

Research Report: Race and Gender Bias in Student Evaluations of Teaching

Compiled by **Therese Huston, Ph.D.**

*Director of the Center for Excellence in Teaching & Learning
Seattle University*

October 31, 2005

Updated with additional references

Introduction

This report was prompted by a faculty member's request to summarize the research on the impact of race and gender in the evaluation of university teaching.

Unfortunately, there is relatively little in the way of empirical, quantitative research on the intersection of race and gender as they pertain to course evaluations. Numerous studies have examined the possibility of a gender bias in student evaluations of teaching and small-scale studies tend to find very different results from large-scale studies. The possibility of a race bias is a relatively new area of inquiry, however, as evidenced by the fact that several of the relevant papers are published in 2005. The silver lining, of course, is that the findings are recent and are likely to reflect current institutional trends.

This document summarizes the key empirical findings. The report is intended for information purposes only, and is not intended to provide guidance toward a particular observation, conclusion, or policy decision. Further information about most of the studies listed here can be obtained through CETL.

This report may be updated periodically as new information becomes available.

Findings by sub-topic

INTERACTION OF RACE & GENDER

- **Researchers found lower final course evaluation ratings for female minority faculty members, but not for male minority instructors.** In their study, women instructors received significantly lower course evaluations than male instructors (nearly ½ standard deviation lower), and faculty of color received lower course evaluations than white faculty. There was also an interaction of race and gender such that female faculty of color received particularly low course evaluations.

Source: Hamermesh & Parker (2005)

- **The same study found that while there was a main effect of language such that non-native English speakers had significantly lower course evaluations overall, male instructors who were non-native English speakers were judged more harshly on course evaluations than female instructors who were non-native English speakers,** revealing a different pattern of interaction between gender and language.

Source: Hamermesh & Parker (2005)

- **Two recent studies have shown that faculty of color receive either the most favorable or the least favorable ratings when compared to Anglo faculty** (Anderson & Smith, 2005; Smith & Anderson, 2005). Both studies only compared Latino/a and Anglo faculty and did not include African-American instructors. In one study of a politically charged social science course, there was an interaction between ethnicity, gender, and teaching style (Anderson & Smith, 2005). Female Latina professors who had a lenient teaching style were viewed as more

warm than female Anglo professors with a lenient teaching style. Likewise, female Latina professors who had a strict teaching style were viewed as more strict than their Anglo female peers with a comparable style. In summary, perceptions of warmth and strictness were magnified for female faculty of color.

Sources: Anderson & Smith (2005) and Smith & Anderson (2005)

(The summary of these last two articles was based on abstracts for these articles– the full articles have been requested through Inter-Library Loan.)

RACE

- Two studies have found a main effect of race such that **faculty of color received lower course evaluations than their white peers** (Hamermesh & Parker, 2005; DiPietro & Faye, 2005). In a study of one university, DiPietro and Faye (2005) found that Hispanic faculty received the lowest course evaluation ratings. Asian-American faculty received slight better course evaluations than their Hispanic colleagues, but their scores were, on average, still worse than the scores of White faculty. The number of African-American faculty in their sample was too small to draw any conclusions.

Source: DiPietro & Faye (2005) and Hamermesh & Parker (2005)

- Researchers have looked at other measures besides course evaluations to gauge racial bias in the classroom. For example:
 - **Using a survey separate from the final course evaluation, Rubin (1998) found that students rate Asian-American instructors as less credible and intelligible than white instructors.**
 - In a series of semi-structured interviews, Hendrix (1998) found that students in a predominantly White university did not believe that a professor's race influenced their perceptions of that instructor's credibility, yet the students simultaneously described a different set of criteria for evaluating the credibility of their Black instructors for courses on certain topics (relative to the criteria applied to their White instructors).
 - **Students' comments revealed that Black instructors had more credibility when they taught courses that had an ethnic or racial focus, and students reported that they would more readily question and challenge the credibility of Black instructors for courses that lacked an ethnic / racial component to them.**

Sources: Hendrix (1998) and Rubin (1998)

- There is also literature that demonstrates a feedback bias for minorities, but these studies have been done outside of instructional settings. One of the interesting dimensions of this research is that the direction of the bias depends on whether the feedback is perceived as going directly to a person of color or whether the feedback is perceived as going to some other third party.
 - **When Whites rate the performance of a person of color with the understanding that their judgments would be communicated to a third party for the purposes of evaluation, Whites consistently rate performance negatively** (e.g. Henderson-King & Nisbett, 1997; Lambert, Cronen, Chasteen & Lickel, 1996).
 - **When Whites rate the performance of a person of color for the purposes of giving feedback to that person directly, however, the Black person being evaluated actually receives significantly more favorable marks than the White person being evaluated** (Harber, 1998).
 - In light of this research, it may be helpful to consider whether students view their evaluations for administrative, summative evaluation purposes or for the instructor to improve his or her teaching.

Sources: Harber (1998); Henderson-King & Nisbett (1997); and Lambert, Cronen, Chasteen & Lickel (1996)

GENDER

- The issue of gender bias in teaching evaluations is a huge area of controversy in the literature. Some authors, such as Hamermesh & Parker (2005; see also Anderson & Miller, 1997; Helgeson, 1994; Miller & Chamberlin, 2000), have found that female faculty members receive lower course evaluation ratings, but other authors report that male faculty members receive lower evaluation scores (Costa, Terracciano, & McCrae, 2001; Kierstead, D'Agostino, & Dill, 1988). Small scale studies that only examine one discipline are more likely to find a gender bias in course evaluations (Santhanam & Hicks, 2002). **However, in the most comprehensive studies that carefully control for other biases and that compare student evaluations across multiple disciplines, researchers have not found a statistically significant difference in the average course evaluations for male and instructors** (Centra & Gaubatz, 2000; Feldman, 1993; Franklin & Theall, 1992; Santhanam & Hicks, 2002; Seldin, 1999). Female faculty, for example, are more likely to teach large, required courses than male faculty (DiPietro & Faye, 2005). Not surprisingly, large, required courses receive lower course evaluation scores than small, elective courses (Petchers & Chow, 1988; DiPietro & Faye, 2005). When such confounds are controlled, female course evaluation scores are not significantly different from those of male faculty.
- There are, however, certain findings that are consistent across multiple studies of gender bias, and perhaps most importantly, this subset of findings is consistent across studies that differ on other empirical findings. In other words, even when researchers disagree about the big question of an overall gender bias in the evaluation of teaching, they often agree about certain issues.
 - **There are same-gender preferences, such that female students tend to give higher ratings to female instructors** (Basow & Silberg, 1987; Feldman, 1993; Centra & Gaubatz, 2000). It is less clear whether male students also give higher ratings to male instructors – some evidence supports this claim but other empirical studies challenge it (Basow, 1995; Basow & Silberg, 1987; Centra & Gaubatz, 2000; Hicks & Santhanam, 2002). Same-gender preferences have been attributed to theories of gender differences in “ways of knowing,” such that both women instructors and women students prefer teaching methods that emphasize “connected classrooms” (Belenky, Clinchy, Goldberger, & Tarule, 1986) that value “understanding and acceptance over evaluation and collaboration over debate” (Centra & Gaubatz, p. 31, 2000).
 - Students have different expectations for how warm and friendly male and female faculty will be (Linse, 2003). **Female faculty who are rated as being low on warmth, interpersonal contact, and interest in their students were penalized on their course evaluations (relative to other female faculty) whereas male faculty who were rated as being low on these same interpersonal qualities did not receive lower course evaluations** (Kierstead, D'Agostino, & Dill, 1988; Bennett, 1982; Downs & Downs, 1993).
- Several studies have found disciplinary differences in how students evaluate male and female faculty. These gender biases seem to appear in fields that are traditionally dominated by men (i.e. science, engineering, etc.). Some studies have found that in science and engineering classrooms, which are traditionally male dominated, both males and female students give higher ratings to female instructors (Centra & Gaubatz, 2000; Hicks & Santhanam, 2002). Other studies, however, report lower ratings for women faculty in the fields of math, science, and engineering (Basow, 1992; Street, Kimmel, & Kromrey, 1996). Such disciplinary differences in gender bias warrant further investigation.
- The question still remains as to why there could still be such dramatic differences in the research findings, such that one steady line of research has female faculty at a disadvantage

and another line has male faculty at a disadvantage. Not surprisingly, the underlying issue might concern the types of questions that are asked on the student evaluations.*

- **If some of the questions on the course evaluation focus on warmth, such as the instructor's ability to understand or respond to students' needs (qualities usually associated with feminine expressiveness) then female instructors have higher scores** (Costa, Terracciano, & McCrae, 2001; Kierstead et al., 1988).
 - **If however, some of the questions ask students to rate instructors' enthusiasm or eagerness in the classroom (qualities usually associated with masculine expressiveness), then male instructors have higher scores** (Helgeson, 1994; Mulac & Lundell, 1982).
- Hamermesh & Parker (2005) also found an interaction between gender and course level (i.e. whether the course being evaluated was an upper or lower division course). Female faculty teaching upper-division courses received course evaluations that were about average for the sample, but female faculty teaching lower-division courses received course evaluations that were far below average.

***Note:** Because the course evaluation forms used by separate colleges within a University differ, it may be valuable to take a look at the specific questions on each evaluation form. If one of the questions suggests that masculine (or feminine) qualities are valued in teaching a particular discipline, then such a question could create what is known as a "halo effect." A halo effect occurs when a person fails to discriminate between distinct aspects of a person's behavior, such that one question on a course evaluations contaminates nearby questions and leads to an artificial increase or decrease in the ratings (Feeley, 2002).

-
- Sources: *Belenky, Clinchy, Goldberger, & Tarule (1986); Centra & Gaubatz (2000); Costa, Terracciano, & McCrae (2001); D'Agostino, & Dill (1988); (Feeley, 2002); Feldman (1993); Helgeson (1994); Kierstead, Mulac & Lundell (1982); Hamermesh & Parker (2005); Basow & Silberg, (1987); Hicks & Santhanam, (2002); Basow, (1995); Anderson & Miller, (1997); Miller & Chamberlin, (2000); Kierstead, D'Agostino, & Dill, (1988); Bennett (1982); Downs & Downs, (1993); Basow (1992); Street, Kimmel, & Kromrey, (1996); Seldin (1999)*

Summary of Key Findings

The bulleted list below summarizes the key findings on race and gender bias in student evaluations of teaching more succinctly.

- There are very few studies that examine the combined issues of race and gender in the evaluation of teaching. One study indicates that relative to their colleagues, female faculty of color receive lower course evaluations, and other studies indicate that White students are more likely to perceive faculty of color in extremely positive or extremely negative terms.
- Likewise, few studies have investigated the main effect of an instructor's race in course evaluations, but two recent studies found that faculty of color received significantly lower ratings on average than White faculty.
- Students are able to simultaneously deny any impact of race on judgments of teaching and articulate the different criteria they use to judge Black and White faculty.
- In non-instructional settings, Whites are biased to evaluate the performance of a person of color more positively if they think they are providing direct feedback, but they are also biased to evaluate performance more negatively if they think the evaluation is for a third-person's evaluation.

- Many researchers have examined the issue of gender bias in teaching evaluations without controlling for race, the ratio of male to female students evaluating the instructor, or a number of other factors. Although individual studies can be found that support a gender bias favoring either male or female faculty, a reasonable conclusion based on the most comprehensive and carefully controlled studies is that there is no general gender bias in course evaluations. There is consistent evidence, however, for a same-gender preference, such that female students prefer female professors. There is also consistent evidence that to receive high course evaluations, students require female faculty to demonstrate more warmth and interpersonal contact than they require of male faculty.

Limitations of the Research

There are conceivably many limitations to the research on race and gender bias in evaluations of teaching, particularly given the small collection of empirical studies that examine the issue of race. A few of these limitations are listed below.

- The issue of cognitive bias (either conscious or unconscious bias) has not been applied or addressed in these studies.
- Although one study shows that students confer more credibility to faculty of color who teach courses related to race and ethnicity, it is not clear which academic disciplines, if any, are subject to greater racial bias. Given that there is a gender bias within the sciences, there may be comparable racial biases in certain fields.
- If students provide different types of feedback depending upon the audience that they envision when they write the feedback, it would be important to know whether White students envision a different audience when they evaluate faculty of color compared to when they evaluate white faculty. Students may perceive that they are communicating with the instructor when they evaluate someone of their race, but they may perceive they are communicating with the administration when they evaluate someone of another race.
- Students' perceptions and evaluations of faculty are also affected by physical factors such as age and formality of dress. These issues have not been taken into consideration within the research on racial bias in course evaluations.

References

- Anderson, K., & Miller, E. D. (1997). Gender and student evaluations of teaching. *PS: Political Science and Politics*, 30 (2), 216-219.
- Anderson, K. J., & Smith, G. (2005). Students' preconceptions of professors: Benefits and barriers according to ethnicity and gender. *Hispanic Journal of Behavioral Sciences*, 27, (2), 184-201.
- Arbuckle, J., & Williams, B. D. (2003). Students' perceptions of expressiveness: Age and gender effects on teaching evaluations. *Sex Roles*, 49, 507-516.
- Basow, S. A. (1994). Student ratings of professors are not gender blind. *AWM Newsletter*, 24(5), Sept.-Oct. 1994. <http://www.awm-math.org/newsletter/199409/basow.html>.
- Basow, S. A. (1995). "Student evaluations of college professors: When gender matters." *Journal of Educational Psychology*, 87(4), 656-665.
- Basow, S. A., and Silberg, N. T. (1987). "Student evaluation of college professors: Are female and male professors rated differently?" *Journal of Educational Psychology*. 79(3), 308-314.
- Belenky, M. F., Clinchy, B. M., Goldberger, N. R., & Tarule, J. M. (1986). *Women's Ways of Knowing*. New York: Basic Books.
- Bennett, S. K. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. *Journal of Educational Psychology*, 74 (2): 170-179.
- Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *Journal of Higher Education*, 71, p. 17.
- Costa, P., Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Cognition*, 81, 322-331.
- DiPietro, M. & Faye, A. (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.
- Downs, V. C., & Downs, T. M. (1993) An exploratory and descriptive study identifying communicative behaviors associated with effective college teaching. Paper presented at the annual meeting of the International Communication Association. May 1993, Washington, D.C.
- Feeley, T. H. (2002). Evidence of halo effects in students' evaluations of communication instruction. *Communication Education*, 51, 225-236.
- Franklin, J., & Theall, M. (1992). Student ratings of instructors and gender differences revisited. Paper presented at the 75th Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Feldman, K. A. (1993). College students' views of male and female faculty college teachers: Part II – Evidence from students' evaluations of their classroom teachers. *Research in Higher Education*, 34, 151-211.
- Hamermesh, D. S. & Parker, A. M. (Forthcoming, 2005). Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review*.
- Harber, K. D. (1998). Feedback to minorities: Evidence of a positive bias. *Journal of Personality and Social Psychology*, 74, (3), 622-628.

- Helgson, V. S. (1994). Prototypes and dimensions of masculinity and femininity. *Sex Roles, 31*, 653-682.
- Henderson-King, E. I., & Nisbett, R. E. (1997). Anti-Black prejudice as a function of exposure to the negative behavior of a single Black person. *Journal of Personality and Social Psychology, 71*, 654-664.
- Hendrix, K. G. (1998). Student perceptions of the influence of race on professor credibility. *Journal of Black Studies, 28*, 738-764.
- Kierstead, D., D'Agostino, P., & Dill, P. (1988). Sex role stereotyping of college professors: Bias in students' ratings of instructors. *Journal of Educational Psychology, 80*, 342-344.
- Lambert, A. J., Cronen, S., Chasteen, A. L., & Lickel, B. (1996). Private vs. public expressions of prejudice. *Journal of Experimental Social Psychology, 32*, 437-459.
- Linse, A. R. (2003, August). Student ratings of women faculty: Data and strategies. Retrieved Oct. 31, 2005 from University of Washington, Center for Engineering Learning and Teaching Web site: <http://depts.washington.edu/celtweb/>.
- Miller, J. & Chamberlin, M. (2000). Women are teachers, men are professors: A study of student perceptions. *Teaching Sociology, 28*, 283-298.
- Mulac, A., & Lundell, T. L. (1982). An empirical test of the gender-linked language effect in a public-speaking setting. *Language and Speech, 25*, 243-256.
- Petchers, M. K. & Chow, J. C. (1988). Sources of variation in students' evaluation of instruction in a graduate social work program. *Journal of Social Work Education, 24* (1), 35-42.
- Rubin, D. L. (1998). Help! My professor (or doctor or boss) doesn't talk English. In J. N. Martin, T. K. Nakayama, L. A. Flores (Eds.), *Readings in Cultural Contexts* (pp. 149 – 160). Mountain View, CA: Mayfield Publishing Company.
- Santhanam, E. & Hicks, O. (2002). Disciplinary, gender, and course year influences on student perceptions of teaching: Explorations and implications. *Teaching in Higher Education, 7* (1), 17-31.
- Seldin, P. and Associates (1999). *Changing Practices in Evaluating Teaching: A Practical Guide to Improved Faculty Performance and Promotion/Tenure Decisions*. Bolton, MA: Anker Publishing.
- Smith, G., & Anderson, K. J. (2005). Students' ratings of professors: The teaching style contingency for Latino/a professors. *Journal of Latinos and Education, 4*, (2), 115-136.
- Street, S., Kimmel, E., Kromrey, J. D. (1996). Gender role preferences and perceptions of university students, faculty and administrators. *Research in Higher Education, 37*: 615-632.

NOTE: We recommend the following reference for citations of this work:

Huston, T. (2005). Research report: Race and gender bias in student evaluations of teaching. Retrieved ***DATE*** from Seattle University, Center for Excellence in Teaching and Learning Web site: http://www.seattleu.edu/cetl/cetl_006.htm.