

Bias in Student Evaluations

Susan A. Basow and Julie L. Martin

Lafayette College

The question of whether student evaluations can be biased is a critical one for those using them, whether for formative or summative purposes. If student evaluations reflect more than an instructor's actual teaching ability, such as how attractive the professor is, this information must be taken into account when such evaluations are used. Given the important role student evaluations play in many academic employment decisions--such as hiring, promotion, tenure, salary, and awards--it is vital to understand potential sources of bias (see McCarthy, this volume). In this chapter, we will examine potential biasing factors involving the professor--such as gender, race/ethnicity, attractiveness, and age--as well as the course, such as course difficulty and expected grade.

Instructor Factors

Social psychologists have documented how a rater's perception of and reaction to another person can be affected by bias, either consciously or unconsciously (Biernat, 2003; Eagly & Karau, 2002; Phelan, Moss-Racusin, & Rudman, 2008.) In particular, cultural stereotypes, such as for gender and race, may create different expectations for different individuals. For example, because women are expected to be nurturant and caring, a woman's interpersonal skills may be viewed more critically in a rater's overall evaluation than would skills of her male counterpart. Furthermore, women and minorities often must work harder to be perceived as equally competent as White men (the normative group), and it is far easier for them to "fall from grace" as well (Biernat, Fuegen, & Kobrynowicz, 2010; Foschi, 2000). Thus, students might perceive the same behavior, such as grading harshly, more negatively if the professor is a woman or African American or Hispanic (who "should" be "nice" and "caring") than if the professor is a White man (who has greater legitimacy due to both race and gender). Such indeed seems to be the case as we discuss below.

Documenting potential sources of bias in field research is challenging because professors vary on many factors, all at the same time. That is, professors not only have a gender, they also have a race, an ethnicity, a certain personality and speaking style; they vary in age and type of course taught in terms of size, level, and discipline. Because it is likely that many of these factors interact to create a particular impression, it is difficult to tease apart only the effects of age or race or gender, etc. In order to do that, laboratory studies often are performed, but their external validity then may be questioned. We will examine both laboratory and field research on this issue in order to explore the possible biasing effects of a professor's gender, race/ethnicity, age, and level of attractiveness on student ratings of teacher effectiveness. We will also attend to the actual questions being asked, because different aspects of teaching effectiveness (e.g., knowledge, dynamism, concern about students) may show different patterns.

Instructor gender

Many research studies have examined whether faculty gender affects student ratings and generally results suggest the negative (e.g., Bennett, 1982; Feldman, 1993); that is, women faculty do not appear to get lower evaluations than do male faculty across the board. This seemingly reassuring result, however, is deceptive because gender appears to operate in interaction with other variables, such as the gender of the rater, the gender-typing of the field in which one teaches, one's gender-typed

characteristics, and status cues. Although gender effects, when found, generally are small in size, they still may have an impact on the ratings received by some women faculty.

The most frequent finding is that teacher gender interacts with student gender to influence student ratings. Whereas male faculty tend to be rated similarly by their male and female students, female faculty tend to be rated lower by their male students and sometimes higher by their female students (Basow, 1995; Basow & Silberg, 1987; Centra & Gaubatz, 2000; Feldman, 1993). The male students who are most likely to devalue their female professors tend to be business and engineering majors, students who tend to hold the most traditional attitudes toward women. Although male students are more likely to rate their female professors lower than their male professors and are less likely to consider them one of their “best” professors, they are not more likely to consider them their “worst” professor (Basow, 2000; Basow, Phelan, & Capotosto, 2006). In contrast, female students often do choose women faculty as “best” and rate them higher than male faculty, especially on qualities related to “fairness” and “providing a comfortable classroom environment.” In general, men faculty often are rated higher than women on questions related to scholarship/knowledge and dynamism/enthusiasm, while women faculty often are rated higher than men on questions relating to faculty-student interactions and quality (Bachen, McLoughlin, & Garcia, 1999; Basow & Montgomery, 2005; Bennett, 1982). For example, in Centra and Gaubatz’s (2000) study of 741 classes at 21 institutions, male professors were evaluated similarly by their male and female students, but female professors were rated higher by their female students overall and on questions relating to communication and faculty-student interaction.

The subject matter that a professor teaches also plays a role in student ratings. Overall, humanities professors tend to get higher ratings with natural science and engineering professors getting the lowest ratings (Basow, 1995; Basow & Montgomery, 2005). Teacher gender tends to interact with student gender in the humanities and social sciences, with female faculty receiving lower ratings from their male students than they do from their female students. But in the natural sciences, all students tend to rate female faculty lower than male faculty, especially on questions such as “demonstrates knowledge.” This result may be due to the fact that the sciences are considered traditionally masculine fields.

Gender as well as discipline may affect a particular professor’s teaching style. For example, men are more likely than women to use a lecture-based teaching style, perhaps because they are more likely to be teaching fact-based courses, such as the physical sciences (Basow & Montgomery, 2005; Brady & Eisler, 1999; Canada & Pringle, 1995). Conversely, women are more likely than men to use a more discussion-based teaching style, perhaps because they are more likely to be teaching humanities courses. Still, even when faculty members are matched in terms of rank and discipline, female faculty are found to be more student-oriented and to engage students more in discussions than their male counterparts (Statham, Richardson, & Cook, 1991). In contrast, male faculty appear more likely than female faculty to assert their authority in the classroom through public reprimands and corrections. It may be that these different teaching styles appear gendered to students as well such that women who use a lecture-based teaching style are evaluated more negatively than men who do so.

Teacher personality characteristics also may affect student evaluations in gendered ways. Because women are expected to be caring, they are judged more critically than their male counterparts when they appear to violate students’ gendered expectations, such as by being demanding, grading harshly, not accepting student excuses, and not being always available (Basow et al., 2006; Bennett, 1982; Sinclair & Kunda, 2000). Sprague and Massoni (2005) in their qualitative study found that students expect more of their women professors compared to their men professors in terms of time and help and react with greater hostility if these expectations are not met.

In general, women faculty bear the burden of higher expectations. Because gender expectations of men overlap considerably with expectations of professors (e.g., competence, knowledge, high status), a male professor has credibility regardless of age or appearance. For example, Arbuckle and Williams (2003) found that students rated a “young” male professor higher than they rated a “young” female professor in a laboratory study that used the exact same lecture but varied the description of the professor in terms of age and gender. Thus age (and other low status cues) may interact with gender to affect student ratings of women and not men. This may be because gender expectations of women do not overlap much with expectations of professors. Therefore, women professors often have to “prove” that they are credible by more clearly displaying expected “professor” qualities of knowledge, competence, and assertiveness, but must do so while also displaying expected “feminine” qualities, such as warmth and nurturance.

In summary, women faculty are expected to be more available and more nurturing than men faculty, and they typically are. But these qualities only result in comparable evaluations, not higher ones. If, however, female professors are not *more* available and nurturant than their male counterparts, such as by having more office hours or requiring less work, they will be rated *lower* than similar male colleagues. Thus, comparable ratings of male and female faculty may mask a differential set of student expectations for faculty behavior. For particular women (e.g., those who appear young, who are in a stereotypically masculine field, who have a no-nonsense teaching style and teach primarily male students), gender variables can have a negative impact on their student evaluations.

Professor race and ethnicity

Relative to professor gender, the effects of professor race and ethnicity on student evaluations have not been widely studied. This may be due to the relatively low percentage of non-White faculty at institutions of higher education in the U.S. Similar to instructor gender, however, it is likely that racial and ethnic stereotypes affect students’ perceptions of and reactions to minority faculty. In particular, African American and Hispanic professors are likely to have to “prove” their knowledge and competence in ways that White professors do not. It also is likely that instructor race/ethnicity interacts with professor gender as well as with student race/ethnicity but very few studies have examined these interactions.

Some research suggests that minority professors may enact their roles differently than White professors. For example, Harlow’s (2003) interviews with White and African American faculty members at a large predominantly-White state university found Black professors showed disproportionate amounts of doubt, questioning of their own status, and feeling the need to prove their abilities, as compared to the White professors. Minority faculty, especially women, also appear to have a greater service and mentoring burden than their White counterparts (Griffin & Reddick, 2011) that may translate into higher student expectations of these qualities in minority faculty, similar to those found for White women.

In general, African American and Hispanic faculty appear to receive lower evaluations than White and Asian faculty (Hamermesh & Parker, 2005). For example, using student evaluations of faculty at the top 25 liberal arts colleges in the U.S. posted on the website ratemyprofessor.com, Reid (2010) found that Black faculty, especially Black men, were evaluated more critically and given lower ratings on quality, helpfulness, and clarity than their White counterparts. Similarly, Smith (2007) found that White faculty were rated consistently higher than Black faculty on global measures of overall teaching at a large university in the southern United States. Because these are naturalistic studies, it is not clear whether such differential ratings are due to bias, to actual differences in teaching effectiveness, or to other

factors, such as the subject matter being taught or teaching style. Some indication that bias may be involved comes from a study by Ho, Thomsen, and Sidanius (2009) who examined how ratings of professor intellectual competence and sensitivity to students related to ratings of overall teaching effectiveness in a sample of 5,655 randomly-selected students. Although the overall performance ratings of African American faculty did not differ from those of White faculty, students' perceptions of intellectual competence were a bigger factor in the overall performance evaluations of Black compared to White faculty. This finding is consistent with social psychological research findings that lower status groups (e.g., women, African Americans) must "prove" competence when evaluated for high status positions (e.g., Foschi, 2000).

In one of the few laboratory studies on the effects of professor race (White, Asian, or African American) and gender on student evaluations, Bavishi, Madera, and Hebl (2010) asked entering college students to rate a Curricula Vita as to the competency, legitimacy, and interpersonal skills of the hypothetical professor. Results revealed that African American professors, especially women, were rated the lowest on all three dimensions and Asian professors were rated lower than the White professors on interpersonal skills. Therefore, although Asian professors may be viewed more positively than Black professors, they still may experience some negative perceptions based on race.

The type of course that minority faculty members teach may also affect how students perceive and rate them. Given that White males are viewed as the "normative" professor, both women and minority faculty may be viewed both as less legitimate and less objective. In race-focused diversity courses, most likely to be taught by minority faculty, students appear to view African American faculty as more biased and subjective, although more knowledgeable, than White faculty teaching the same course (Anderson & Smith, 2005; Littleford, Ong, Tseng, Milliken, & Humy, 2010). In general, faculty teaching about White privilege to White students often receive lower student evaluations in those courses than in their other courses (Boatright-Horowitz & Soeung, 2009), a finding that may contribute to the lower ratings of African American and Hispanic faculty frequently found.

Given the paucity of minorities among the professorate as well as the student body, it is not clear whether minority students react the same way to minority professors, especially of their own race/ethnicity, as White students do. In a few naturalistic studies, students have been found to receive more attention and support by a faculty member of the same race than by a faculty member of another race (Dee, 2005; Ehrenberg, Goldhaber, & Brewer, 1995). It is possible that teachers' in-group support may influence students' perceptions and evaluations such that a student in-group preference is found, at least on some questions. As noted above, this frequently is the pattern found regarding how student gender and faculty gender interact. In other contexts, such as doctor-patient relationships and peer friendships, race-concordance is associated with more positive ratings (Cooper-Patrick et al., 1999; Verkuyten, 2007). Research is needed on this question.

Overall, it is likely that race/ethnicity of the professor affects student ratings, with White faculty generally rated higher than minority faculty, but this effect may depend on the specific question (e.g., overall, interpersonal) and other variables, such as professor gender, student race/ethnicity, and type of course being evaluated.

Professor age

Another demographic variable that has been insufficiently studied is professor age. This may be particularly salient as the baby-boom generation, now entering retirement, moves through the

professorate. As with other professor factors, it is likely to operate in interaction with other variables, such as professor gender, student age, and type of course taught.

Age discrimination has been found in the workplace, with older job applicants often viewed more negatively than younger and middle-aged applicants (e.g., Kite, Stockdale, Whitley, & Johnson, 2005; Krings, Sczesny, & Kluge, 2011). In order to examine the effect of professor age on student evaluations, several laboratory studies have been conducted, wherein a professor's age is varied by either pictures and/or written descriptions. In two studies using a hypothetical male professor of three different ages (25, 53, 73), college students tended to rate the oldest professor most negatively (Levin, 1988; Stolte, 1996). In Arbuckle and Williams' (2003) experimental study, female and male students watched slides of an age- and gender-neutral stick figure and listened to a neutral voice presenting a lecture. Students then evaluated the stick-figure professor on teacher evaluation forms that indicated one of four different age and gender conditions—male or female; “old” (over 55) or “young” (under 35). The “young” male professor received significantly higher ratings on questions tapping “enthusiasm” and “meaningful voice tone” than did the “young” female professor and both “old” professors. There was no effect of age or gender on ratings of “seemed to be relaxed and confident.” Thus age and gender may interact to the benefit of younger males, at least on some evaluative questions. Because these are all laboratory studies and two of the three are over 15 years old, it is not clear how well they relate to contemporary evaluation contexts.

In the only quasi-naturalistic study found that explored the effect of professor age, Radmacher and Martin (2001) recruited a volunteer sample of 13 professors (2 male, 11 female) from different disciplines to collect mid-term evaluations from their students ($N = 351$). Results indicated that teacher age was significantly negatively correlated with ratings of teacher effectiveness, but the size of the correlation was small. Given the double-standard of aging in U.S. culture wherein older women are viewed even more negatively than older men (Kite et al., 2005), and the predominance of women professors in the Radmacher and Martin study, it is possible that instructor age and gender may interact.

Overall, professor age is understudied and likely to operate in interaction with other factors in affecting student evaluations. There is some evidence that older faculty may receive “lower” evaluations relative to “younger” ones. Whether this relationship is curvilinear (i.e., perhaps the best-rated professor is in her/his 40s rather than 30s or 50s) needs to be examined further.

Professor attractiveness

Professor attractiveness is a more seemingly-subjective variable than the other categories of potential bias (gender, race/ethnicity, age) we have discussed. Considerable research has demonstrated that attractive people generally are liked better and perceived more positively than their less-attractive peers (Dion, Berscheid, & Walster, 1972; Langlois, Kalakanis, Rubenstein, Larson, Hallam, & Smoot, 2000). This preferential treatment of attractive people holds true for both females and males and has been found in student evaluations as well.

Economists Hamermesh and Parker (2005) took pictures of professors from departmental websites and various other sources at the University of Texas in Austin and had six students (3 male, 3 female) rate the attractiveness of each picture. Using each professor's average attractiveness rating along with other demographic and course variables, the researchers examined which factors significantly predicted the end-of-term ratings of course excellence. Results indicated that the effects of professor attractiveness on average course ratings were significant and strong, especially in lower division courses. Moving from

one standard deviation below the mean in attractiveness to one standard deviation above led to nearly a full standard deviation increase in the average class rating of teacher effectiveness. Interestingly, attractiveness ratings accounted for more of the variance in the ratings of male compared to female professors, perhaps because women were rated lower than men on teaching effectiveness, or because other factors may contribute more to ratings of women than to men (e.g., personal traits, teaching style, type of course, gender of student).

Other research has found similar results. For example, using student ratings from the website *ratemyprofessor.com*, Riniolo and colleagues (2006) found that those teachers with higher scores on the attractiveness scale also had higher scores on ratings of teacher effectiveness. This significant positive correlation was found for both men and women. Although data on this website is not a random sample, the findings are consistent with other research. Overall, physical attractiveness of a professor may bias student evaluations in a positive direction. Further research is needed to examine the possible interaction between professor gender and professor attractiveness on students' evaluations.

Course-Related Factors

As student evaluations have increased in importance in employment-related decisions, so has concern over whether faculty can improve the ratings they receive by reducing the rigor of the course material or by grading more leniently. These two factors have received extensive study and the answer is complicated (see also Keeley, this volume and Wilson & Ryan, this volume).

Expected grade

Some relationship between grades students receive and student evaluations of teacher effectiveness is to be expected. Indeed, one way to assess the validity of student evaluations is to ascertain if the teachers with the highest ratings also have students who learn more, as assessed by their academic performance and the grades they receive. The best test of this relationship is in multiple-section courses taught by different instructors where all students take a common final exam. In such studies, there is a modest but significant positive correlation between student academic performance and student evaluations (Cohen, 1981).

In other research designs, however, where there is no common course material or common exam, the relationship between student grade and teacher evaluations may reflect something other than teaching effectiveness—it may reflect student “liking,” or gratefulness for a high grade, or student unhappiness and dislike for a low grade, regardless of actual student learning. Especially for untenured faculty, this presumed relationship has been viewed as contributing to grade inflation (Eizler, 2002)—an increase in grading leniency in order to obtain more positive student evaluations.

This assumed grading effect is a source of great concern for those who have to interpret and utilize student evaluations. In order to assess the research on whether such a grading effect occurs, it is important to make a distinction between actual and expected grade. Since course evaluations typically are completed before students receive their final course grade, a grading effect would be demonstrated by a greater correlation between students' expected grade and student ratings of their professor than between students' actual grade and their ratings. Such indeed is the case (Felton, 2008; Greenwald & Gillmore, 1997; Millea & Grimes, 2002). For example, Ducette and Kenney's (1982) examination of 456 classes at an eastern university over the course of three years found that students who expected higher grades gave their instructor higher ratings on effectiveness than those who expected lower grades. However, there was no significant relationship between actual grades and ratings of teacher effectiveness.

Eizler (2002) investigated whether the use of student evaluations of teaching effectiveness contributed to grade inflation by examining student evaluations in more than 37,000 course sections between 1980 and 1999 in a mid-sized, public university in the upper Midwest. The percentage of students expecting A/A– grades was relatively stable during the 1980s but increased steadily by more than 10% over the next 10 years; the same pattern was found in student ratings of teaching (i.e., student ratings were relatively stable during the 1980s but increased steadily in the '90s.) The correlations between expected grade and student ratings remained significant even after controlling for variables tapping alternative explanations, such as prior achievement, course popularity and instructor appeal. Thus, it seems likely that grading leniency can bias student ratings in the positive direction.

Course difficulty

Course difficulty is another course factor that may affect student ratings either directly (if students give higher ratings to professors of “easier” courses than to those of more challenging ones) or indirectly through its effect on grades (i.e., students in “easier” courses may expect higher grades, and it is this higher expected grade that is associated with higher student ratings). The research on this question does not present a clear picture.

Some studies find a direct relationship (e.g., Addison, Best, & Warrington, 2006) wherein courses considered easier than expected receive higher ratings than courses viewed as harder than expected, regardless of student grade. Other studies find evidence of an indirect relationship between ratings of course difficulty and student evaluations, mediated by expected course grade (e.g., Ducette & Kenney, 1982). Still other studies find either no relationship or a more complicated one (Heckert, Latier, Ringwald-Burton, & Drazen, 2006; Marsh & Roche, 2000; Millea & Grimes, 2002; Zabaleta, 2000). For example, there may be a relationship between perceived course difficulty and student ratings only for students expecting lower grades. In some cases, greater student effort is associated with higher rather than lower student ratings.

Summary and Implications

Although there is still a considerable amount of research needed to understand all the ways that student evaluations can be biased, this chapter suggests that not only is some bias possible but it is likely. As a human activity reliant upon person perception and interpersonal judgment, student ratings are affected by the same factors that can potentially affect any rater’s judgment: stereotypes based on gender, race/ethnicity, age, and other qualities (such as professor sexual orientation); the equation of “what is beautiful is good;” more positive feelings towards those who seem to reward us (e.g., with good grades). Even though the size of individual effects may be small, for specific professors these small effects may add up to make a meaningful difference on the ratings they receive. Although the average-looking young-to-middle-aged White male professor teaching traditional courses may receive student ratings that are relatively unbiased reflections of his teaching effectiveness, other professors (women, minorities, older, unattractive-looking, teaching diversity-related courses) may receive evaluations that reflect some degree of bias. It behooves those who use such ratings for evaluative purposes to understand the subtle ways such variables may operate, especially in interaction with each other.

References

Addison, W.E., Best, & Warrington, J.D. (2006). Students' perceptions of course difficulty and their ratings of the instructor. *College Student Journal*, 40, 409-416.

- Anderson, K. L., & Smith, G. (2005). Students' preconceptions of professors: Benefits and barriers according to ethnicity and gender. *Hispanic Journal of Behavioral Sciences*, 27, 184-201. [10.1177/0739986304273707](https://doi.org/10.1177/0739986304273707)
- Arbuckle, J., & Williams, B. D. (2003). Students' perceptions of expressiveness: Age and gender effects on teacher evaluations. *Sex Roles*, 49, 507-516. [10.1023/A:1025832707002](https://doi.org/10.1023/A:1025832707002)
- 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. [10.1080/03634529909379169](https://doi.org/10.1080/03634529909379169)
- Basow, S. A. (1995). Student evaluations of college professors: When gender matters. *Journal of Educational Psychology*, 87, 656-665. [10.1037/0022-0663.87.4.656](https://doi.org/10.1037/0022-0663.87.4.656)
- Basow, S. (2000). Best and worst professors: Gender patterns in students' choices. *Sex Roles*, 43, 139-149. [10.1023/A:1026655528055](https://doi.org/10.1023/A:1026655528055)
- Basow, S. A., & Montgomery, S. (2005). Student evaluations of professors and professor self-ratings: Gender and divisional patterns. *Journal of Personnel Evaluation in Education*, 18, 91-106. [10.1007/s11092-006-9001-8](https://doi.org/10.1007/s11092-006-9001-8)
- Basow, S. A., Phelan, J., & Capotosto, L. (2006). Gender patterns in college students' choices of their best and worst professors. *Psychology of Women Quarterly*, 30, 25-35. [10.1111/j.1471-6402.2006.00259.x](https://doi.org/10.1111/j.1471-6402.2006.00259.x)
- Basow, S. A., & Silberg, N. T. (1987). Student evaluations of college professors: Are female and male professors rated differently? *Journal of Educational Psychology*, 79, 308-314. [10.1037/0022-0663.79.3.308](https://doi.org/10.1037/0022-0663.79.3.308)
- Bavishi, A., Madera, J. M., & Hebl, M. R. (2010). The effect of professor ethnicity and gender on student evaluations: Judged before met. *Journal of Diversity in Higher Education*, 3, 245-256. [10.1037/a0020763](https://doi.org/10.1037/a0020763)
- Bennett, S. K. (1982). 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, 170-179. [10.1037/0022-0663.74.2.170](https://doi.org/10.1037/0022-0663.74.2.170)
- Biernat, M. (2003). Toward a broader view of social stereotyping. *American Psychologist*, 58, 1019-1027. [10.1037/0003-066X.58.12.1019](https://doi.org/10.1037/0003-066X.58.12.1019)
- Biernat, M., Fuegen, K., & Kobrynowicz, D. (2010). Shifting standards and the inference of incompetence: Effects of formal and informal evaluation tools. *Personality and Social Psychology Bulletin*, 36, 855-868. [10.1177/0146167210369483](https://doi.org/10.1177/0146167210369483)
- Boatright-Horowitz, S. L., & Soeung, S. (2009). Teaching White privilege to White students can mean saying good-bye to positive student evaluations. *American Psychologist*, 64, 574-575. [10.1037/a0016593](https://doi.org/10.1037/a0016593)
- Brady, K. L., & Eisler, R. M. (1999). Sex and gender in the college classroom: A quantitative analysis of faculty-student interactions and perceptions. *Journal of Educational Psychology*, 91, 127-145. [10.1037/0022-0663.91.1.127](https://doi.org/10.1037/0022-0663.91.1.127)
- Canada, K., & Pringle, R. (1995). The role of gender in college classroom interactions: A social context approach. *Sociology of Education*, 68, 161-186. [10.2307/2112683](https://doi.org/10.2307/2112683)
- Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *Journal of Higher Education*, 71, 17-33. [10.2307/2649280](https://doi.org/10.2307/2649280)
- Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, 51, 281-309.
- Cooper-Patrick, L., Gallo, J. J., Gonzales, J. J., Vu, H. T., Powe, N. R., & Nelson, C., et al. (1999). Race, gender, and partnership in the patient-physician relationship. *Journal of the American Medical Association*, 282, 583-589. [10.1001/JAMA.282.6.583](https://doi.org/10.1001/JAMA.282.6.583)
- Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *The American Economic Review*, 95, 158-165. [10.1257/000282805774670446](https://doi.org/10.1257/000282805774670446)
- Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, 2, 285-290. [10.1037/h0033731](https://doi.org/10.1037/h0033731)

- DuCette, J., & Kenney, J. (1982). Do grading standards affect student evaluations of teaching? Some new evidence on an old question. *Journal of Educational Psychology*, 74, 308-314. [10.1037/0022-0663.74.3.308](https://doi.org/10.1037/0022-0663.74.3.308)
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Bulletin*, 108, 233-256. [10.1037/0033-295X.109.3.573](https://doi.org/10.1037/0033-295X.109.3.573)
- Ehrenberg, R. G., Goldhaber, D.D., & Brewer, D.J. (1995). Do teachers' race, gender, and ethnicity matter? Evidence from the national educational longitudinal study of 1988. *Industrial and Labor Relations Review*, 48, 547-561. <http://www.nber.org/papers/w4669>
- Eizler, C. F. (2002). College students' evaluations of teaching and grade inflation. *Research in Higher Education*, 43, 483-501. [10.1023/A:1015579817194](https://doi.org/10.1023/A:1015579817194)
- 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, 151-211. [10.1007/BF00992161](https://doi.org/10.1007/BF00992161)
- Felton, J. (2008). Attractiveness, easiness and other issues: Student evaluations of professors on ratemyprofessors.com. *Assessment and Evaluation in Higher Education*, 33, 45 - 61. [10.1080/02602930601122803](https://doi.org/10.1080/02602930601122803)
- Foschi, M. (2000). Double standards for competence: Theory and research. *Annual Review of Sociology*, 26, 21-42. [10.1146/annurev.soc.26.1.21](https://doi.org/10.1146/annurev.soc.26.1.21)
- Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 52, 1209-1217. [10.1037/0003-066X.52.11.1209](https://doi.org/10.1037/0003-066X.52.11.1209)
- Griffin, K. A., & Reddick, R. J. (2011). Surveillance and sacrifice: Gender differences in the mentoring patterns of Black professors at predominantly White research universities. *American Educational Research Journal*, 48, 1032-1057. [10.3102/0002831211405025](https://doi.org/10.3102/0002831211405025)
- Hamermesh, D. S., & Parker, A. M. (2005). Beauty in the classroom: Professors' pulchritude and putative pedagogical productivity. *Economics of Education Review*, 24, 369-376. [10.1016/j.econedurev.2004.07.013](https://doi.org/10.1016/j.econedurev.2004.07.013)
- Harlow, R. (2003). "Race doesn't matter, but...": The effect of race on professors' experiences and emotion management in the undergraduate college classroom. *Social Psychology Quarterly*, 66, 348-363. [10.2307/1519834](https://doi.org/10.2307/1519834)
- Heckert, T. M., Latier, A., Ringwald-Burton, A., & Drazen, C. (2006). Relations among student effort, perceived class difficulty appropriateness, and student evaluations of teaching: Is it possible to "buy" better evaluations through lenient grading? *College Student Journal*, 40, 588-596.
- Ho, A. K., Thomsen, L., & Sidanius, J. (2009). Perceived academic competence and overall job evaluations: Students' evaluations of African American and European American professors. *Journal of Applied Social Psychology*, 39, 389-406. [10.1111/j.1559-1816.2008.00443.x](https://doi.org/10.1111/j.1559-1816.2008.00443.x)
- Kite, M. E., Stockdale, G. D., Whitley, B. E., & Johnson, B. T. (2005). Attitudes toward younger and older adults: An updated meta-analytic review. *Journal of Social Issues*, 61, 241-266. [10.1111/j.1540-4560.2005.00404.x](https://doi.org/10.1111/j.1540-4560.2005.00404.x)
- Krings, F., Sczesny, S., & Kluge, A. (2011). Stereotypical inferences as mediators of age discrimination: The role of competence and warmth. *British Journal of Management*, 22, 187-201. [10.1111/j.1467-8551.2010.00721.x](https://doi.org/10.1111/j.1467-8551.2010.00721.x)
- Langlois, J. H., Kalakanis, L., Rubenstein, A.J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, 126, 390-423. [10.1037/0033-2909.126.3.390](https://doi.org/10.1037/0033-2909.126.3.390)
- Levin, W. C. (1988). Age stereotyping: College student evaluations. *Research on Aging*, 10, 134-148. [10.1177/0164027588101007](https://doi.org/10.1177/0164027588101007)

- Littleford, L. N., Ong, K. S., Tseng, A., Milliken, J. C., & Humy, S. L. (2010). Perceptions of European American and African American instructors teaching race-focused courses. *Journal of Diversity in Higher Education*, 3, 230–244. [10.1037/a0020950](https://doi.org/10.1037/a0020950)
- Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology*, 92, 202-228. [10.1037/0022-0663.92.1.202](https://doi.org/10.1037/0022-0663.92.1.202)
- Millea, M., & Grimes, P. W. (2002). Grade expectations and student evaluation of teaching. *College Student Journal*, 36, 582-590. [0146-3934](https://doi.org/10.146-3934)
- Phelan, J. E., Moss-Racusin, C. A., & Rudman, L. A. (2008). Competent yet out in the cold: Shifting criteria for hiring reflect backlash toward agentic women. *Psychology of Women Quarterly*, 32, 406-413. [10.1111/j.1471-6402.2008.00454.x](https://doi.org/10.1111/j.1471-6402.2008.00454.x)
- Radmacher, S. A., & Martin, D. J. (2001). Identifying significant predictors of student evaluations of faculty through hierarchical regression analysis. *Journal of Psychology*, 135, 259-268. [10.1080/00223980109603696](https://doi.org/10.1080/00223980109603696)
- Reid, L. 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, 137-153. [10.1037/a0019865](https://doi.org/10.1037/a0019865)
- Riniolo, T. D., Johnson, K. C., Sherman, T. R., & Misso, J. A. (2006). Hot or not: Do professors perceived as physically attractive receive higher student evaluations? *Journal of General Psychology*, 133, 19–35. [16475667](https://doi.org/10.16475667)
- Sinclair, L., & Kunda, Z. (2000). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. *Personality & Social Psychology Bulletin*, 26, 1329-1342. [10.1177/0146167200263002](https://doi.org/10.1177/0146167200263002)
- Smith, B. P. (2007). Student ratings of teaching effectiveness: An analysis of end of course faculty evaluations. *College Student Journal*, 41, 788-800. [0146-3934](https://doi.org/10.146-3934)
- Sprague, J., & Massoni, K. (2005). Student evaluations and gendered expectations: What we can't count can hurt us. *Sex Roles*, 53, 779-793. [10.1007/s11199-005-8292-4](https://doi.org/10.1007/s11199-005-8292-4)
- Statham, A., Richardson, L., & Cook, J. (1991). *Gender and university teaching: A negotiated difference*. Albany, NY: State University of New York Press.
- Stolte, J. F. (1996). Evaluation of persons of varying ages. *Journal of Social Psychology*, 136, 305-309. [10.1080/00224545.1996.9714009](https://doi.org/10.1080/00224545.1996.9714009)
- Verkuyten, M. (2007). Ethnic in-group favoritism among minority and majority groups: Testing the self-esteem hypothesis among preadolescents. *Journal of Applied Social Psychology*, 37, 486-500. [10.1111/j.1559-1816.2007.00170.x](https://doi.org/10.1111/j.1559-1816.2007.00170.x)
- Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in Higher Education*, 12, 55-76. [10.1080/13562510601102131](https://doi.org/10.1080/13562510601102131)

Contact Information

Susan Basow may be contacted at basows@lafayette.edu