6-15.
64. Schuster, M. R. & S. R. Van Dyne, The changing classroom. In M. R. Schuster & S. R. Van Dyne (Eds.), *Women's place in the academy*, (pp. 161-171). Totowa, New Jersey: Rowman & Allanheld (1985).
65. Sidanius, J. & M. Crane, Job evaluation and gender: The case of university faculty, *Journal of Applied Social Psychology*, 19(2) 174-197, (1989).
 - a. 9005 undergraduates at a public university were asked to complete an instructor evaluation known as the “Course-Instructor Survey: General Questionnaire”. After analyzing the evaluations, the researchers found that male faculty were given significantly higher evaluations on teacher effectiveness and perceived competence than female instructors.
66. Sinclair, L., Z. Kunda. Motivated stereotyping of women: She’s fine if she praised me but incompetent if she criticized me. *Personality and Social Psychology Bulletin*, 26(11), 1329-1342. <http://dx.doi.org/10.1177/0146167200263002>

Suggested Citation: Holman, Mirya, Ellen Key and Rebecca Kreitzer. 2019. “Evidence of Bias in Standard Evaluations of Teaching.” <http://www.rebeccakreitzer.com/bias/>

- a. This research study asked students to fill out a teaching evaluation the week after final grades for courses were released at a university. The teaching evaluations were completed on an 11 point scale to determine whether there was a potential for test motivated stereotyping. 580 course evaluations were collected, 334 from students who received high grades and 236 from students who received low grades (142 were disqualified for insufficient information) for 318 instructors. The research found that male and female instructors typically gave students the same grades. Both male and female professors received higher ratings from students if they gave higher grades. However, female instructors were given significantly lower ratings from students who received lower grades. The evaluations of female instructors was more heavily influenced by the grades they gave in comparison to the effect of grades on the evaluations of male instructors. Students were also found to be more upset about a low grade when they received a low grade from a female instructor.
67. Smith, B. and B. Hawkins. Examining Student Evaluations of Black College Faculty: Does Race Matter?, *The Journal of Negro Education*, 2011, 80(2): 149-162.
- a. The researchers analyzed data from 13,702 students evaluations compiled over the course of three years. The evaluations contained 36 questions which were measured on a scale from 1 to 5. The demographics of the university at which the study was conducted were predominantly white with 87% of the student population identified as Causasian and only 6.5% identified as African American. The purpose of the study was to identify whether the race of the student impacted the evaluation of Black faculty members. According to the results, the faculty in general had above average scores on most items tested in the questionnaire however, black faculty received the lowest mean scores on the multidimensional questions and “global” items (overall value of course). The researchers suggested that evaluations should focus on multidimensional questions as opposed to global items to get a more accurate evaluation for each faculty member.
68. Spooren, P., B. Brockx, & D. Mortelmans, *On the validity of student evaluation of teaching: The state of the art*, *Review of Educational Research*, 83(4) 598-642, (2013).
69. Sprague, J. & K. Massoni, Student evaluations and gendered expectations: What we can't count can hurt us. *Sex Roles*, 53, 779–793 (2005).
- a. The researchers used a qualitative analysis of words used in student evaluations of instructors to describe the best and worst rated professors. 288 college students responded to an optional survey that asked them behavioral questions about their professors. The questions asked about students’ opinions on social issues, confidence in social institutions, satisfaction and other items. The students were also asked to write down four adjectives to describe the best teacher they had, and repeat the same for the worst teacher they’ve had. Out of the 288 students, 198 were women and 90 were men and the majority of the respondents were white. 135 of the students responded that their “best teacher” was a man and 153 of the students responded that their “best teacher” was a woman. Male instructors were seen as performers and were described as “funny” while female instructors were seen as “caring and nurturing”. The worst male instructors were described as “boring and self-centered” while the worst female instructors were “rigid, mean, and unfair”. According to the differences in descriptive words used for male and

female instructors, it is evident that students hold teachers to certain gender expectations but the burdens are more prevalent for females.

70. Stark, P. B. and Richard Freishtat. An Evaluation of Course Evaluations, ScienceOpen, 26 September 2014.
 - a. The researcher reviewed various pieces of literature to compile the data for their study. The goal of the study was to use statistical information from course evaluations to prove that student evaluations of instructor effectiveness are not accurate measures of teaching effectiveness. The publication emphasizes that due to the nature of course evaluations, universities should not rely on student evaluations to make decisions regarding promotions and tenure for faculty. According to Stark, using averages of evaluations to make employment decisions is not an accurate way to measure instructor effectiveness.
71. Statham, A., L. Richardson, and J. Cook, Gender and University Teaching: A Negotiated Difference, SUNY Press (1991).
 - a. The publication analyzes various aspects of teaching in universities and how students respond to differences in teaching methods. Data was collected from observing professors in classrooms and from student evaluations and feedback. The authors found that the gender of the professor alters how critical students are of the instructor due to the fact that gender is created and influences social expectations.
72. Storage, D., Z. Horne, A. Cimpian, & S-J. Leslie, The Frequency of “Brilliant” and “Genius” in Teaching Evaluations Predicts the Representation of Women and African Americans across Fields, PLOS ONE, March 3, 2016.
 - a. The researchers created a “brilliance language score” by standardizing the frequencies of words such as “brilliant” and “genius” for instructors within each discipline to get a number representation of the usage of these words to refer to female and male instructors. The data of the words usage was collected from RateMyProfessor.com. The results provided that male instructors were more likely to be described as “brilliant” or “genius” than female faculty. However, there did not seem to be a large discrepancy in the usage of the words “excellent” and “amazing”. There was a strong positive correlation between the use of “brilliant” and “genius” to describe instructors and rigorous courses. According to the data, the researchers noticed that women were given less praise on their intellectual ability and intelligence.
73. Subtirelu, N. “She does have an accent but ...” : Race and language ideology in students’ evaluations of mathematics instructors on RateMyProfessors.com. Language in Society February 2015, 44(1), 35-62.
74. Uttl, B., C. A. White, and D. W. Gonzalez, Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related, Studies in Educational Evaluation, 54, September 2017, 22-42.
75. Wagner, N., M. Rieger, and K. Voorvelt, Gender, ethnicity and teaching evaluations: Evidence from mixed teaching teams, Economics of Education Review, vol. 54, 79-94 (October 2016).
 - a. The researchers analyzed data from the International Institute of Social Studies of Erasmus University in the Netherlands. The data was compiled from student evaluations from 2010 to 2015, 688 teaching evaluations were used in the study. According to the results, female instructors were 11 percentage points less likely to reach the line

Suggested Citation: Holman, Mirya, Ellen Key and Rebecca Kreitzer. 2019. “Evidence of Bias in Standard Evaluations of Teaching.” <http://www.rebeccakreitzer.com/bias/>

necessary to receive a promotion in comparison to male instructors. The researchers attempted to reduce potential bias by controlling for various variables such as leadership and teacher experience however, female instructors were still rated significantly lower than males. The data supports the idea that students rate women more negatively in evaluations which disadvantages female faculty looking for promotions.

76. Young, S., L. Rush, & D. Shaw, Evaluating gender bias in ratings of university instructors' teaching effectiveness, *International Journal of Scholarship of Teaching and Learning*, 3, 1–14 (2009)
 - a. Data was collected from 765 undergraduate and graduate students at a university in the U.S. The students were asked to rate instructors using a 25 question survey where they were expected to rate the instructors on three factors-- interpersonal characteristics, pedagogical characteristics, and course content characteristics on a scale from 1 to 9. From the data collected, there seemed to be a significant interaction between student gender and professor gender in regards to teaching characteristics and course content ratings. However, there did not seem to be a significant difference in the ratings of interpersonal characteristics. This insinuates that gender bias does play a role in student evaluations of effective teaching and course content. Female students rated female instructors and male students rated male instructors higher on these specific characteristics.

Bibliography of Research Showing Negligible or No Race or Gender Bias in Student Evaluations

1. Hancock, G., D. Shannon, and L. Trentham. Student and Teacher Gender in Ratings of University Faculty: Results from Five Colleges of Study. *Journal of Personnel Evaluation in Education* 6:235-248, 1993.
 - a. The data is comprised of student evaluations from 29,519 students who attended Auburn University. The goal of the study was to attempt to identify potential gender-oriented biases in student evaluations. The data was collected from student ratings of instructors using an eight question survey that required students to rate statements on a 5-point Likert scale. According to the results of the study, student gender seemed to influence student evaluation however, instructor gender did not seem to have an effect. The researchers found that female students rated instructors higher in general than male students did and that students gave higher ratings to instructors with the same gender as them. The researchers did acknowledge that in order to get a better understanding of external variables that cause differences in student evaluations, the interpretation of rankings must be changed.
2. Wallisch, P. and J. Cachia. 2019. "Determinants of perceived teaching quality: The role of divergent interpretations of expectations." 10.31234/osf.io/dsvqg
 - a. Analyzing a comprehensive set of data from RateMyProfessors.com. We find that gender differences in perceived quality are minute, and that the overall ratings can be well predicted by specific instructor attributes. We conclude that student derived evaluations of teaching are largely unbiased and reflect instructor qualities.

Suggested Citation: Holman, Mirya, Ellen Key and Rebecca Kreitzer. 2019. "Evidence of Bias in Standard Evaluations of Teaching." <http://www.rebeccakreitzer.com/bias/>

3. Wright, S. L. & M. A. Jenkins-Guarnieri, Student evaluations of teaching: Combining the meta-analyses and demonstrating further evidence for effective use, *Assessment & Evaluation in Higher Education*, 37(6) 683-699, (2012).
 - a. The researchers combined 9 meta-analyses that covered 193 studies in various articles. Through analyzing the data, the researchers were able to find a correlation coefficient of .264, indicating a small medium relationship between the variables being studied. Gender of professor was only one of the many variables studied, making it difficult to identify whether the gender of professor was a significant variable in the study. From this study, it is not clear whether there is a relationship between gender of professor and student evaluations of teaching.

Bibliography of Research Showing Women Receive Higher Student Evaluations

1. Bachen, C. M., M. M. McLoughlin, & S. S. Garcia, Assessing the role of gender in college students' evaluations of faculty, *Communication Education*, 48, 193–210 (1999).
 - a. 486 undergraduate students (out of 3500) from a mid-size private university in CA responded to a survey of closed-ended questions about perceptions of male and female faculty characteristics / practices and one open-ended question to write about the differences between the two they've experienced. "Female students tend to evaluate female faculty very highly across both the more traditionally feminine dimensions of caring-expressiveness and an interactive teaching style, as well as the more masculine dimensions of professional-ism-challenge and organization. They also tend to evaluate male faculty considerably lower on these same dimensions. In contrast, the male students evaluated male and female faculty more equally across the five measures of teaching style." P 201 "The most immediate surprise when reviewing the open-ended responses was that the majority of students elected to write (either in positive or negative terms) about their female professors. Despite the wording of the question, asking students for observations about male or female instructors, their responses were most often structured so as to point out how female faculty differed from male faculty. Their comments on how female faculty depart from male faculty (for better or worse) appear to reinforce this understanding of the female professor's anomalous or outsider status." P 203
2. Centra, J. A. and N. B. Gaubatz, Is there gender bias in student evaluations of teaching? *The Journal of Higher Education*, 71(1), (2000) 17-33.
 - a. This study investigated gender differences in student evaluation of teaching through two analyses, based upon data from 741 college classes. In the first analysis, researchers compared female and male student ratings in the same classes for female and male instructors. In the second analysis, researchers examined how student ratings differed for male and female instructors. Female students gave higher ratings to female instructors on three of eight scales for all disciplines combined. Male students gave higher ratings to male instructors on only one scale. Male and female students showed no difference in their rankings of male teachers. For the total sample of classes, when students gave more favorable ratings, such ratings were largely given by female students to female instructors.

Suggested Citation: Holman, Mirya, Ellen Key and Rebecca Kreitzer. 2019. "Evidence of Bias in Standard Evaluations of Teaching." <http://www.rebeccakreitzer.com/bias/>

3. Rowden, G. V. and R. E. Carlson, Gender issues and students' perceptions of instructors' immediacy and evaluation of teaching and course, *Psychological Reports*, 78, 835-839, (1996).
 - a. This study evaluated three different aspects-- teacher immediacy, the instructor overall, and course evaluations. Data was collected from student evaluations from 197 undergraduate students. The data supported a positive correlation between students' evaluations of instructors and the overall course evaluations. There was no significant difference in immediacy ratings between male and female instructors. However, female instructors received higher teacher and course ratings than male instructors.
4. Whitworth, J. E., B. A. Price, & C. H. Randall, Factors that affect college of business student opinion of teaching and learning. *Journal of Education for Business*, May/June (2002), 282-289.
 - a. In this study, the authors analyzed faculty evaluations from 12,153 students to investigate the effects of faculty gender, course type, and course level (graduate versus undergraduate) on the faculty evaluations. They found that female instructors received higher ratings than male instructors and that ratings differed significantly by course type and by students' perceived amount of learning.
5. Wigington, H., N. Tollefson, & E. Rodriguez, Students' ratings of instructors revisited: Interactions among class and instructor variables. *Research in Higher Education*, 30(3), 331-34, (1989).
 - a. The researchers attempted to analyze how situational variables influence a student when they complete evaluations of professors. The interactions between three variables related to the class (type, level, and size) and three variables related to the instructor (reputation, rank, and sex) were analyzed in relation to student evaluations. The study found that the reputation of the instructor prior to the student having him/her influenced whether the student liked the instructor. The study also found that the size of the class influenced course evaluation results and that the sex of the teacher had an impact on the types of evaluations-- male instructors received better course evaluations.