

The Role of Perceived Race and Gender in the Evaluation of College Teaching on RateMyProfessors.com

Landon D. Reid

Colgate University

The present study examined whether student evaluations of college teaching (SETs) reflected a bias predicated on the perceived race and gender of the instructor. Using anonymous, peer-generated evaluations of teaching obtained from RateMyProfessors.com, the present study examined SETs from 3,079 White; 142 Black; 238 Asian; 130 Latino; and 128 Other race faculty at the 25 highest ranked liberal arts colleges. Results showed that racial minority faculty, particularly Blacks and Asians, were evaluated more negatively than White faculty in terms of overall quality, helpfulness, and clarity, but were rated higher on easiness. A two-stage cluster analysis demonstrated that the very best instructors were likely to be White, whereas the very worst were more likely to be Black or Asian. Few effects of gender were observed, but several interactions emerged showing that Black male faculty were rated more negatively than other faculty. The results of the present study are consistent with the negative racial stereotypes of racial minorities and have implications for the tenure and promotion of racial minority faculty.

Keywords: student evaluations of teaching, race, gender, cluster analysis, RateMyProfessors.com

Research for more than half a century has examined the factors underlying student evaluations of college teaching. A growing body of research has shown that student evaluations of teaching (SETs) are influenced by the demographic characteristics of teachers (Arbuckle & Williams, 2003; Basow, 1990; Liddle, 1997). Remarkably little empirical research to date, however, has examined the effects of an instructor's race on SETs. As Beran and Violato (2005) noted, the omission of race from discussions of student evaluations of teaching is particularly problematic because student evaluations are the most commonly used metric for evaluating teaching in promotion and tenure cases (see also Basow, 1998; Marsh, 2007; McKeachie, 1997). As the professorate be-

comes more racially diverse, it becomes correspondingly more important to understand how faculty race combines with previously examined demographic characteristics of instructors to affect SETs (Beran & Violato, 2005; Williams, 2007). The present study, therefore, investigated the effects of an instructor's perceived race and gender on student evaluations of teaching as observed on RateMyProfessors.com (RMP).

The Evaluation of College Teaching

A large body of prior research articulated the factors associated with positive SETs. Professors viewed by students as more knowledgeable (Babad, Darley, & Kaplowitz, 1999), rational (Jenkins & Downs, 2001), helpful (Van Giffen, 1990), fair (Marlin & Gaynor, 1989), organized (Fortson & Brown, 1998), and clear (Ogier, 2005) are seen as quality instructors and receive more favorable SETs (see also, Barth, 2008). Further, professors perceived to be warmer, more expressive, and who show greater immediacy receive more positive SETs (Best & Addison, 2000; Kelley, 1950; Widmeyer & Loy, 1988; Wilson & Taylor, 2001). In addition, faculty who use more humor in their instruction are more likely to be favorably evalu-

Landon D. Reid, Department of Psychology, Colgate University.

I thank Rachel Goldfarb and Sue LeBarron and the members of the Racism, Experience and Perception Lab at Colgate University for their invaluable help. Thank you also to the Editor and anonymous reviewers for their comments on a draft of this article.

Correspondence concerning this article should be addressed to Landon D. Reid, Department of Psychology, Colgate University, 13 Oak Drive, Hamilton, NY 13346. E-mail: lreid@colgate.edu

ated on SETs (Fortson & Brown, 1998; Perry, Abrami, Leventhal & Check, 1979). Finally, research has shown that more physically attractive faculty also receive better SETs (Felton, Mitchell, & Stinson, 2004; Goebel & Cashen, 1985; but see also Campbell, Gerdes, & Steiner, 2005, for an opposing view).

Student evaluations of college teaching are subject to a number of systematic contaminants (Greenwald & Gilmore, 1997). One of the most significant factors impacting SETs is the grade that students expect to receive in a course. A variety of studies showed that students who expected to receive higher grades provided more favorable SETs (DuCette & Kenney, 1982; Millea & Grimes, 2002). This has been described as a reciprocity effect where students reward faculty for good grades and punish them for bad ones (Clayson, Frost, & Sheffet, 2006). Further, instructors who are perceived as easier, also receive more favorable SETs (Cashin, 1995; McKeachie, 1997).

Gender and Teaching Evaluations

Although some prior studies investigated faculty characteristics such as age (Arbuckle & Williams, 2003) and sexual orientation (Liddle, 1997), the most researched demographic characteristic in relation to SETs is gender (Heckert, Latier, Ringwald, & Silvey, 2006). Although some work found that women receive less favorable SETs than their male colleagues (Heckert et al., 2006; Tatro, 1995), other studies found no effect of faculty gender (Blackhart, Peruche, DeWall, & Joiner, 2006; Feldman, 1993; Liddle, 1997). Nevertheless, negative SETs were related to burnout for female faculty (Lackritz, 2004).

The inconsistency of the effect of gender on SETs has been explained as the result of gender interacting with other factors. For instance, prior studies found a gender symmetry effect such that male students rated male faculty more favorably and female students rated female faculty more favorably (Basow & Silberg, 1987; Martin, 1984). Other research found that SETs related to the congruity, or incongruity, between a faculty member's gender and the gender-stereotype of the academic discipline they teach (Basow, 1990, 1995; Bennett, 1982; Sprinkle, 2008). In other words, women receive less favorable SETs in traditionally masculine disci-

plines (e.g., Physics) and more favorable SETs in traditionally feminine disciplines (e.g., English).

Gender also imposes a different set standards on female faculty that affect their SETs. For example, whereas a male faculty member can demonstrate competence and be unfriendly toward students and still be considered intellectually competent, a female faculty member must demonstrate competence and friendliness to be judged as intellectually competent (Kierstead, D'Agostino, & Dill, 1988). Similarly, male faculty simply need to be perceived as helpful to get positive SETs, female faculty must be helpful and funny (Van Giffen, 1990).

Race and Teaching Evaluations

Racial minority faculty struggle to be seen as intellectually competent and credible in the classroom (Nast, 1999; Williams, 2007). Hendrix (1997, 1998) found that students of all races apply more stringent criteria for credentialing Black faculty as intellectually competent than they do White faculty. Similarly, Harlow (2003) found that students rarely challenged the academic credibility of White faculty in the classroom. Conversely, Black faculty, even those who believed that their race had no effect on how students treated them, reported having students regularly challenge their intellectual authority and academic competence in the classroom. A recent study by Ho, Thomsen, and Sidanius (2009) found that students' perceptions of intellectual competence were a bigger factor in the SETs of Black compared to White faculty. This was true for both Black and White students and high and low prejudice students.

Experimental studies of the effect of a faculty member's race on SETs yield contradictory results. In these studies, students evaluate the credentials and behavior of a fictitious faculty member whose race is changed across experimental conditions. Anderson and Smith (2005) found that students evaluated a fictitious Latino faculty member as having been less competent and warm than a White faculty member. Conversely, using a similar paradigm, Ludwig and Meacham (1997) found no evidence that students rated a fictitious racial minority professor differently than a White one.

The preponderance of studies utilizing actual SETs found that racial minority faculty are eval-

uated more negatively than White faculty (Boatright-Horowitz & Soeung, 2009; McPherson & Jewell, 2007; Smith, 2007). Chowdhary (1988) conducted a unique study examining the effects of race on SETs. In two different sections of the same course taught in the same semester, she either wore traditional Indian clothing or traditional Western clothing. She found that she received more negative SETs in the section where she wore traditional Indian clothing. Two studies, however, found no evidence of racial bias in the overall evaluation of Black versus White faculty (Ho, Thomsen, & Sidanius, 2009; Sidanius & Crane, 1989).

Faculty from different racial minority groups may be evaluated in different ways. The small populations of racial minorities at most colleges and universities mean that researchers may have to aggregate data for all racial minority faculty. This could obscure potentially meaningful group differences (Worthington, Navarro, Loewy, & Hart, 2008). Smith (2007) found that White faculty were rated more favorably than racial minority faculty, particularly Black faculty, on global evaluations of teaching quality. More specifically, she also found that the racial category of "Other" composed of Latino, Asian, and Native American faculty scored higher than Black faculty in all but one of the 25 categories examined. The lone rating where Black faculty scored higher than Other race faculty and close to the scores of White faculty pertained to the easiness of the course.

Race, Gender, and Teaching Evaluations

The interactive effects of race and gender on SETs are unclear. Although prior research indicates that racial minority women, particularly Black women, report being less satisfied in the professorate than racial minority men or Whites of either gender (Allen, Epps, Guillory, Suh, & Bonous-Hammarth, 2000), no research to date has examined both race and gender in relation to SETs. Moreover, race and gender have contradictory stereotypes (Landrine, Klonoff, Alcaraz, Scott, & Wilkins, 1995).

Consider the case of racial minority, male faculty. Stereotypically, men are considered more intellectually competent, a factor associated with favorable SETs (Basow, 1995, 2000). Conversely, racial minorities are stereotypically considered less intellectually competent (Banaji

& Greenwald, 1994). In addition to the generalized perception of academic incompetence, racial minority male faculty may have to contend with the additional burden of fear responses from students (Jackson & Crowley, 2003) who implicitly associate men of color with violence, hostility, and crime (Bodenhause & Lichtenstein, 1987; Devine, 1989).

The case of racial minority female faculty is also ambiguous. Women are generally perceived as higher on expressive characteristics such as warmth (Glick & Fiske, 1999; Kierstead, D'Agostino, & Dill, 1988; Swim & Cohen, 1997) that have been linked with favorable evaluations of teaching (Basow, 2000; Kelley, 1950). Conversely, the stereotypes of some women of color, Black women in particular, indicate that they may be seen as more angry or hostile than their White female peers (Landrine, 1999). Further, racial minority female faculty could also be subject to the combined burdens of racism and sexism (Evans & Cokely, 2008; Myers, 2005). Therefore, it remains unclear whether racial minority women would be evaluated more positively than their male peers.

The Present Study

The present study used the Website RMP to examine the effects of perceived faculty race and gender on student evaluations of teaching at 25 of the nation's leading liberal arts colleges. Relative to large research universities, selective liberal arts colleges demand both quality scholarship and exemplary teaching from faculty (Aries, 2008; Boyer, 1997; Ruscio, 1987). Consequently, there should be less variance in the quality of instruction both between and within institutions.

The previous research examining the effect of faculty race on SETs was limited by a number of factors. Because many institutions have very small populations of faculty of color (Allen et al., 2000; Turner & Myers, 2000), researchers must use data obtained from a single institution over time (e.g., Boatright-Horowitz & Soeung, 2009; Ho, Thomsen, & Sidanius, 2009; McPherson & Jewell, 2007; Sidanius & Crane, 1989; Smith, 2007) or aggregate ratings across all racial minority groups (Boatright-Horowitz & Soeung, 2009; McPherson & Jewell, 2007). The first approach, examining SETs collected over a number of semesters, produces a number

of confounds. It amplifies effects that may be because of a small set of individuals as well as effects that might be specific to a particular institution. The second approach, aggregating across racial minority groups, obscures potentially meaningful differences between racial groups (Worthington et al., 2008). Finally, the small sample sizes used in previous studies precluded the possibility of quantitative comparisons of SETs across both race and gender.

Why RMP?

Increasingly, students have come to use and rely on Websites that facilitate anonymous, peer-generated evaluations of teaching. The most popular of these Websites is RMP with over 10 million ratings of over one million faculty at more than 6,000 colleges and universities (<http://RateMyProfessors.com>). When deciding what courses to take, students rely heavily on peer recommendations based on the desirability, expertise, and credibility of prospective faculty (Beatty & Zahn, 1990; Felton, Mitchell, & Stinson, 2004; Harlow, 2003; Hendrix, 1997). Although students typically do not have access to institutionally administered evaluations of teaching, they strongly favor having SETs made publicly available (Howell & Symbaluk, 2001). Sites like RMP make SET data available to students.

A growing body of research used RMP to address a variety of questions related to SETs (Bowling, 2008; Kindred & Mohammed, 2005; Otto & Sanford, 2008; Riniolo, Johnson, Sherman, & Misso, 2006; Timmerman, 2008). Prior research finds that ratings made by students on RMP are consistent with end of term SETs (Silva et al., 2008; Timmerman, 2008) to the extent that some researchers have argued that they may be a useful supplement to traditional SETs (Otto & Sanford, 2008).

Research using RMP has a number of advantages. First, RMP provides a common metric for evaluating college teaching across both institutions and disciplines. Official evaluations of teaching are rarely seen outside of the institution where they are collected. As a result, it is difficult to evaluate how teaching is perceived across institutions. This addresses the issues that might be caused by studying one particular college or university. Second, RMP makes it possible to create a large enough sample of

racial minority faculty to quantitatively examine SETs in relation to both the race and gender of faculty.

The present study collected and evaluated ratings for every faculty member listed on RMP at the top 25 liberal arts colleges according to the 2006 U.S. News and World Report rankings (2005). Because the RMP site does not include demographic information on the faculty rated, the race and gender of faculty were added for the present study. RMP includes a global rating of overall instructor quality as well as ratings of instructor easiness, helpfulness, and clarity.

Predictions

The classroom experiences of faculty of color, and the preponderance of evidence on studies examining race and SETs, suggest that White faculty in the present study will be evaluated more favorably than racial minority faculty. The prior research, however, suggests that there will be no overall differences in the ways that female and male faculty are evaluated. It also remains unclear how race and gender interact with respect SETs. These predictions were tested in the present study.

Method

Data Set

Data in the present study were collected from the ratings of 5,630 faculty at the top 25 liberal arts colleges as listed in the 2006 edition of America's Top Colleges published by U.S. News and World Report (2005). The colleges examined are: Williams College (MA), Amherst College (MA), Swarthmore College (PA), Wellesley College (MA), Middlebury College (VT), Carleton College (MN), Bowdoin College (ME), Pomona College (CA), Haverford College (PA), Davidson College (NC), Wesleyan University (CT), Vassar College (NY), Claremont McKenna College (CA), Grinnell College (IA), Harvey Mudd College (CA), Colgate University (NY), Hamilton College (NY), Washington and Lee University (VA), Smith College (MA), Colby College (ME), Bryn Mawr College (PA), Oberlin College (OH), Bates College (ME), Macalester College (MN), and Mount Holyoke College (MA). Ratings

were obtained for every faculty member listed on the RMP Website for each of the colleges.

Professor Demographics

RMP does not list the race and gender of instructors. The method used in the present study to identify the race and gender of faculty approximates the kinds of guesses that students might make about the race and gender of faculty. Typically, faculties do not explicitly disclose their race and/or gender in the classroom. Students are therefore forced to make assumptions about those characteristics of their instructors based on appearance, first or last name, and the use of gendered pronouns.

To obtain information about the race and gender of faculty, a multiracial group of 12 undergraduate student coders used photographs obtained from publicly available sources (e.g., college and departmental Websites, professional meetings, Google image search) to evaluate each faculty member's race. The instructor's gender was determined by examining the professor's name, the gender of pronouns used by students in the qualitative comments section of each teacher's evaluation, and visual inspection of the instructor's photograph. All coders were enrolled as students at one of the institutions included in the study. Using this procedure, the present study was able to identify the races of 66.59% of the faculty. A random sample of 10% of the data taken from individuals where the race information was previously identified was recoded to obtain a reliability estimate. The level of interrater agreement for the race of faculty was 84%. Gender was identified for 99.14% of the sample. The level of interrater agreement for the gender of faculty was 97%.

The final sample included ratings of 3,717 faculty (1,493 female, 2,224 male) where race and gender were known. It was composed of ratings of 3,079 White (1,177 female, 1,902 male), 142 Black (61 female, 81 male), 238 Asian (137 female, 101 male), 130 Latino (60 female, 70 male), and 128 Other (58 female, 70 male) faculty. Faculty in the Other race category were Native American, Arab/Middle Eastern, and Bi/Multiracial.

Measures

RMP allows students to rate faculty on several dimensions that have previously been used to assess quality instruction. Students assess the overall quality of instruction (Barth, 2008; Marsh, 2007), easiness of the instructor (Cashin, 1995; Clayson, Frost, & Sheffet, 2006), clarity of the instructor (Fortson & Brown, 1998; Ogier, 2005), and helpfulness of the instructor (Babad, Darley, & Kaplowitz, 1999; Best & Addison, 2000; Wilson & Taylor, 2001) on a scale from 1 (Not at All) to 5 (Very Much).

Results

Analyses in the present study were only performed on data where information on faculty race and gender was available.

Preliminary Analyses

Table 1 shows the correlations and means of the focal variables for the entire sample. Overall Quality was strongly, positively correlated with perceptions of Helpfulness and Clarity. Overall Quality was also significantly positively related to perceptions of Easiness. Both Helpfulness and Clarity was also positively correlated with Easiness. Finally, Helpfulness and Clarity were strongly, positively correlated.

Race of Instructor

Preliminary analyses examined whether there were overall differences in SETs between White and Racial Minority faculty. As demonstrated in Table 2, Racial Minority faculty were rated significantly less favorably than White faculty on Overall Quality, Helpfulness, and Clarity.

Table 1
Correlations and Means of Overall Teaching Quality, Easiness, Clarity, and Helpfulness

Variable	1	2	3	4	Mean
1. Overall Quality	—				3.87 (.88)
2. Helpfulness	.96**	—			3.93 (.92)
3. Clarity	.96**	.84**	—		3.80 (.93)
4. Easiness	.15**	.18**	.11**	—	2.95 (.78)

Note. SDs in parentheses.
* $p < .05$. ** $p < .01$.

Table 2
Instructor Ratings for Racial Minority and White Faculty

Rating	Racial minority	White	<i>F</i> (1, 3550)
Overall Quality	3.72 (.94)	3.89 (.87)	17.64**
Helpfulness	3.81 (.98)	3.95 (.90)	11.03**
Clarity	3.64 (.99)	3.83 (.92)	20.89**
Easiness	3.03 (.81)	2.94 (.77)	7.11**

Note. *SDs* in parentheses.

* $p < .05$. ** $p < .01$.

Racial Minority faculty were, however, rated by students as easier than White faculty.

Table 3 shows ratings disaggregated by race. Upon closer inspection, many of the previously noted effects of faculty race were driven by disparities between Black faculty versus faculty from other racial groups. Across racial groups, means with different subscripts are significantly different, $p < .05$. For example, for Overall Quality, ratings of White faculty (subscript a) were different than those of Black faculty (subscript c), but not differently Latino or Other faculty (subscript ab). For ratings of Overall Quality, White faculty were not perceived differently than Latino or Other faculty, but were rated more favorably than Asian faculty who were rated more favorably than Black faculty. Similarly, faculty of all other races were rated as more Helpful than Black faculty. For ratings of Clarity, White faculty were not perceived differently than Latino or Other faculty but were rated more favorably than Asian faculty who were again rated more favorably than Black faculty. Finally, Black faculty were perceived to be significantly easier than Asian, Latino, White faculty, or Other faculty. In turn, Asian and Latino faculty were perceived to be easier than Other faculty.

Table 3
Instructor Ratings by Instructor Race

Rating	White	Black	Asian	Latino	Other	<i>F</i> (4, 3010)
Overall Quality	3.89 (.87) _a	3.48 (.99) _c	3.75 (.89) _b	3.87 (.89) _{a,b}	3.88 (.88) _{a,b}	8.32**
Helpfulness	3.95 (.90) _a	3.53 (1.03) _b	3.87 (.99) _a	3.97 (.89) _a	3.93 (.92) _a	7.31**
Clarity	3.83 (.92) _a	3.43 (1.04) _c	3.63 (1.00) _b	3.78 (.96) _{a,b}	3.84 (.86) _a	8.38**
Easiness	2.94 (.77) _{b,c}	3.18 (.81) _a	3.01 (.83) _b	3.07 (.75) _{a,b}	2.81 (.74) _c	5.42**

Note. Within a variable, racial group means with different subscripts differ significantly at $p < .05$. *SDs* in parentheses.

* $p < .05$. ** $p < .01$.

Two-step cluster analysis provides an exploratory method for identifying groups of faculty that may have been perceived in similar ways by students (for a review of this technique, see Punj & Stewart, 1983). This method is helpful for examining groupings across both continuous and categorical variables in large data sets. Two-step cluster analysis creates cluster groupings that are internally similar, but maximally dissimilar from the other clusters. In the present analysis, cluster membership was based on the continuous variables Overall Quality, Clarity, Easiness, Helpfulness, and the categorical variable Instructor Race.

The two-step cluster procedure yielded four clusters based on both Schwarz's Bayesian Information Criterion ($BIC = 7,856.50$) and the highest Log-likelihood distance measure (ratio of distance measures = 2.1). Figure 1 shows a plot of the cluster centroids on each of the continuous variables (Overall Quality, Clarity, Easiness, and Helpfulness). All differences between cluster centroids are significant, $p < .001$ across all variables other than Easiness, where Clusters 1 through 3 were not different from each other, but were all different from Cluster 4. In addition, Figure 1 also shows the percentage within each racial group represented in each of the clusters.

As shown in Figure 1, Cluster 1 ($n = 1624$) represents those professors regarded most favorably by students. These faculty were rated as best in Overall Quality, Helpfulness, and Clarity. This cluster was populated exclusively by White faculty. Cluster 2 ($n = 528$) represents most faculty of color. Although they were rated favorably in terms of Overall Quality, Helpfulness, and Clarity, they were rated less positively than the White faculty in the first cluster. Cluster 3 ($n = 1137$) appeared to represent subpar

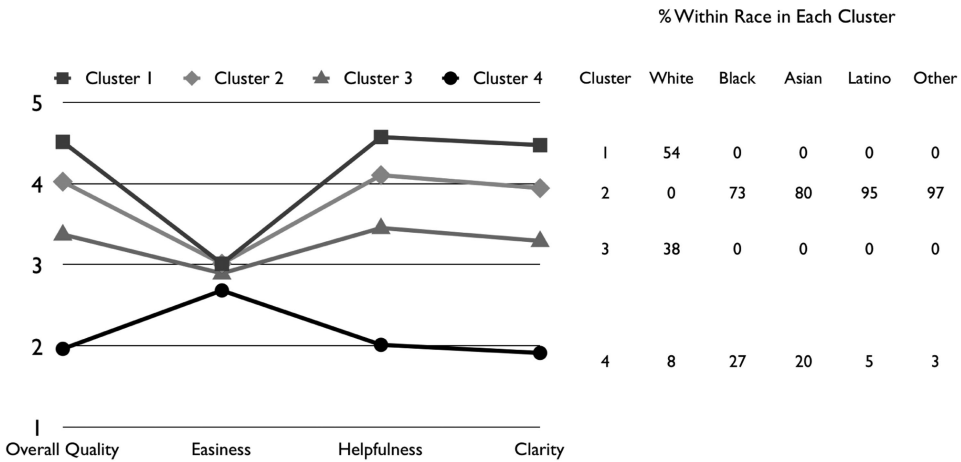


Figure 1. Plot of cluster centroids for each type of evaluation. The table shows the percentage of instructors of each race within each adjacent cluster.

White faculty. These faculty were rated neither particularly positively nor particularly negatively on any of the criteria. This cluster was also composed exclusively of White faculty. Of note is the finding that the centroids for the first three clusters were similar with respect to perceptions of Easiness. Despite differences in the perception of the other criteria, instructor Easiness was perceived in nearly identical ways across the first three clusters. Cluster 4 ($n = 319$) represents those faculty judged by students as the worst instructors. These faculty were rated much more negatively than the best instructors in Cluster 1, and significantly more negatively than the average instructors in Cluster 3. This was true for Overall Quality, Helpfulness, and Clarity. Of note is the finding that, in addition to being perceived more negatively on other dimensions, faculty in Cluster 4 were also perceived by students as more difficult than other instructors. Although faculty from every racial group were represented in Cluster 4, it included relatively few White, Latino, and Other race faculty. Cluster 4 did, however, include more than a quarter of the Black and one fifth of the Asian faculty.

Race and Gender

Instructor gender. First, the present study examined the overall effect of gender. ANOVAs were performed on the focal dependent

variables with gender as the between-subjects factor. As shown in Table 4, there was no effect of gender on evaluations of Overall Quality, Helpfulness, and Clarity. A gender difference was, however, observed for ratings of Easiness such that men were rated as easier than women.

Racial minority status and gender. A series of ANOVAs were performed to assess the possibility of an interaction between racial minority status (i.e., Racial Minorities vs. Whites) and gender. As shown in Figure 2, there are a number of interactions between racial minority status and gender. Interactions were observed between racial minority status and gender for Overall Quality, $F(1, 3550) = 3.89, p < .05$ and Clarity $F(1, 3550) = 3.88, p < .05$ such that although Racial Minority faculty were rated less favorably with respect to both Overall Quality and Clarity than White faculty, this effect was

Table 4
Instructor Ratings by Faculty Gender

Rating	Female	Male	$F(1, 3600)$
Overall Quality	3.86 (.90)	3.87 (.87)	1.82
Helpfulness	3.93 (.93)	3.93 (.90)	2.05
Clarity	3.78 (.95)	3.82 (.92)	1.14
Easiness	2.94 (.78)	2.96 (.77)	4.64*

Note. SDs in parentheses.
* $p < .05$. ** $p < .01$.

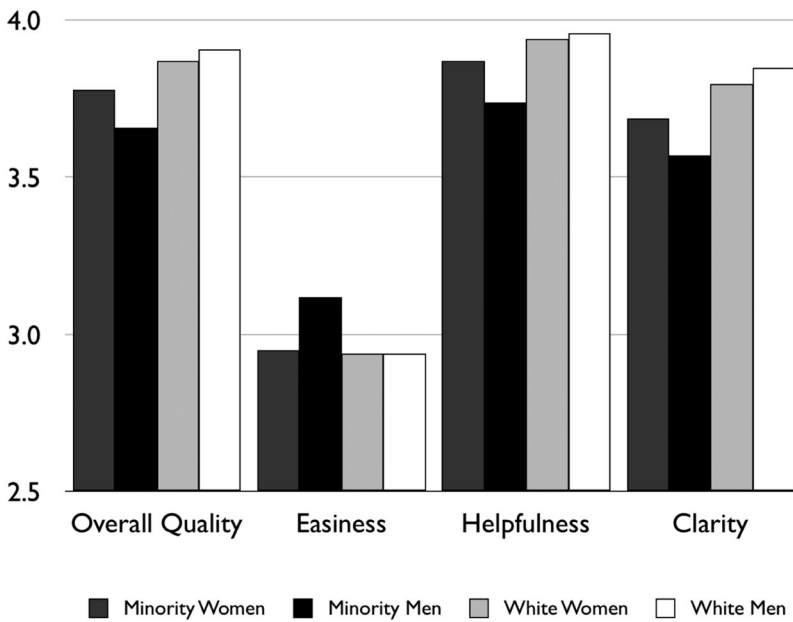


Figure 2. Mean ratings of teaching evaluations by minority and majority status and gender.

particularly pronounced for Racial Minority male faculty. Racial Minority male faculty were also rated as Easier than Racial Minority female faculty or White faculty of either gender, $F(1, 3550) = 4.93, p < .05$. This interaction was not significant for Helpfulness $F(1, 3550) = 2.99, ns$.

Race and gender. Table 5 shows student ratings disaggregated by race and gender. This analysis compares female versus male faculty of

each race on each of the focal dependent variables. Within a particular race and dependent variable (e.g., Overall Quality), genders with different subscripts are significantly different ($p < .05$). Generally, there were no gender differences within racial groups (all F s $< 2.44, ns$). This was true for Whites, Asians, and Latinos. Gender differences were, though, observed for Black faculty. Although Black male faculty were considered easier than Black female fac-

Table 5
Instructor Ratings by Instructor Race and Gender

Rating, gender	White	Black	Asian	Latino	Other
Overall Quality					
Women	3.87 (.89) _a	3.67 (1.06) _a	3.76 (.93) _a	3.92 (.87) _a	3.87 (.87) _a
Men	3.91 (.85) _a	3.35 (.91) _a	3.72 (.96) _a	3.84 (.92) _a	3.89 (.83) _a
Helpfulness					
Women	3.94 (.93) _a	3.69 (1.12) _b	3.89 (.95) _a	4.05 (.86) _a	3.93 (.93) _a
Men	3.96 (.89) _a	3.41 (.95) _a	3.85 (1.03) _a	3.90 (.89) _a	3.94 (.91) _a
Clarity					
Women	3.80 (.94) _a	3.64 (1.06) _a	3.64 (.98) _a	3.80 (.96) _a	3.81 (.88) _a
Men	3.85 (.90) _a	3.28 (1.00) _b	3.60 (1.05) _a	3.76 (.98) _a	3.86 (.85) _a
Easiness					
Women	2.93 (.77) _a	2.99 (.79) _b	2.99 (.80) _a	3.11 (.72) _a	2.67 (.74) _b
Men	2.94 (.77) _a	3.32 (.81) _a	3.01 (.87) _a	3.04 (.77) _a	2.93 (.73) _a

Note. Within racial groups, within variables, genders with different subscripts differ significantly at $p < .05$. SDs in parentheses.

ulty $F(1, 136) = 5.78, p < .05$, Black female faculty were perceived as more clear than Black male faculty $F(1, 137) = 4.06, p < .05$. In addition, Other male faculty were rated as easier than Other female faculty, $F(1, 120) = 3.89, p < .05$.

Easiness and Overall Quality

The present results indicate that, contrary to much of the existing literature, ratings of an instructors Easiness may not necessarily be related to ratings of their Overall Quality. For example, Black faculty were rated as Easier than White faculty, but White faculty are rated higher on Overall Quality. To examine this possibility, correlations between Overall Quality and Easiness were performed separately by faculty race and gender. Table 6 shows that although Overall Quality and Easiness are positively related for both White and Latino faculty, they are unrelated for Black, Asian, and Other faculty.

Discussion

The present study examined whether student evaluations of college teaching at selective liberal arts schools reflected biases based on the perceived race and gender of faculty. Using anonymous, peer-generated evaluations of teaching obtained from RMP, the present study found support for the idea that racial minority faculty, particularly Black faculty, were evaluated more negatively than White faculty in terms of Overall Quality, Helpfulness, and Clarity, but were rated higher in Easiness. Although there were few overall effects of gender, several interactions between faculty race and gender were observed such that Black male faculty

were rated more negatively than others. Finally, whereas student perceptions of an instructor's Easiness were related to their perceptions of Overall Quality for White and Latino faculty, contrary to the majority of previous studies, this relation was not observed for Black, Asian, and Other race faculty. The results of the present study suggest that both race and gender have an interactive effect on SETs that should be considered in the tenure and promotion cases of racial minority faculty.

Race and the Evaluation of College Teaching

Students evaluated racial minority faculty more negatively than White faculty across a variety of dimensions. Consistent with both experimental studies examining the effect of race on teaching evaluations (Anderson & Smith, 2005) and research on actual SETs (Boatright-Horowitz & Soeung, 2009; McPherson & Jewell, 2007; Smith, 2007), the present study found that racial minority faculty were rated more poorly than Whites overall. This effect, though, was not limited to overall ratings of the quality of instruction. Racial minority faculty were also rated lower on Helpfulness and Clarity, factors related to students' overall perceptions of instruction (Babad, Darley, & Kaplowitz, 1999; Fortson & Brown, 1998; Ogier, 2005). The finding that racial minority faculty were rated lower on all of these factors suggests that student perceptions may have been influenced by systemic biases like prejudice or racial stereotyping.

The present findings suggest that racial stereotypes and the continued existence of racism also affect racial minority faculty SETs. Racial minority faculty may represent a double-violation of stereotype-based expectancies.

The first violation is that faculty of color deviate from the stereotypical expectation that professors are bearded, bespectacled, White men (Messner, 2000). This violation of stereotype-based expectancies may create psychological discomfort (Lepore & Brown, 1997; Macrae & Bodenhausen, 2000). This discomfort could then be associated with racial minority faculty members in ways that could negatively affect student perceptions of teaching.

The second violation is related to what some racial minority faculty are. The mere presence

Table 6
Correlations Between Overall Quality and Easiness by Race and Gender

Race	Group	Women	Men
Whites	.17**	.19**	.17**
Blacks	-.02	.00	.03
Asians	.04	.07	-.01
Latinos	.31**	.32**	.31**
Others	.05	.19	-.07

* $p < .05$. ** $p < .01$.

of a racial minority professor in the classroom is sufficient to activate the negative racial stereotypes directly implicated in the perception of quality instruction like intellectual competence (Brigham, 1993; Devine, 1989; Dovidio & Gaertner, 2004; Greenwald & Banaji, 1994; Steele, 1997) because race is one of the dimensions that humans use to instantly, automatically categorize others (Lepore & Brown, 1997; Zarate & Smith, 1990). The feelings of hostility and threat evoked by the negative stereotypes of Blacks and Latinos (Bodenhausen & Lichtenstein, 1987; Devine, 1989) help explain how students can describe being physically afraid in the classrooms of Black male professors (Jackson & Crawley, 2003; Maddox & Gray, 2002). Further, faculty of color make up a very small proportion of the professorate (Wilds, 2000). Therefore, when students encounter a racial minority professor, it may be the only one that they have encountered during college. They may be correspondingly more likely to rely on racial stereotypes to evaluate racial minority faculty (Macrae & Bodenhausen, 2000). The rapid activation of negative racial stereotypes is particularly problematic because first impressions of a faculty member predict their end of term evaluations (Babad, Avni-Babad, & Rosenthal, 2004; Buchert, Laws, Apperson, & Bregman, 2008).

Faculty of color were not perceived by students to be among the very best teachers. Although they could certainly be evaluated in positive terms, racial minority instructors were not rated in the same cluster as the most highly regarded White faculty. Racial minority faculty could be perceived as good, but not great. At the same time, racial minority faculty appeared to be overrepresented among those faculty judged most negatively. Of particular note is the result that instructors in the poor teaching category were also perceived as more difficult than other instructors. Contemporary research suggests that students are unlikely to assert that a racial minority faculty member is a bad instructor because of their race (Bonilla-Silva, 2006; Dovidio & Gaertner, 2004). Instead, prejudicial biases are more likely to be expressed as principled, and therefore socially defensible, evaluations of an instructor's teaching.

Racial stereotypes also support the finding that Whites were rated by students as the very best teachers. The racial stereotype of Whites is

that they are knowledgeable, intellectually competent (Brigham, 1993; Katz & Braly, 1933) and rational (Kincheloe & Steinberg, 1998), all characteristics associated with quality teaching (Babad, Darley, & Kaplowitz, 1999; Jenkins & Downs, 2001). Based on racial stereotypes alone, White faculty may have an advantage over racial minority faculty on SETs (Messner, 2000).

One of the important findings of the present study was that the effects of race were not the same across all racial minority groups. At the group level, Latino and Other Race faculty were not perceived in significantly different ways than White faculty. It remains unclear why this was the case. These groups were also perceived differently from Asian and Black faculty. Research that aggregates across these differences would have obscured these findings (Sue, 2004; Worthington et al., 2008).

The finding that Black and Asian faculty were evaluated in similar ways is counterintuitive considering their respective racial stereotypes. Racial stereotypes suggest that Blacks are not academically competent (Banaji & Greenwald, 1994; Devine, 1989) whereas Asians are academically competent (Lin, Kwan, Cheung, & Fiske, 2005; Sue, Sue, & Sue, 1975). Accordingly, Asians would be expected to excel as college instructors relative to Blacks. It is therefore surprising that Blacks and Asians would be perceived in similar ways. Although stereotypically dissimilar, Blacks and Asians may represent the groups most easily identifiable as racial minorities. It is possible that instructors who are the most visually distinct from Whites are evaluated most negatively. Although inferences about an instructor's race may be made from a surname (e.g., Chan) or departmental affiliation (e.g., African American Studies), it is likely that most information about an instructor's race comes from visual identification (see Macrae & Bodenhausen, 2000). Previous research has demonstrated that the more phenotypically representative an individual is of her or his racial group, the more likely they will be characterized according to the negative stereotypes of that group (Dixon & Maddox, 2005; Maddox, 2004). As a result, individuals who are more easily recognized as racial minorities may be correspondingly more likely to bear the burden of their group's most negative stereotypes. For darker complexioned, more readily identifiable,

Blacks this means being stereotypically labeled as intellectually incompetent. For phenotypically featured Asians, this could mean being stereotypically labeled as poor speakers of English.

It is also possible that Asian faculty were negatively evaluated because they were disproportionately perceived as non-native English speakers. For example, an analysis of the Asian faculty at one of the institutions included in the present study revealed that more than half were non-native English speakers. In the present study, students' perceptions of the instructor's clarity were strongly related to their overall evaluations. To the extent that faculty were, or were perceived to be, non-native English speakers, they would likely be perceived as having been less clear. This idea is supported by a study by Ogier (2005) who found that instructors for whom English was not their first language were evaluated more negatively by students.

One of the most striking findings in the present study is that student perceptions of Overall Quality were independent of their perceptions of an instructor's Easiness for most racial minority faculty. This is unexpected because the positive relationship between perceived easiness and overall evaluation is so pervasive that it is considered a universal contaminant in all SETs (Cashin, 1995; Greenwald & Gilmore, 1997; McKeachie, 1997; but cf. Smith, 2007). Racial minority faculty, particularly Black and Asian faculty, were perceived as easier overall than White faculty. At the same time, the grouping of faculty judged by students to be the worst instructors (Cluster 4) that contained a substantial percentage of the Black and Asian faculty, was judged as more difficult.

It is possible that students may perceive Black and Asian faculty in two contradictory ways. Students may perceive these faculty as easier because, in stereotypical terms, they are supposed to be less intellectually rigorous than Whites. Alternatively, some racial minority faculty may use easy grading as a strategic tactic to mitigate the effects of being perceived as hostile or threatening in the classroom. If, however, faculty of color uphold a rigorous grading standard, they may be correspondingly more likely to be punished by students expecting an easy class (e.g., Clayson, Frost, & Sheffet, 2006). This could be exacerbated by the idea that racial

minority faculty are less likely to be automatically granted intellectual deference and academic credibility (Harlow, 2003; Hendrix, 1998). Instructor race appears to effectively serve as a boundary condition for an effect previously assumed to be ubiquitous.

Race, Gender, and SETs

Race and gender interact to affect SETs. Consistent with previous research, the present study found no main effects of gender (Basow, 2000; Basow & Silberg, 1987; Kierstead, D'Agostino, & Dill, 1988). We also found few interactions between an instructor's gender and race. Those interactions that were observed were primarily related to differences between Black male versus female faculty. Generally, Black men were rated more negatively by students than all other faculty. This suggests that Black men were not necessarily benefiting from gender stereotypes of men that suggest that they are more academically inclined than women (Basow, 1995). Correspondingly, the results of the present study suggest that racial minority female faculty may not necessarily be doubly punished for students for being both female and racial minority. This possibility is, however, more difficult to assess given the complex and interactive way that gender affects SETs (Basow & Silberg, 1987; Martin, 1984).

Limitations and Future Directions

The current research is subject to a number of important caveats. First, the present study aggregates data across a number of demographic and instructor characteristics. For example, the study does not include information about the instructor's department or teaching style. Both of these factors have been strongly associated with SETs (Basow, 2000; McKeachie, 1997). In addition, the anonymous nature of RMP makes it impossible to know anything about the race or gender of the student raters, factors that could affect evaluations (Basow & Silberg, 1987; Martin, 1984). Future research should examine how these factors affect the evaluation of racial minority versus White faculty.

Second, data were collected from a Website where anyone can anonymously post information about instructors. Consequently, it is impossible to verify whether raters were enrolled

in the classes of the faculty rated or were even students at these institutions. This issue is inherent in the anonymous nature of the Website. It is precisely this characteristic, however, that makes RMP attractive. Students consult RMP for guidance when choosing classes because it ostensibly provides direct and honest feedback about instructors. Moreover, prior research found that ratings on RMP are consistent with institutionally administered SETs (Silva et al., 2008; Timmerman, 2008).

Third, the exploratory nature of the present study made it difficult to explain the findings for every racial group. For example, it remains unclear why Latino faculty, who have many of the same negative stereotypes as Black faculty and linguistic stereotypes as Asian faculty (Bodena-hausen & Lichtenstein, 1987) were perceived positively. Future research should more carefully examine why students perceive the members of different racial minority groups differently.

The present study was also limited by a focus on Predominantly White Institutions (PWIs). It is possible that the present results may not replicate at Minority Serving Institutions (MSIs). This would serve as way of generalizing the findings of the present study to all faculty. Research by Ho et al., (2009) did, however, find that racial minority and White students viewed racial minority faculty in identical ways. Similarly, another study found that grades were related to SETs at a MSI (Guinn & Vincent, 2006). It is, therefore, possible, that similar perceptual processes could be guiding SETs at PWIs and MSIs.

In addition, it is possible that the results of the present study do not generalize beyond selective, liberal arts colleges. Relative to research-intensive universities, selective liberal arts colleges place a much greater emphasis on teaching (Aries, 2008; Ruscio, 1987). The focus on undergraduate education suggests that selective liberal arts colleges represent a more conservative test of the effect of faculty race and gender because selective liberal arts colleges attempt to hire and help develop the best teachers (Boyer, 1997). It is, though, possible that the racial differences in faculty SETs could be minimized at institutions where teaching matters less. This possibility should be investigated by future research.

Finally, it is not only possible, but likely that mistakes were made in the coding of the instructor's race. It is, though, also likely that the racial assumptions made by coders in the present study were similar to those made by the students. From that perspective, the professor's actual race is less important than that individual's perceived race. Future research should further examine both between-groups racial differences in the evaluation of teaching, and within-group differences based on appearance (e.g., skin tone, racially phenotypic features).

Conclusion

Despite the limitations of the present research, it highlights the importance of considering how an instructor's demographic characteristics can affect student evaluations of teaching. The findings of the present study suggest that, a priori, SETs may place some racial minority faculty at an evaluative disadvantage compared to their White peers. This problem can be compounded by institutions that demand excellent, not merely good, teaching for promotion and tenure (Boyer, 1997; Nast, 1999; Williams, 2007).

It is important to consider that the interaction between racial majority students and racial minority faculty is an intergroup contact experience (for a review, see Tropp & Pettigrew, 2005). The institutions examined in the present study had White majorities at both the student and faculty levels. The prospect of interacting across racial lines can induce enough anxiety that it can impair cognitive functioning (Richeson & Shelton, 2007). From this view, students are negotiating a social interaction burdened by the racial legacy of a nation. Despite these difficulties, interaction with racial minority faculty remains an important route for helping students to overcome their biases.

References

- Allen, W. R., Epps, E. G., Guillory, E. A., Suh, S. A., & Bonous-Hammarth, M. (2000). The black academic: Faculty status among African Americans in U.S. higher education. *Journal of Negro Education, 69*, 112–127.
- Anderson, K. J., & Smith, G. (2005). Students' preconceptions of professors: Benefits and barriers according to ethnicity and gender. *Hispanic Journal of Behavioral Sciences, 27*, 184–201.

- Arbuckle, J., & Williams, B. D. (2003). Students' perceptions of expressiveness: Age and gender effects on teacher evaluations. *Sex Roles, 49*, 507–516.
- Aries, E. (2008). *Race and class matters at an elite college*. Philadelphia: Temple University Press.
- Babad, E., Avni-Babad, D., & Rosenthal, R. (2004). Prediction of students' evaluations from brief instances of professors' nonverbal behavior in defined instructional situations. *Social Psychology of Education, 7*, 3–33.
- Babad, E., Darley, J. M., & Kaplowitz, H. (1999). Developmental aspects in students' course selection. *Journal of Educational Psychology, 91*, 157–168.
- Banaji, M., & Greenwald, A. G. (1994). Implicit stereotyping and prejudice. *The psychology of prejudice: The Ontario symposium, 7*, 55–76.
- Barth, M. M. (2008). Deciphering student evaluations of teaching: A factor analysis approach. *Journal of Education for Business, 84*, 40–46.
- Basow, S. A. (1990). Effects of teacher expressiveness: Mediated by teacher sex-typing? *Journal of Educational Psychology, 82*, 599–602.
- Basow, S. A. (1995). Student evaluations of college professors: When gender matters. *Journal of Educational Psychology, 87*, 656–665.
- Basow, S. A. (1998). Student evaluations: The role of gender bias and teaching styles. In L. H. Collins, J. C. Chrisler, & K. Quina (Eds.), *Career strategies for women in academe: Arming Athena* (pp. 135–156). Thousand Oaks, CA: Sage.
- Basow, S. A. (2000). Best and worst professors: Gender patterns in students' choices. *Sex Roles, 45*, 405–417.
- 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.
- Beatty, M. J., & Zahn, Ch. J. (1990). Are student ratings of communication instructors due to "easy" grading practices? An analysis of teacher credibility and student-reported performance levels. *Communication Education, 39*, 275–291.
- Bennett, S. K. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. *Journal of Educational Psychology, 74*, 170–179.
- Beran, T., & Violato, C. (2005). Ratings of university teacher instruction: How much do student and course characteristics really matter? *Assessment & Evaluation in Higher Education, 30*, 593–601.
- Best, J. B., & Addison, W. E. (2000). A preliminary study of perceived warmth of professor and student evaluations. *Teaching of Psychology, 27*, 60–62.
- Blackhart, G. C., Peruche, B. M., DeWall, C. N., & Joiner, T. E. J. (2006). Factors influencing teaching evaluations in higher education. *Teaching of Psychology, 33*, 37–39.
- Boatright-Horowitz, S., & Soeung, S. (2009). Teaching white privilege to white students can mean saying good-bye to positive student evaluations. *American Psychologist, 64*, 574–575.
- Bodenhausen, G. V., & Lichtenstein, M. (1987). Social stereotypes and information-processing strategies: The impact of task complexity. *Journal of Personality and Social Psychology, 52*, 871–880.
- Bonilla-Silva, E. (2006). *Racism without racists: Color-blind racism and the persistence of racial inequality in the United States*. Lanham, MD: Rowman & Littlefield Publishers.
- Bowling, N. A. (2008). Does the relationship between student ratings of course easiness and course quality vary across schools? The role of school academic rankings. *Assessment & Evaluation in Higher Education, 33*, 1–12.
- Boyer, E. L. (1997). *Scholarship reconsidered: Priorities of the professorate*. New York: Jossey-Bass.
- Brigham, J. C. (1993). College students' racial attitudes. *Journal of Applied Social Psychology, 23*, 1933–1967.
- Buchert, S., Laws, E. L., Apperson, J. M., & Bregman, N. J. (2008). First impressions and professor reputation: Influence on student evaluations of instruction. *Social Psychology of Education, 11*, 397–408.
- Campbell, H. E., Gerdes, K., & Steiner, S. (2005). What's looks got to do with it? Instructor appearance and student evaluations of teaching. *Journal of Policy Analysis and Management, 24*, 611–620.
- Cashin, W. E. (1995). *Student ratings of teaching: The research revisited* (IDEA Paper No. 32). Manhattan: Kansas State University, Center for Faculty Evaluation and Development.
- Chowdhary, U. (1988). Instructor's attire as a biasing factor in students' ratings of an instructor. *Clothing & Textiles Research Journal, 6*, 17–22.
- Clayson, D. E., Frost, T. F., & Sheffet, M. J. (2006). Grades and the student evaluation of instruction: A test of the reciprocity effect. *Academy of Management Learning & Education, 5*, 52–85.
- Devine, P. D. (1989). Stereotypes and prejudices: Their automated and controlled components. *Journal of Personality and Social Psychology, 56*, 5–18.
- Dixon, T., & Maddox, K. B. (2005). Skin tone, crime news, and social reality judgments: Priming the stereotype of the dark and dangerous black criminal. *Journal of Applied Social Psychology, 35*, 1555–1570.

- Dovidio, J. F., & Gaertner, S. L. (2004). Aversive Racism. *Advances in experimental social psychology*, *36*, 1–52.
- 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.
- Evans, G. L., & Cokley, K. O. (2008). African American women and the academy: Using career mentoring to increase research productivity. *Training and Education in Professional Psychology*, *2*, 50–57.
- Feldman, K. A. (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.
- Felton, J., Mitchell, J., & Stinson, M. (2004). Web-based student evaluations of professors: The relations between perceived quality, easiness, and sexiness. *Assessment & Evaluation in Higher Education*, *29*, 91–108.
- Fortson, S. B., & Brown, W. E. (1998). Best and worst university instructors: The opinions of graduate students. *College Student Journal*, *32*, 572–576.
- Glick, P., & Fiske, S. T. (1999). The ambivalence toward men inventory: Differentiating hostile and benevolent beliefs about men. *Psychology of Women Quarterly*, *23*, 519–536.
- Goebel, B. L., & Cashen, V. M. (1985). Age stereotype bias in student ratings of teachers: Teacher, age, sex, and attractiveness as modifiers. *College Student Journal*, *19*, 404–410.
- Greenwald, A., & Banaji, M. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27.
- Greenwald, A. G., & Gilmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, *52*, 1209–1217.
- Guinn, B., & Vincent, V. (2006). The influence of grades on teaching effectiveness ratings at a Hispanic-serving institution. *Journal of Hispanic Higher Education*, *5*, 313–321.
- 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.
- Heckert, T. M., Latier, A., Ringwald, A., & Silvey, B. (2006). Relation of course, instructor, and student characteristics to dimensions of student ratings of teaching effectiveness. *College Student Journal*, *40*, 195–203.
- Hendrix, K. G. (1997). Student perceptions of verbal and nonverbal cues leading to images of black and white professor credibility. *Howard Journal of Communications*, *8*, 251–273.
- Hendrix, K. G. (1998). Student perceptions of the influence of race on professor credibility. *Journal of Black Studies*, *28*, 738–763.
- 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.
- Howell, A. J., & Symbaluk, D. G. (2001). Published student ratings of instruction: Revealing and reconciling the views of students and faculty. *Journal of Educational Psychology*, *93*, 790–796.
- Jackson, R., & Crawley, R. (2003). White student confessions about a black male professor: A cultural contracts theory approach to intimate conversations about race and worldview. *The Journal of Men's Studies*, *12*, 25–41.
- Jenkins, S. J., & Downs, E. (2001). Relationship between faculty personality and student evaluation of courses. *College Student Journal*, *35*, 636–640.
- Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *The Journal of Abnormal and Social Psychology*, *28*, 280–290.
- Kelley, H. H. (1950). The warm-cold variable in first impressions of persons. *Journal of Personality*, *18*, 431–439.
- 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*, 342–344.
- Kincheloe, J. L., & Steinberg, S. R. (1998). Addressing the crisis of whiteness: Reconfiguring white identity in a pedagogy of whiteness. In J. Kincheloe, S. Steinberg, N. Rodriguez, & R. Chennault (Eds.), *White reign: Deploying whiteness in America* (pp. 3–29). New York: St. Martin's Press.
- Kindred, R., & Mohammed, S. (2005). "He will crush you like an academic ninja!": Exploring teacher ratings on ratemyprofessors.com. *Journal of Computer-Mediated Communication*, *10*, Article 9.
- Lackritz, J. R. (2004). Exploring burnout among university faculty: Incidence, performance, and demographic issues. *Teaching and Teacher Education*, *20*, 713–729.
- Landrine, H. (1999). Race x class stereotypes of women. In L. A. Peplau, S. C. DeBro, R. C. Veniegas, & P. L. Taylor (Eds.), *Gender, culture, and ethnicity: Current research about women and men* (pp. 38–61). Mountain View, CA: Mayfield Publishing Co.
- Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology*, *72*, 275–287.
- Liddle, B. J. (1997). Coming out in class: Disclosure of sexual orientation and teaching evaluations. *Teaching of Psychology*, *24*, 32–35.

- Lin, M. H., Kwan, V. S. Y., Cheung, A., & Fiske, S. T. (2005). Stereotype content model explains prejudice for an envied outgroup: Scale of anti-Asian American stereotypes. *Personality and Social Psychology Bulletin*, *31*, 34–47.
- Ludwig, J. M., & Meacham, J. A. (1997). Teaching controversial courses: Student evaluations of instructors and content. *Educational Research Quarterly*, *21*, 27–38.
- Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120.
- Maddox, K. B. (2004). Perspectives on racial phenotypicity bias. *Personality and Social Psychology Review*, *8*, 383–401.
- Maddox, K. B., & Gray, S. (2002). Cognitive representations of Black Americans: Reexploring the role of skin tone. *Personality and Social Psychology Bulletin*, *28*, 250–259.
- Marlin, J. W., & Gaynor, P. E. (1989). Do anticipated grades affect student evaluations? A discriminant analysis approach. *College Student Journal*, *23*, 184–192.
- Marsh, H. W. (2007). Do university teachers become more effective with experience? A multilevel growth model of students' evaluations of teaching over 13 years. *Journal of Educational Psychology*, *99*, 775–790.
- Martin, E. (1984). Power and authority in the classroom: Sexist stereotypes in teaching evaluations. *Journal of Women in Culture and Society*, *9*, 482–492.
- McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist*, *52*, 1218–1225.
- McPherson, M. A., & Jewell, R. T. (2007). Leveling the playing field: Should student evaluation scores be adjusted? *Social Science Quarterly*, *88*, 868–881.
- Messner, M. A. (2000). White guy habitus in the classroom: Challenging the reproduction of privilege. *Men and Masculinities*, *2*, 457–469.
- Millea, M., & Grimes, P. W. (2002). Grade expectations and student evaluation of teaching. *College Student Journal*, *36*, 582–590.
- Myers, L. W. (2005). Black women coping with stress in academia. *The psychology of prejudice and discrimination: Bias based on gender and sexual orientation* (Vol. 3., pp. 133–149). Westport, CT: Praeger/Greenwood Publishing Group.
- Nast, H. J. (1999). 'Sex,' 'race,' and multiculturalism: Critical consumption and the politics of course evaluations. *Journal of Geography in Higher Education*, *23*, 102–115.
- Ogier, J. (2005). Evaluating the effect of a lecturer's language background on a student rating of teaching form. *Assessment & Evaluation in Higher Education*, *30*, 477–488.
- Otto, J., Sanford, D. A. J., & Ross, D. N. (2008). Does ratemyprofessor.com really rate my professor? *Assessment & Evaluation in Higher Education*, *33*, 1–16.
- Perry, R. P., Abrami, P. C., Leventhal, L., & Check, J. (1979). Instructor reputation: An expectancy relationship involving student ratings and achievement. *Journal of Educational Psychology*, *71*, 776–787.
- Punj, G., & Stewart, D. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, *20*, 134–148.
- Richeson, J., & Shelton, J. (2007). Negotiating interracial interactions: Costs, consequences, and possibilities. *Current Directions in Psychological Science*, *16*, 316–320.
- Riniolo, T. C., 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.
- Ruscio, K. P. (1987). The distinctive scholarship of the selective liberal arts college. *The Journal of Higher Education*, *58*, 205–222.
- Sidanius, J., & Crane, M. (1989). Job evaluation and gender: The case of university faculty. *Journal of Applied Social Psychology*, *19*, 174–197.
- Silva, K. M., Silva, F. J., Quinn, M. A., Draper, J. N., Cover, K. R., & Munoff, A. A. (2008). Rate my professor: Online evaluations of psychology instructors. *Teaching of Psychology*, *35*, 71–80.
- Smith, B. P. (2007). Student ratings of teacher effectiveness: An analysis of end-of-course faculty evaluations. *College Student Journal*, *41*, 788–800.
- Smith, G., & Anderson, K. J. (2005). Students' ratings of professors: The teaching style contingency for Latino/a professors. *Journal of Latinos and Education*, *4*, 115–136.
- Sprinkle, J. E. (2008). Student perceptions of effectiveness: An examination of the influence of student biases. *College Student Journal*, *42*, 276–293.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613–629.
- Sue, D. W. (2004). Whiteness and ethnocentric monoculturalism: Making the "invisible" visible. *American Psychologist*, *59*, 761–769.
- Sue, S., Sue, D. W., & Sue, D. W. (1975). Asian Americans as a minority group. *American Psychologist*, *30*, 906–910.
- Swim, J. K., & Cohen, L. L. (1997). Overt, covert, and subtle sexism: A comparison between the Attitudes Toward Women and Modern Sexism Scales. *Psychology of Women Quarterly*, *21*, 103–118.

- Tatro, C. N. (1995). Gender effects on student evaluations of faculty. *Journal of Research & Development in Education, 28*, 169–173.
- Timmerman, T. (2008). On the validity of Rate-MyProfessors.com. *Journal of Education for Business, 84*, 55–61.
- Tropp, L., & Pettigrew, T. (2005). Relationships between intergroup contact and prejudice among minority and majority status groups. *Psychological Science, 16*, 951–957.
- Turner, C. S. V., & Myers, S. L. (2000). *Faculty of color in academe: Bittersweet success*. New York: Pearson.
- U.S. News & World Report. (2005). *U.S. News Ultimate College Guide 2006*. Napierville, IL: Sourcebooks.
- Van Giffen, K. (1990). Influence of professor gender and perceived use of humor on course evaluations. *Humor: International Journal of Humor Research, 3*, 65–73.
- Widmeyer, W. N., & Loy, J. W. (1988). When you're hot, you're hot: Warm-cold effects in first impressions of persons and teaching effectiveness. *Journal of Educational Psychology, 80*, 118–121.
- Wilds, D. J. (2000). *Minorities in higher education 1999–2000: Seventeenth annual Status report*. Report No. ED457768. Washington, DC: American Council on Education.
- Williams, D. A. (2007). Examining the relation between race and student evaluation of faculty members: A literature review. *Profession, 168*–173.
- Wilson, J. H., & Taylor, K. W. (2001). Professor immediacy as behaviors associated with liking students. *Teaching of Psychology, 28*, 136–138.
- Worthington, R. L., Navarro, R. L., Loewy, M., & Hart, J. (2008). Color-blind racial attitudes, social dominance orientation, racial-ethnic group membership and college students' perceptions of campus climate. *Journal of Diversity in Higher Education, 1*, 8–19.
- Zarate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition, 8*, 161–185.

Received October 7, 2008

Revision received January 4, 2010

Accepted January 4, 2010 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://notify.apa.org/> and you will be notified by e-mail when issues of interest to you become available!