
Professor Gender, Age, and “Hotness” in Influencing College Students’ Generation and Interpretation of Professor Ratings

Sara L. Sohr-Preston^{*,a}, Stefanie S. Boswell^b, Kayla McCaleb^a, and Deanna Robertson^a

^a Southeastern Louisiana University, USA

^b University of the Incarnate Word, USA

Submitted: June 9, 2016 | **Peer-reviewed:** September 18, 2016

Revisions required: September 21, 2016 | **Accepted:** September 29, 2016

Author version published: September 30, 2016 | **Last updated:** November 3, 2016

Abstract: A sample of 230 undergraduate psychology students rated their expectations of a bogus professor (who was randomly designated a man or woman and “hot” versus “not hot”) based on ratings and comments found on RateMyProfessors.com. Five professor qualities were derived using principal components analysis: dedication, attractiveness, enhancement, fairness, and clarity. Participants rated current actual psychology professors on the same qualities. Current professors were divided based on gender (man or woman), age (under 35 or 35 and older), and attractiveness (at or below the median or above the median). Using a multivariate analysis of covariance (MANCOVA), students expected hot professors to be more attractive but lower in clarity. They rated current professors who were male and 35 or older as lowest in clarity. Current professors scored significantly lower in dedication, enhancement, fairness, and clarity when rated at or below the median on attractiveness. Results, along with previous research, suggest numerous factors (largely out of professors’ control) influencing how students interpret and create professor ratings. Caution is therefore warranted in using online ratings to inform a variety of decisions, including students’ course selection or even administrators’ hiring and promotion decision making.

Keywords: professor ratings, bias, higher education, psychology faculty, teaching

Introduction

University instructors have long been regularly evaluated by their own students. In recent decades, opportunities for college students to express their opinions of their professors have expanded beyond formal university-administered group evaluations to posts on informal websites like RateMyProfessors.com (RMP; Johnson & Crews, 2013). RMP is a website on which university students may post anonymous evaluations of their professors. Since its inception in 1999, the service has become immensely popular throughout the world; users have created over 17 million ratings for 1.6 million professors in Canada, the United Kingdom, and the United States (RMP, 2016a). Sites similar to RMP operate in other countries; for example, Rate My Teachers for Republic of Ireland, Australia, and New Zealand (i.e., ratemyteachers.com). In addition to writing narrative comments about their professors, students use RMP to rate their professors on helpfulness, instructional clarity, and course easiness using a rating scale of 1 (*low*) to 5 (*high*). Scores ranging from 3.5 to 5 are considered “good,” scores ranging from 2.5 to 3.4 are considered

* Corresponding author (sara.sohr-preston@selu.edu)

“average,” and scores ranging from 1 to 2.4 are considered “poor” (RMP, 2016b). Helpfulness and instructional clarity ratings are averaged to produce an overall quality score (RMP, 2016b). Moreover, students also rate professors as “hot” or “not hot”; a chili pepper icon appears on the RMP profile of professors whose aggregate hot ratings exceed their not-hot ratings (RMP, 2016c). While RMP does not explicitly state that the chili pepper icon represents a professor’s physical attractiveness, many assume this to be true (Landry, Kurkul, & Poirier, 2010).

The impetus for RMP and similar websites is to provide students with a forum to exchange information about their professors and courses (RMP, 2016a). Given that the results of university-administered teaching evaluations are typically inaccessible to students (Kindred & Mohammed, 2005), RMP offers students a publically available platform for the sharing of course and professor data. By providing potential students with otherwise inaccessible ratings from former students (Kindred & Mohammed, 2005), those considering the professor may inform their enrollment choices in hope of receiving a higher quality college education (Davison & Price, 2009; Johnson & Crews, 2013). Not surprisingly, concerns have been voiced about potential bias in online ratings, and many professors doubt the utility of sites like RMP for students truly seeking a higher quality education (Boswell, 2016; Davison & Price, 2009).

Professors cite several sources of concern regarding the validity of RMP ratings (e.g., Davison & Price, 2009; Hartman & Hunt, 2013; Sonntag, Bassett, & Snyder, 2009). First, there is no guarantee that ratings have actually been posted by former students of the professor (Johnson & Crews, 2013; Montell, 2006; Otto, Sanford, & Ross, 2008; Timmerman, 2008). For example, the first and second authors both have at least once been rated on RMP for classes they have never taught, possibly a result of students not correctly remembering the names of the actual instructors. While instances such as this may produce laughter and seem fairly harmless, more alarming cases have been recorded involving negative postings made by rivals or disgruntled colleagues instead of students (see Carnevale, 2006). Second, even when postings are crafted by actual students, concerns remain about the validity of such postings as reflections of teaching quality or as windows into what potential students may expect from taking a class with a particular professor (Legg & Wilson, 2012). For example, students self-selecting to participate in RMP posting may harbor deeply felt or extreme views and may not represent a professor’s general student body (Boswell, 2015; Legg & Wilson, 2012). Further concerns exist about possible biasing factors shaping how online professor ratings are both interpreted and created.

The purpose of the current investigation was to examine potential sources of bias in both interpreting and posting online professor ratings. It should be noted that there is evidence for a variety of sources of bias in university-administered teaching evaluations (which would presumably involve fewer concerns about the identity of the evaluator than RMP-type postings). For example, university-administered teaching evaluations may be subject to bias from whether the evaluated course is required versus elective (Divoky & Rothermel, 1988; Feldman, 1978; Patrick, 2011; Petchers & Chow, 1988; Scherr & Scherr, 1990), whether the course is higher level versus lower level (Goldberg & Callahan, 1991; Moritsch & Suter, 1988; Patrick, 2011), and whether it is a humanities or social sciences course versus a math or science course (Cashin, 1992; Patrick, 2011). Students’ perceptions of the instructor’s personality characteristics or interaction style also impact evaluations, which could reflect influence on actual quality of teaching or on more peripheral features like likability (Ahmadi, Helms, & Raiszadeh, 2001; Clayson & Sheffet, 2006; Feldman, 1986; Hart & Driver, 1978; Jenkins & Downs, 2001; Patrick, 2011; Widmeyer & Loy, 1988; R. Wilson, 1998). Moreover, other teaching-irrelevant qualities such as style of dress (Eadie, 1996; Sebastian & Bristow, 2008), formality of name (e.g., title and last name versus first name; Sebastian & Bristow, 2008), and ability to be entertaining (Gottlieb, 2011) affect students’ evaluations of professors. These teaching-irrelevant qualities exert such an

influence upon professors' evaluations that it is possible for students to predict how a professor will be evaluated simply by watching a muted video clip of the professor; entertaining individuals, even with no demonstrated knowledge of the topic, benefit from bias and receive higher teaching evaluations (Ambady & Rosenthal, 1993).

Given that RMP provides students with a similar opportunity to offer opinions of professors, it is likely that RMP ratings will be affected by the same sources of bias that affect their university-administered counterparts. Moreover, the lack of quality control (Johnson & Crews, 2013) inherent in ratings on RMP and similar sites likely allows for additional sources of bias to shape the content of ratings, for example, course easiness (Felton, Koper, Mitchell, & Stinson, 2008; Felton, Mitchell, & Stinson, 2004) and professor race (Reid, 2010).

Investigating potential sources of bias or distortion of online professor ratings is particularly important because, while RMP presents itself as an entertainment site, its data is commonly used for numerous purposes well beyond pure entertainment (Landry et al., 2010). Research has demonstrated that many students use such sites before signing up for classes with specific professors (Hossain, 2009), and exposure to online professor ratings may actually influence students' expectations and motivations for their own performance in a course (Edwards, Edwards, Qing, & Wahl, 2007; Edwards, Edwards, Shaver, & Oaks, 2009). Exposure to positive or negative online ratings has even been shown to influence students' teaching evaluations of an actual classroom lecture (Lewandowski, Higgins, & Nardone, 2012). Moreover, RMP evaluations may influence students' in-class behaviors such as notetaking and participation in class discussions and activities (Kowai-Bell, Guadagno, Little, Preiss, & Hensley, 2011). Taken together, these findings suggest that RMP content may exert a significant influence upon students' learning and academic achievement.

In addition to its effects on students, RMP's consequences also extend to institutions and professors. For example, RMP scores and narrative comments at least partially contribute to some college rankings (Howard, 2014; Johnson & Crews, 2013) as well as promotion and hiring decisions (Johnson & Crews, 2013; Montell, 2006; Pannapacker, 2007). Professors, too, appear affected by online rating content in terms of their affect and self-efficacy, and the effects do not differ from those of reading more respected university-administered student evaluations of teaching (Boswell, 2016). Past research regarding targeted possible influences on interpretation and generation of online professor ratings will be described in detail below.

Gender of Professor

Gender expectations may lead students to interpret professor ratings or rate professors differently based on whether they are men or women. Prior research has indicated that college students tend to rate men professors more favorably than women professors (Abel & Meltzer, 2007; Basow & Silberg, 1987; Joye & Wilson, 2015). This bias extends even to the online classroom, where students and professors do not interact in person. For example, assistant instructors, working under both a stereotypically male and stereotypically female pseudonym, were evaluated more favorably when using the male identity (MacNell, Driscoll, & Hunt, 2015). Other research has revealed, however, that a main effect of professor gender is likely best interpreted within the context of additional moderating factors. For example, college students evaluate women professors more harshly when they do not conform to gender-based expectations of helpfulness and flexibility (Bennett, 1982), and women are evaluated more severely than men when they have high grading standards and teach academically rigorous courses (Sinclair & Kunda, 2000). Other research has suggested that student raters interpret professor qualities differently when rating men versus women, attributing lack of clarity to low

effort from men professors but low ability in women professors (Stuber, Watson, Carle, & Staggs, 2009). Furthermore, students rating a supposed applicant for a university teaching position appear to expect different qualities from men versus women applicants (Burns-Glover & Veith, 1995). Based on this body of findings, the following hypotheses were included:

Hypothesis 1: College students' interpretations of online professor ratings will differ by professor gender (whether the professor is described as a man versus a woman).

Hypothesis 2: College students will rate their own professors differently depending on whether they are a man versus a woman.

Age of Professor

In addition to evidence for gender bias, studies indicate a professor's age also affects students' evaluations of teaching (Bianchini, Lissoni, & Pezzoni, 2013; Feldman, 1983; Kinney & Smith, 1992). For example, Arbuckle and Williams (2003) demonstrated that students would evaluate a lecture more positively if they were led to believe that it was delivered by a young (i.e., under 35) man, supporting the notion that students expect college professors to be or at least resemble younger men (Messner, 2000). Moreover, in an analysis of RMP ratings, Stonebraker and Stone (2015) found that elevated age negatively affects students' ratings of professors' teaching; these effects begin as early as professors' mid-forties. Other recent findings also indicate that students rate older professors most harshly (Joye & Wilson, 2015; J. Wilson, Beyer, & Monteiro, 2014; Zabaleta, 2007), possibly because of the professors' dissimilarity to students who are typically in their late teens or early twenties (Gehrt, Louie, & Osland, 2015). The following hypotheses were generated from this evidence supporting a potential main effect of professor age and, based on the findings of Arbuckle and Williams (2003), a moderating effect of professor gender on this age effect:

Hypothesis 3: College students will rate their own professors differently depending on whether they are under 35 versus 35 or older.

Hypothesis 4: Professor age and gender will interact in influencing students' evaluations of their own professors.

While college students may interpret online professor ratings differently depending on whether the professor is described as younger versus older, this hypothesis was not tested in the current study. This was largely because of concerns that online comments on professor rating sites do not often include information regarding professor age.

The Chili Pepper (or "Hotness")

Past studies also have indicated that professors higher in hotness (defined in various ways, but generally encompassing students' subjective appraisal of professors' physical features) are perceived more positively than those lacking hotness (Bonds-Raacke & Raacke, 2007; Buck & Tiene, 1989; Felton et al., 2004; Felton et al., 2008; Freng & Webber, 2009; Hamermesh & Parker, 2005; Liu, Hu, & Furutan, 2013; Riniolo, Johnson, Sherman, & Misso, 2006; Romano & Bordieri, 1989). Teaching at the college level may represent one of the countless situations in which people attribute higher degrees of socially desirable traits to attractive people but not less attractive individuals (Dion, Berscheid, & Walster, 1972), often summarized with the phrase "what is beautiful is good" (Eagly, Ashmore, Makhijani, & Longo, 1991, p. ●●●).

RMP provides information on hotness with the chili pepper feature. Professors with a chili pepper are hot; professors without a chili pepper lack hotness (RMP, 2016c). Research on the effects of the dichotomous chili pepper feature has yielded mixed results; some studies indicate little influence on perceptions of teaching quality (Coladarci & Kornfield, 2007) and even lack of user respect for its meaningfulness (Kindred & Mohammed, 2005), but others demonstrate that professors with a chili pepper receive more favorable ratings (Lawson & Stephenson, 2005). For example, hotness may impact ratings of professors' clarity and helpfulness (Bonds-Raacke & Raacke, 2007). The pool of evidence, while mixed, led the authors to include the following hypotheses:

Hypothesis 5: College students will interpret online professor ratings differently depending on whether the professor is noted to be hot or not.

Hypothesis 6: College students will rate current professors differently depending on how physically attractive they find them.

Given the evidence for self-reported females rating their instructors differently than self-reported males (Burns-Glover & Veith, 1995; Kohn & Hatfield, 2006), exhibiting preference for professors of the same gender (Gehrt et al. 2015) and being more influenced by hotness in evaluating professors (Liu et al., 2013), all hypotheses were tested while controlling for participants' reported sex. Because some have proposed perceived similarity as the driving feature in age effects on professor ratings (Gehrt et al., 2015), participant age was also included as a control variable for both hypotheses addressing professor age.

Method

Participants

The convenience sample consisted of 230 college undergraduate students enrolled in introductory psychology and lifespan developmental psychology classes at a public university in the southern United States. Participants were recruited from 18 class sections ranging in size from 40 to 100 students. Professors teaching these sections during the year of study recruitment included six men and six women. Students received course credit for participation; the option to participate in other studies and an alternate assignment were available to students who chose not to participate or did not meet inclusion requirements for this investigation. This study specified that participants must have been 18 years of age or older at the time of participation.

Participants providing their sex ($n = 228$) included 43 males (18.70%) and 185 females (80.43%). The average age was 19.54 years ($SD = 1.62$). The sample was predominantly White (70.00%), with 20.43% African American, 3.91% Hispanic or Latino, 2.61% Asian, and 0.87% Native American, Aleut, or Aboriginal peoples. Participants' self-reported grade point averages were somewhat high ($M = 3.18$; $SD = .60$), averaging in the range of a letter grade of B. The vast majority of participants (91.74%) had previously visited RMP, and 76.50% reported using the site to make decisions about enrolling in classes at least some of the time. Out of the full sample, 218 participants' (41 males, 177 females) data were complete on all proposed covariates and independent and dependent variables and were included in final statistical analyses. The Institutional Review Board reviewed and approved this study prior to recruitment.

Procedure

Data collection occurred online. Participants accessed the questionnaire using a web link posted by the primary investigator on the psychology department's participant recruitment site. Upon opening the survey, participants were randomly assigned to view one of the different versions of the bogus professor's online rating summary. Before viewing this material and follow-up questions regarding their expectations of the bogus professor, their ratings of a current actual professor, and demographics, participants initialed an informed consent form. They were required to complete the questionnaires in one session. Instructions stated that participants were allowed to skip any items with which they felt uncomfortable. Participants were assured that all information would be kept confidential and that no responses would be shared with their psychology professors. Data were managed by a graduate research assistant not teaching any psychology classes so as to maintain strict confidentiality.

Measures

Demographics. Participants completed closed-ended items addressing their own sex and race. In addition, they completed open-ended items regarding their own current age and grade point average on a 5-point scale (0 = F, 4 = A).

Previous exposure to RMP. Participants reported whether they had ever visited RMP (a yes/no item) and how often they made decisions about enrolling in classes based on RMP ratings using a rating scale from 1 (*never*) to 5 (*always or almost always*).

Perceived demographics of current professor. Participants were asked how old they believed their current psychology professor to be (*under 35 years old* versus *35 years old or older*) using a closed-ended item. They additionally responded as to whether their current psychology professor was a man or a woman. Out of the full sample of responses, 106 participants reported that their professor was a man, while 115 identified their professor as a woman. In addition, 95 participants perceived their professor to be under 35, and 126 perceived their professor to be 35 or older.

Interpretation of online professor ratings. Because the investigators desired to learn more about students' expectations from *and* creation of professor ratings, professor ratings were assessed in two distinct ways. First, to measure students' expectations of a professor based on reading professor ratings, participants read an online rating of a bogus professor and then rated the professor on a series of teaching and personal qualities (see Table 1). The stimulus professor's rating was presented in the format used on RMP; ratings were in the average range for helpfulness (3.2 out of 5) and in the good range for clarity (4.2 out of 5), easiness (4.0 out of 5), and overall quality (3.8 out of 5). The researchers intended the scores to indicate neutral to good quality. Extreme scores were avoided so that the scores themselves would not overwhelmingly command attention. The description also listed whether the professor had a chili pepper to indicate being hot. There were four different versions of the described professor: a woman with no chili pepper, a woman with a chili pepper, a man with no chili pepper, and a man with a chili pepper. The professor was always listed with the intentionally gender-neutral name "Alex Johnson," but pronouns differed between female and male versions.

After reading the bogus RMP listing, participants were asked how often they would predict the professor would display a series of qualities. Items were scaled from 1 (*never or almost never*) to 4 (*always or nearly always*). The items presented (see Table 1) were crafted for the present investigation. The content of the items was generated by the investigators based on frequent content of their own teaching evaluations and online professor ratings. Additionally, students

enrolled in independent research were asked to contribute any additional items they deemed appropriate.

To examine potential subcategories of teaching qualities, a principal components analysis (PCA) was conducted using promax rotation because items were anticipated to be correlated. The Kaiser–Meyer–Olkin measure of sampling adequacy was .92, and Bartlett’s test of sphericity was statistically significant ($p = .00$), indicating the data was suitable for PCA (Field, 2013). Factor loadings obtained from the pattern matrix are summarized in Table 1. The authors employed a criterion of a factor loading of .40 or higher for inclusion of an item in a particular subcategory. The total number of factors or subcategories was determined using the Kaiser criterion of an eigenvalue of at least 1.00 (Field, 2013).

As seen in Table 1, the first factor, *Dedication*, accounting for 36.77% of variance, included items centered on the theme of professors behaving in a professional and respectful manner and conveying general enjoyment of teaching. The second factor, *Attractiveness*, explaining 12.95% of variance, included items about the professor’s physical appearance. The third factor, *Enhancement*, accounting for 7.69% of variance, included items referring to enrichment of teaching, what some might label “the little extras” in obtaining and maintaining student interest. Next, the fourth factor, *Fairness*, representing 4.02% of variance, reflected a theme of evenhandedness and consistency in teaching and grading. Finally, the fifth factor, *Clarity*, explaining 3.42% of variance, centered on making oneself understood by students.

Table 1. Summary of Factor Loadings From Principal Components Analysis of Professor Rating Items (Promax Rotation)

Item	Factor				
	1	2	3	4	5
Arrives on time	.45	.16	-.29	.27	.28
Is helpful	.63	-.15	-.02	.23	-.09
Is polite to students	.67	.08	-.10	.26	-.22
Smiles	.48	.23	.03	.30	-.32
Knows students by name or face	.40	.07	.18	.31	-.32
Exhibits respect toward students	.57	.03	-.08	.29	.14
Appears to be in a good mood	.70	.15	.15	-.03	-.14
Displays confidence	.66	.30	.08	-.22	.06
Seems to enjoy teaching	.90	-.01	.03	-.26	.03
Hands back tests and assignments quickly	.73	-.16	.07	-.08	.17
Is available outside of class	.89	-.19	.04	-.16	.10
Uses class time discussing relevant topics	.70	-.15	.12	-.06	.19
Dresses professionally	.12	.77	-.25	.11	.16
Is physically attractive	-.17	.96	.02	.01	.04
Is physically fit	-.13	.96	.03	-.03	.04
Has attractive or appropriate hairstyle	-.08	.93	.11	-.04	.03
Wears flattering clothing	.00	.91	.09	-.04	.00
Uses examples from popular culture	.11	.07	.71	-.02	.09
Shows videos	-.19	.02	.94	.13	.02
Performs in-class demonstrations	.02	-.01	.84	.03	.08
Gives extra credit	.33	-.06	.46	.10	-.16
Tells funny jokes	.22	.21	.55	-.23	.05
Shares relevant examples in class	.25	.02	.46	-.03	.25

Gives handouts	.02	-.13	.65	.25	.06
Has clear grading policies	.21	-.06	-.10	.57	.28
Tests material covered in lectures	.06	.03	-.01	.75	.03
Tests material covered in readings	-.37	-.07	.21	.89	.05
Is consistent	.13	.06	.12	.56	.28
Explains topics clearly	.08	.10	-.04	.18	.70
Speaks loudly enough to be heard	-.11	.13	.28	.00	.58
Breaks down complicated topics	.40	-.11	.14	.10	.44

Note. Bolded values indicate inclusion in the composite variable. Factor 1 = Dedication; Factor 2 = Attractiveness; Factor 3 = Enhancement; Factor 4 = Fairness; Factor 5 = Clarity.

Based on these factor loadings, five composite variables were created with the mean scores for included items. The mean was used in place of a sum or other calculation so as to maintain the item scaling of 1 (*never or almost never*) to 4 (*always or nearly always*). Descriptive statistics, internal consistency within factors, and correlations among the computed variables are listed in Table 2. As seen in Table 2, internal consistency, as assessed with Cronbach's α , was high (>.70) for all but the clarity variable. This composite variable included the fewest individual items, and scales with fewer items often exhibit lower consistency as calculated with Cronbach's formula (Peterson, 1994).

Table 2. Descriptive Statistics and Correlations Among Variables Assessing Expectations of a Bogus Professor

Item	M (SD)	Cronbach's α	1	2	3	4
1. Dedication	2.83 (.60)	.91	—			
2. Attractiveness	2.50 (1.01)	.93	.38**	—		
3. Enhancement	2.44 (.71)	.88	.63**	.37**	—	
4. Fairness	3.13 (.65)	.79	.61**	.05	.32**	—
5. Clarity	2.43 (.65)	.65	.62**	.06	.52**	.49**

** $p < .01$.

Evaluation of current professor. Participants were also asked to evaluate their current psychology professor. The psychology professor was selected as the target because all participants were enrolled in a psychology class; however, their other coursework would vary. Participants rated the professor using the same items used to assess interpretations of online ratings of a bogus professor (see original stem items listed in Table 1). Specifically, students indicated how often their current professor displayed the characteristic in question using a scale from 1 (*never or almost never*) to 4 (*always or almost always*). For consistency, the same composite variables were computed. The investigators did conduct PCA with these items as well to ensure that similar factors emerged. Considerable overlap with the original PCA conducted with the responses for the bogus professor was evident in these results. Descriptive statistics, internal consistency, and bivariate correlations for these variables addressing evaluation of the current psychology professor are presented in Table 3.

Table 3. Descriptive Statistics and Correlations Among Variables Assessing Evaluation of an Actual Professor

Item	M (SD)	Cronbach's α	1	2	3	4
1. Dedication	3.61 (.46)	.90	—			
2. Attractiveness	2.65 (.79)	.85	.40**	—		

3. Enhancement	3.05 (.66)	.81	.56**	.54**	—	—
4. Fairness	3.54 (.55)	.76	.72**	.34**	.48**	—
5. Clarity	3.58 (.63)	.83	.71**	.42**	.52**	.59**

** $p < .01$.

Results

Interpretation of Online Professor Ratings

To examine hypothesized differences in participants' expectations of a bogus professor based on the professor's gender and designated hotness, a multivariate analysis of covariance (MANCOVA) was conducted with the five composite dependent variables (dedication, attractiveness, enhancement, fairness, and clarity) and two factor or grouping variables: professor gender (man versus woman) and hotness (chili pepper versus no chili pepper). Participant sex was included as a covariate. MANCOVA was conducted in place of a series of analysis of covariance (ANCOVA) due to significant correlation between most of the five dependent variables ($r = .05$ to $r = .63$; $p < .01$ for 8 out of 10 correlations; see Table 2). The assumption of homogeneity of covariances was examined with Box's test of equality of covariance matrices, and results were not significant ($p = .26$), suggesting the assumption had not been violated. However, the assumption of normality did appear to be violated for all dependent variables based on significant Kolmogorov–Smirnov and Shapiro–Wilk tests and evident negative skew (with scores situated at higher values) in histograms. Because ANOVA procedures are considered robust to violations of the normality assumption (Field, 2013; Mertler & Vannatta, 2005), analyses proceeded, but results should be interpreted with caution.

The main effect of professor gender was not statistically significant (Wilks' $\lambda = .99$; $F = .45$; $p = .81$); however, the main effect of hotness was statistically significant (Wilks' $\lambda = .53$; $F = 38.16$; $p = .00$; $\eta^2_p = .47$, or large effect size). The interaction effect for Gender \times Hotness was explored, but it was not significant.

Follow-up tests were conducted using a series of univariate ANCOVA. For the hotness effect, differences at the univariate level were significant for attractiveness ($F = 148.17$; $p = .00$; $\eta^2_p = .40$, or large effect size), which was not surprising. Less anticipated, there were significantly different expectations of professor clarity based on hotness ($F = 9.80$; $p = .00$; $\eta^2_p = .04$, or small effect size). Examination of the means (see Figure 1) revealed that participants expected Alex Johnson to be more attractive and lower in clarity when a chili pepper was included in the online professor rating summary presented.

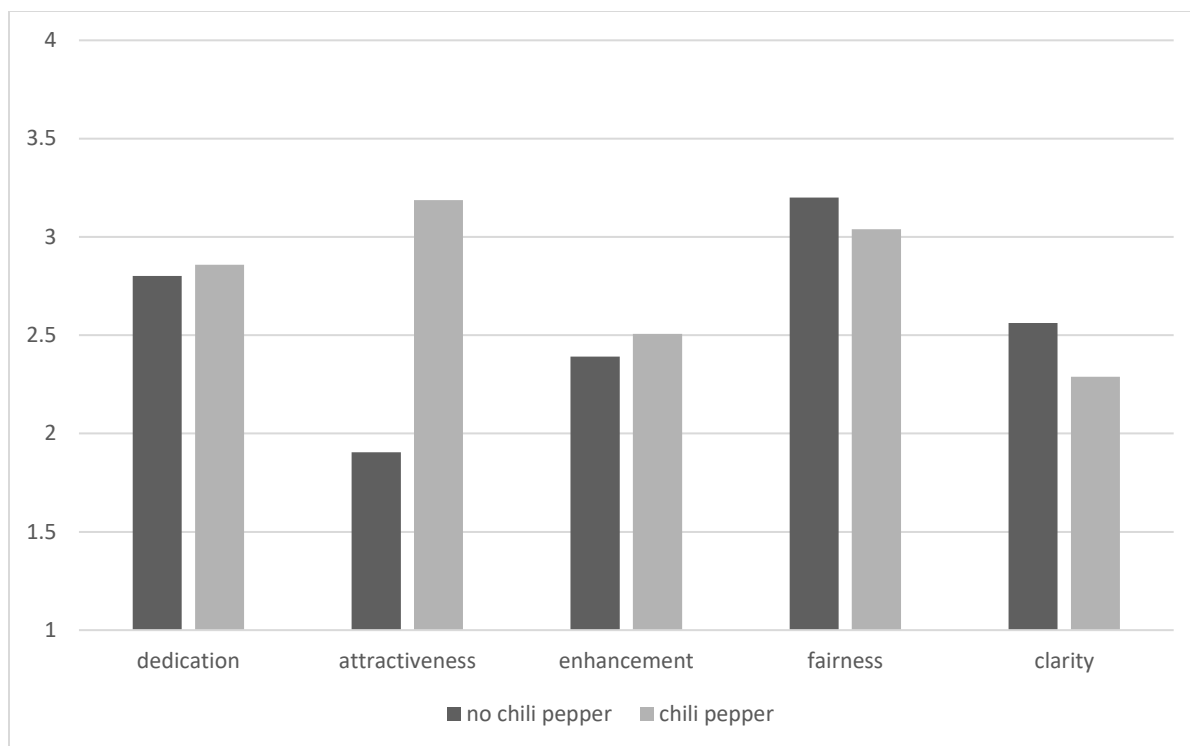


Figure 1. Mean expectations of bogus professor designated as either hot (chili pepper) or not hot (no chili pepper).

Rating of Current Professor

Examination of the effects of professor gender (man versus woman), perceived age (under 35 versus 35 or older), and student-rated attractiveness (at or below the median versus above the median) on ratings of four of the current psychology professor's apparent qualities (dedication, enhancement, fairness, and clarity) also was conducted with MANCOVA, with both participant sex and current age in years included as covariates. For this analysis, attractiveness was excluded as a dependent variable because it was used as a grouping variable. The assumption of homogeneity of covariances was again examined with Box's test of equality of covariance matrices, but results were significant ($p = .00$) for this analysis, suggesting the assumption had been violated. Box's test is known to be highly sensitive and unequal cell sizes can increase the likelihood of obtaining significant results. For this analysis, ensuring equal cell sizes was difficult because existing professor qualities were being evaluated. Specifically, there were 104 participants reporting having men as professors compared to 113 reporting women professors, and 94 participants believing their professor to be under 35 compared to 123 perceiving their professor as 35 or older. In addition, the assumption of normality did once again appear to be violated for all dependent variables based on significant Kolmogorov–Smirnov and Shapiro–Wilk tests and evident negative skew (with scores situated at higher values) in histograms. Because ANOVA procedures are considered robust to these violations (Field, 2013; Mertler & Vannatta 2005), analyses proceeded as planned but results should still be interpreted cautiously.

Participants' ratings of their current psychology professor significantly differed by professor gender (Wilks' $\lambda = .76$; $F = 13.17$; $p = .00$; $\eta^2_p = .24$, or large effect size) and perceived age (Wilks' $\lambda = .82$; $F = 9.39$; $p = .00$; $\eta^2_p = .19$, or large effect size), but the interaction between the two variables also was statistically significant (Wilks' $\lambda = .92$; $F = 3.47$; $p = .01$; $\eta^2_p = .08$, or

medium effect size); therefore, the main effects will not be interpreted further. The follow-up ANCOVA indicated that the significant difference was on ratings of professor clarity ($F = 7.59$; $p = .01$; $\eta^2_p = .04$, or small effect size). As illustrated in Figure 2, younger professors were generally rated higher in clarity, and men in the 35 or older category were rated as less clear than women in that age group. There was also a significant main effect for professor attractiveness (Wilks' $\lambda = .87$; $F = 7.87$; $p = .00$; $\eta^2_p = .13$, or large effect size). As illustrated in Figure 3 and supported by follow-up ANCOVA, professors considered more attractive (above the median of 2.60) were rated significantly higher on dedication, enhancement, fairness, and clarity.

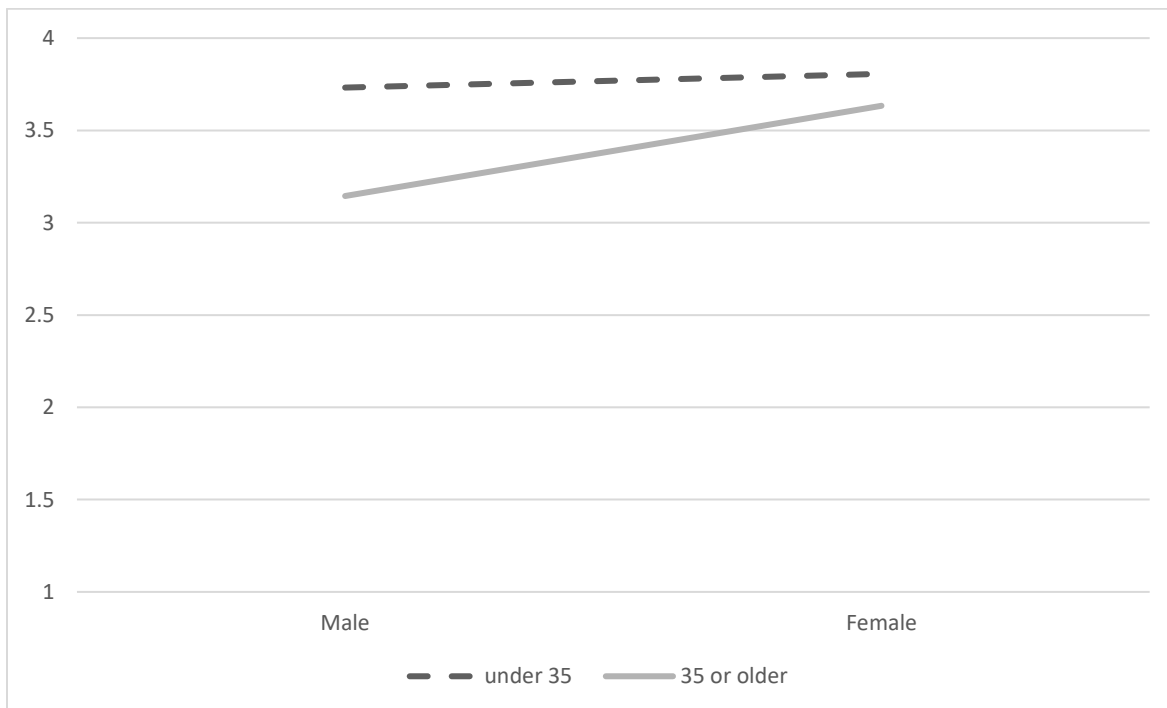


Figure 2. Interaction between current professor sex and apparent age in affecting students' evaluations of clarity.

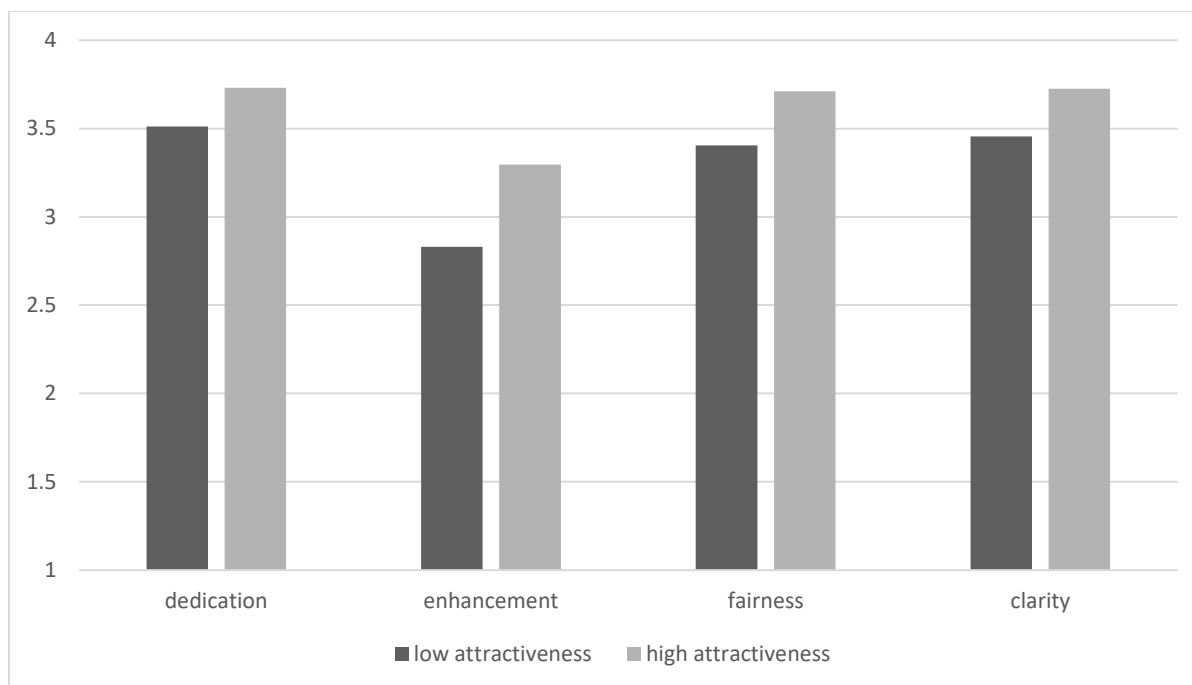


Figure 3. Mean ratings of current professor grouped by professor attractiveness.

Discussion

The current investigation addressed several hypotheses regarding possible sources of bias in college students' interpretation and creation of online professor ratings such as those available on RMP. Findings revealed that online professor ratings may be swayed by a variety of personal features of professors, many of them being, for the most part, out of the professors' control. The investigators obtained support for bias stemming from professor gender, age, and hotness, with specific findings not necessarily matching with previous findings.

Gender and Age of Professor

Results of the current investigation supported student bias in favor of women professors when rating an actual professor but not when interpreting or forming expectations from an online professor rating. Professor gender interacted with professor age when evaluating a current psychology professor on clarity. Specifically, male professors believed to be 35 or older received the lowest clarity ratings compared to younger men and all women evaluated. Given that the sample was largely female and young, this finding may stem from dissimilarity between participants and older men as professors; younger female participants may perceive older men as too dissimilar from themselves. This is consistent with previous research that individuals are more dismissive of information that they receive from others they perceive to be different from themselves (Wheless, 1974).

Additional explanations have been proposed for both gender- and age-related bias against professors. For example, age dissimilarity may produce a communication gap with undergraduates; this gap is especially wide for older faculty members who may have less familiarity with widely used technology and well-known examples from popular culture (Gehrt et al., 2015). Many digital technology-native university students (similar to the current sample) currently expect professors to incorporate PowerPoint slideshows, social media, and occasional

YouTube clips into lectures and to post supplemental material online (Borboa, Joseph, & Spake, 2012; Griesemer, 2011; Saeed, Yun, & Sinnappan, 2009); however, older faculty members may have developed their preferred teaching style and format when such features were not available or commonplace. Studies have indicated that older faculty members are more reluctant to implement computer-based teaching methods (Rousseau & Rogers, 1998), and faculty members in general report little to no formal training in implementing enhancements such as web-based instruction (Vodanovich & Piotrowski, 2005). While it is completely possible to deliver a quality lecture using only chalk or a dry-erase marker and a mounted wall board, students nonetheless may prefer notes and supplements accessible through electronic devices and examples linking course material to the video clips, memes, and streaming programs they commonly view on the same devices outside of class.

In addition to dissimilarity, the bias against older men as professors may exist because there are more women teaching on college campuses than in the past. Previous evidence for preference for professors who are men (see Basow & Silberg, 1987) may reflect an earlier time when women professors were more of a novelty. Current university students in the United States are likely exposed to a mixture of men and women as professors, potentially eroding the once dominant stereotype of the university professor as male authority figure. Furthermore, the higher number of females in the sample may have further swayed findings because women have been shown to particularly value female faculty over male faculty (Bachen, McLoughlin, & Garcia, 1999; Basow, 2000). The body of evidence suggests a complex array of influencing factors at work and beyond professors' command, with physical attractiveness further expanding the list of advantageous characteristics professors may be fortunate to possess out of sheer luck or effort unrelated to teaching ability.

Professor Hotness

Support for bias based on professor hotness emerged as well, without all obtained results fitting the expectations. It was hardly shocking that participants anticipated a professor designated as hot to be higher in attractiveness. Such a result implies that participants noticed the chili pepper and believed it to be accurate. More surprising was the finding that hot professors were expected to be lower in clarity, suggesting a negative effect of hotness. That is, hot professors were expected to look good but not teach as well in at least one domain. This result is in direct opposition to that of Bonds-Raacke and Raacke (2007), who reported higher ratings of clarity in professors rated higher in attractiveness.

The present finding is among a growing body suggesting that the "what is beautiful is good" stereotype may be more complex or in need of exceptions than previously believed. Other evidence for this includes Chia, Allred, Grossnickle, and Lee's (1998) finding that, based on photographs of both attractive and unattractive men and women, unattractive men were rated highest in ability, while unattractive women were rated lowest in ability. In addition, Mehng (2015) observed a negative effect of attractiveness on perceived competence in the presence of low warmth. This finding may reflect what some have labeled the "beauty is beastly" effect (see Johnson, Podratz, Dipboye, & Gibbons, 2010). That is, there is a small body of evidence that attractiveness can be detrimental (for women, specifically) in some situations. In particular, physically attractive women are rated more negatively when applying for jobs perceived as more masculine in nature and for which physical appearance is deemed unimportant. The investigators did not find support for an interaction effect between gender and hotness, though, casting doubt on the "beauty is beastly" effect operating as it is typically described. Replication and further expansion of this research is clearly called for to better understand why college students might expect less clarity from professors having a chili pepper.

The results of professor hotness on ratings of current professors are more consistent with prior work. Professors rated as more attractive were rated higher in all other areas. In addition to the possibility that students indeed perceived more attractive professors as more competent, it is possible that well-liked professors were rated favorably in all areas assessed regardless of their actual level of hotness. This is consistent with previous research indicating that RMP attractiveness ratings are influenced by students' positive illusions of professors (Theyson, 2015). Additionally, because the grouping for this analysis resulted from a quasi-experimental instead of true experimental design, the authors cannot say for certain whether attractiveness influenced the other professor ratings in a unidirectional manner. Furthermore, because attractiveness was not rated by objective outside observers, it is certainly possible that less attractive professors with other strong teaching qualities may have become more attractive to students over time. Still, given that many aspects of hotness are out of one's control, and more so with age, results are concerning considering that professors not graced with hotness may be unduly penalized for a feature not intrinsically linked to actual teaching ability. Likewise, higher ratings assigned to professors lucky enough to possess pleasing physical features may place them at an advantage in competing for student enrollment, faculty positions, teaching awards, tenure, and promotion.

Limitations and Future Directions

This investigation possessed several limitations. First, the sample was limited to students in lower level psychology courses. Given that course level (Goldberg & Callahan, 1991; Moritsch & Suter, 1988; Patrick, 2011) and discipline (Cashin, 1992; Patrick, 2011) may bias student evaluations, results may not generalize to students taking upper level classes or those from other disciplines. The unequal representation of male and female students prevented comparisons based on participant sex, though sex was included as a covariate in all analyses. The current study also did not include professor race as a variable because the current psychology professors being evaluated were all White. Previous research suggests racial-minority faculty members are rated most harshly on RMP, and that Black men as professors are rated particularly negatively (Reid, 2010); therefore, race should be included in analyses when possible.

Convenience sampling presents an additional limitation of the current study. Most participants identified as White. This may limit generalizability of results to larger, more ethnically diverse college groups. Future research could extend the current study to college campuses with greater ethnic and racial diversity in their student body. Moreover, the average participant age in the sample was 19 years old, seating this sample in the traditional college student age range (National Center for Educational Statistics, 2016). Findings from this largely traditional-aged sample may not generalize well to nontraditional students. This may be particularly applicable to findings related to technology use in the classroom; nontraditional college students may differ in the importance that they place on classroom technology use compared to digital-native, traditional-aged college students.

Additionally, geography may limit the generalizability of the findings. Participants were recruited from a public university in the southern United States; however, RMP evaluations represent professors across the United States as well as Canada and the United Kingdom. Future investigations may benefit from recruitment of participants from all regions represented on RMP to determine if findings from an American sample generalize to student populations in other countries with RMP. Additionally, it is important to learn if these findings would also generalize to students who use similar services in countries where RMP is currently unavailable. These students include those who use the Rate My Teachers Republic of Ireland, Rate My Teachers Australia, and Rate My Teachers New Zealand websites. Despite differences in rating site used, these findings suggest that student ratings available through unofficial, non-university-affiliated

sources (e.g., RMP, Rate My Teachers, social media) have the potential to impact students' perceptions of their professors. This should be taken seriously given that these perceptions, formed before students ever meet professors, can impact students' motivations for a course and course-related behavior (Edwards et al., 2007; Kowai-Bell et al., 2011).

Because the measures used in hypothesis testing were developed specifically for this study, it is optimal for replication to take place, including attempts to more firmly establish reliability and validity of measures. Alternate means of measuring relevant variables should be explored, too. For example, when examining the effect of hotness on evaluations of current professors, it would be desirable to have trained outside coders objectively rate physical attractiveness. Another key feature to assess would be professors' actual teaching style, vocal quality, organization, and evident personality features. Again, this information would best be assessed by objective coders unfamiliar with the professors.

Implications

Taken together, and keeping these limitations in mind, these findings indicate a multifaceted and complicated association between teaching competence and student evaluations. Moreover, they suggest cause for concern regarding the widespread and increasing use of online professor ratings by students and university administration. These seemingly fun ratings have the potential to influence course enrollment, sway expectations of students who do enroll in a given course, and tarnish a professor's reputation. Alarming, these negative consequences may not result from poor teaching so much as largely uncontrollable factors such as a professor not being the preferred gender, not looking or acting young enough, or not being as easy on the eyes as peers or competitors. Sites like RMP may rate professors, but findings like these beg us to ask if they truly do rate teaching or have any appropriate place in students' or university administrators' decision making.

Despite these concerns about the validity of RMP ratings, this remains: They can and sometimes do influence decision making. For example, students perceive these evaluations as credible tools to inform their education-related decisions (Field, Bergiel, & Viosca, 2008; Davison & Price, 2009; Hayes & Prus, 2014; Landry et al., 2010). The influence of RMP-style ratings on students' decision making suggests that faculty members should not be hasty and completely reject their content, despite the ratings' biases. Further supporting the case that RMP content merits some faculty attention is evidence that RMP narrative content is pertinent to teaching (Otto et al., 2008) and focused on instruction-related characteristics such as content knowledge, clarity in communication, and organization (Hartman & Hunt, 2013; Kindred & Mohammed, 2005; Silva et al., 2008).

Although RMP ratings are not intended to provide formative feedback to faculty, some may be nonetheless interested in exploring their ratings. We suggest that faculty members seeking to garner formative information from RMP ratings approach them with similar techniques utilized for formal, university-administered student evaluations. For example, Buskist and Hogan (2010) recommend a systematic approach to interpretation of student ratings. First, remove all comments that are irrelevant to course content or teaching (e.g., "Her clothes are ugly"). Next, remove all comments that lack concrete, specific information about teaching or the course (e.g., "She is super" or "I think it's stupid that we have to take these classes to graduate"). Then, group comments into two categories: (a) aspects of teaching and course content that can be changed (e.g., "Assignments would work better in a different sequence") and (b) aspects of teaching and course content that cannot be changed (e.g., "This subject has a lot of technical information"). Aspects of teaching and content that can be changed may also be further categorized as (i) things

useful to change and (ii) things not useful to change. For example, comments such as “I hate that we have to do outside reading” reflect pedagogically useful components of a course despite their unpopularity with some students.

In addition to their influence upon student decision making, RMP ratings may also influence administrator decision making regarding hiring, tenure, and promotion. We caution, however, their use in this way. Given that course easiness influences RMP's overall quality ratings (Felton et al., 2008), it may be tempting to “water down” or reduce the rigor of one's course to improve one's overall quality ratings. This concern regarding reduced rigor extends to the results of formal, university-administered student evaluations of teaching as well (Zabaleta, 2007). When the results of any form of student evaluation, formal or informal, are used in summative hiring, tenure, and promotion decisions, it is important to include other assessments of teaching effectiveness, for example, peer and supervisor teaching evaluations as well as teaching portfolios (Marsh, 1984; Marsh & Roche, 1997; Zabaleta, 2007).

References

- Abel, M. H., & Meltzer, A. L. (2007). Student ratings of a male and female professors' lecture on sex discrimination in the workforce. *Sex Roles, 57*, 173–180. doi:10.1007/s11199-007-9245-x
- Adamson, G., O'Kane, D., & Shevlin, M. (2005). Students' ratings of teaching effectiveness: A laughing matter? *Psychological Reports, 96*, 225–226. doi:10.2466/PRO.96.1.225-226
- Ahmadi, M., Helms, M. M., & Raiszadeh, F. (2001). Business students' perceptions of faculty evaluations. *International Journal of Educational Management, 15*, 12–22. doi:10.1108/09513540110366097
- Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64*, 431–441. doi:10.1037/0022-3514.64.3.431
- Arbuckle, J., & Williams, B. D. (2003). Students' perceptions of expressiveness: Age and gender effects on teacher evaluations. *Sex Roles, 49*, 507–516. doi: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. doi:10.1080/03634529909379169
- Basow, S. A. (2000). Best and worst professors: Gender patterns in students' choices. *Sex Roles, 43*, 407–417. doi:10.1023/A:1026655528055
- 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. doi:10.1037/0022-0663.79.3.308
- 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. doi:10.1037/0022-0663.74.2.170
- Bianchini, S., Lissoni, F., & Pezzoni, M. (2013). Instructor characteristics and students' evaluation of teaching effectiveness: Evidence from an Italian engineering school. *European Journal of Engineering Education, 38*, 38–57. doi:10.1080/03043797.2012.742868

- Bonds-Raacke, J., & Raacke, J. D. (2007). The relationship between physical attractiveness of professors and students' ratings of professor quality. *Journal of Psychiatry, Psychology and Mental Health, 1*, 1–7. Retrieved from <http://www.scientificjournals.org/journals2007/articles/1227.pdf>
- Borboa, D., Joseph, M., & Spake, D. F. (2012). Use of technology in higher education: A preliminary analysis of student views. In K. J. Shanahan (Ed.), *Annals of the Society for Marketing Advances, Volume 1* (pp. 157–158). Montgomery, AL: Society for Marketing Advances. Retrieved from <http://c.ymcdn.com/sites/www.marketingadvances.org/resource/resmgr/imported/2012%20Conference%20Proceedings.pdf>
- Boswell, S. S. (2016). RateMyProfessors is hogwash (but I care): Effects of RateMyProfessors and university-administered teaching evaluations on professors. *Computers in Human Behavior, 56*, 155–162. doi:10.1016/j.chb.2015.11.045
- Buck, S., & Tiene, D. (1989). The impact of physical attractiveness, gender, and teaching philosophy on teacher evaluations. *The Journal of Educational Research, 82*, 172–177. doi:10.1080/00220671.1989.10885887
- Burns-Glover, A. L., & Veith, D. J. (1995). Revisiting gender and teaching evaluations: Sex still makes a difference. *Journal of Social Behavior and Personality, 10*, 69–80.
- Buskist, C., & Hogan, J. (2010). She needs a haircut and a new pair of shoes: Handling those pesky course evaluations. *Journal of Effective Teaching, 10*, 51–56.
- Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology, 76*, 839–855. doi:10.1037/0022-3514.76.5.839
- Cahyadi, V. (2004). The effect of interactive engagement teaching on student understanding of introductory physics at the faculty of engineering, University of Surabaya, Indonesia. *Higher Education Research and Development, 23*, 455–464. doi:10.1080/0729436042000276468
- Carnevale, D. (2006). University of Saskatchewan fires tenured professor accused of maligning colleagues on RateMyProfessors web site. *The Chronicle of Higher Education, 52*, A28.
- Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review, 97*, 19–35. doi:10.1037/0033-295X.97.1.19
- Cashin, W. E. (1992). Student ratings: The need for comparative data. *Instructional Evaluation and Faculty Development, 12*, 146.
- Chia, R. C., Allred, L. J., Grossnickle, W. F., & Lee, G. W. (1998). Effects of attractiveness and gender on the perception of achievement-related variables. *The Journal of Social Psychology, 138*, 471–477. doi:10.1080/00224549809600401
- Clayson, D. E., & Sheffet, M. J. (2006). Personality and the student evaluation of teaching. *Journal of Marketing Education, 28*, 149–160. doi:10.1177/0273475306288402
- Coladarci, T., & Kornfield, I. (2007). RateMyProfessors.com versus formal in-class student evaluations of teaching. *Practical Assessment, Research & Evaluation, 12*, 1–15.
- Davison, E., & Price, J. (2009). How do we rate? An evaluation of online student evaluations. *Assessment & Evaluation in Higher Education, 34*, 51–65. doi:10.1080/02602930801895695
-

-
- Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, 24, 285–290. doi:10.1037/h0033731
- Divoky, J. J., & Rothermel, M. A. (1988). Student perceptions of the relative importance of dimensions of teaching performance across type of class. *Educational Research Quarterly*, 12, 40–45.
- Eadie, B. (1996). Instructor credibility and style of dress: Performance in groups at an Irish gay bar. *Spectra*, 32, 11–13.
- Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, 110, 109–128. doi:10.1037/0033-2909.110.1.109
- Edwards, C., Edwards, A., Qing, Q., & Wahl, S. T. (2007). The influence of computer-mediated word-of-mouth communication on student perceptions of instructors and attitudes toward learning course content. *Communication Education*, 56, 255–277. doi:10.1080/03634520701236866
- Edwards, A., Edwards, C., Shaver, C., & Oaks, M. (2009). Computer-mediated word-of-mouth communication on Ratemyprofessors.com: Expectancy effects on student cognitive and behavioral learning. *Journal of Computer-Mediated Communication*, 14, 368–392. doi:10.1111/j.1083-6101.2009.01445.x
- Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. *Research in Higher Education*, 9, 199–242. doi:10.1007/BF00976997
- Feldman, K. A. (1983). Seniority and instructional experience of college teachers as related to the evaluations they receive from their students. *Research in Higher Education*, 18, 3–124. doi:10.1007/BF00992080
- Feldman, K. A. (1986). The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: A review and synthesis. *Research in Higher Education*, 24, 139–213. doi:10.1007/BF00991885
- Felton, J., Koper, P. T., Mitchell, J., & Stinson, M. (2008). Attractiveness, easiness and other issues: Student evaluations of professors on ratemyprofessors.com. *Assessment & Evaluation in Higher Education*, 33, 45–61. doi:10.1080/02602930601122803
- 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. doi:10.1080/0260293032000158180
- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). London, United Kingdom: Sage.
- Field, J. R., Bergiel, B. J., & Viosca, R. C. (2008). An exploratory study of student use and perceived accuracy of faculty evaluations on Ratemyprofessors.com. *Competition Forum*, 6, 378–383.
- Freng, S., & Webber, D. (2009). Turning up the heat on online teaching evaluations: Does "hotness" matter? *Teaching of Psychology*, 36, 189–193.
- Frenzel, A. C., Goetz, T., Ludtke, O., Pekrun, R., & Sutton, R. E. (2009). Emotional transmission in the classroom: Exploring the relationship between teacher and student enjoyment. *Journal of Educational Psychology*, 101, 705–716. doi:10.1037/a0014695
- Garner, R. L. (2006). Humor in pedagogy: How ha-ha can lead to aha! *College Teaching*, 54, 177–180. doi:10.3200/CTCH.54.1.177-180
-

- Gehrt, K., Louie, T. A., & Osland, A. (2015). Student and professor similarity: Exploring the effects of gender and relative age. *Journal of Education for Business, 90*, 1–9. doi:10.1080/08832323.2014.968514
- Goldberg, G., & Callahan, J. (1991). Objectivity of student evaluations of instructors. *Journal of Education for Business, 66*, 377–378. doi:10.1080/08832323.1991.10117505
- Gotlieb, J. (2011). Classroom entertainment, grades and student evaluations of business professors. *Journal of the Academy of Business Education, 12*, 118–136.
- Griesemer, J. A. (2011). Using social media to enhance undergraduate business students learning experiences. In *Proceedings of the Northeast Business & Economics Association* (pp. 201–204). Philadelphia, PA: Northeast Business & Economics Association, Inc.
- Hamermesh, D., & Parker, A. (2005). Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review, 24*, 369–376.
- Hart, J., & Driver, J. (1978). Teacher evaluation as a function of student and instructor personality. *Teaching of Psychology, 5*, 198–200. doi:10.1207/s15328023top0504_8
- Hartman, K. B., & Hunt, J. B. (2013). What RateMyProfessors.com reveals about how and why students evaluate their professors: a glimpse into the student mindset. *Marketing Education Review, 23*, 151e161. doi:10.2753/MER1052-8008230204
- Hayes, M. W., & Prus, J. (2014). Student use of quantitative and qualitative information on RateMyProfessors.com for course selection. *College Student Journal, 48*, 675–688.
- Hossain, T. M. (2009). Hot or not: An analysis of online professor-shopping behavior of business students. *Journal of Education for Business, 85*, 165–171. doi:10.1080/08832320903252439
- Howard, C. (2014). Ranking America's top colleges 2014. *Forbes*. Retrieved from <http://www.forbes.com/sites/carolinehoward/2014/07/30/ranking-americas-top-colleges-2014/>
- Huntsinger, J. R., & Smith, C. T. (2009). First thought, best thought: Positive mood maintains and negative mood degrades implicit-explicit attitude correspondence. *Personality and Social Psychology Bulletin, 35*, 187–197. doi:10.1177/0146167208327000
- Jenkins, S.J., & Downs, E. (2001). Relationship between faculty personality and student evaluation of courses. *College Student Journal, 35*, 636–640.
- Johnson, R. R., & Crews, A. D. (2013). My professor is hot! Correlates of RateMyProfessors.com ratings for criminal justice and criminology faculty members. *American Journal of Criminal Justice, 38*, 639–656. doi:10.1007/s12103-012-9186-y
- Johnson, S. K., Podratz, K. E., Dipboye, R. L., & Gibbons, E. (2010). Physical attractiveness biases in ratings of employment suitability: Tracking down the “beauty is beastly” effect. *The Journal of Social Psychology, 150*, 301–318. doi:10.1080/00224540903365414
- Joye, S., & Wilson, J. H. (2015). Professor age and gender affect student perceptions and grades. *Journal of the Scholarship of Teaching and Learning, 15*, 126–138. doi:10.14434/josotl.v15i4.13466
- Karpen, S. C., Jia, L., & Rydell, R. J. (2012). Discrepancies between implicit and explicit attitude measures as an indicator of attitude strength. *European Journal of Social Psychology, 42*, 24–29. doi:10.1002/ejsp.849
-

-
- Keiner, L. E., & Gilman, C. S. (2015). Implementation of interactive engagement teaching methods in a physical oceanography course. *Journal of College Science Teaching, 45*, 70–77.
- Kindred, J., & Mohammed, S. N. (2005). "He will crush you like an academic ninja!" Exploring teacher ratings on Ratemyprofessors.com. *Journal of Computer-Mediated Communication, 10*, 9. doi:10.1111/j.1083-6101.2005.tb00257.x
- Kinney, D. P., & Smith, S. P. (1992). Age and teaching performance. *Journal of Higher Education, 63*, 282–302. doi:10.2307/1982016
- Kohn, J., & Hatfield, L. (2006). The role of gender in teaching effectiveness ratings of faculty. *Academy of Educational Leadership Journal, 10*, 121–137. Retrieved from
- Kowai-Bell, N., Guadagno, R. E., Little, T., Preiss, N., & Hensley, R. (2011). Rate my expectations: How online evaluations of professors impact students' perceived control. *Computers in Human Behavior, 27*, 1862–1867. doi:10.1016/j.chb.2011.04.009
- Landry, N. M., Kurkul, K. E., & Poirier, C. R. (2010). Online faculty rating sites: Examining how students perceive and use Ratemyprofessors.com. *Psi Chi Journal of Undergraduate Research, 15*, 139–142.
- Lawson, R. A., & Stephenson, E. F. (2005). Easiness, attractiveness, and faculty evaluations: Evidence from ratemyprofessors.com. *Atlantic Economic Journal, 33*, 485–486. doi:10.1007/s11293-005-2873-z
- Legg, A. M., & Wilson, J. H. (2012). RateMyProfessors.com offers biased evaluations. *Assessment & Evaluation in Higher Education, 37*, 89–97. doi:10.1080/02602938.2010.507299
- Lewandowski Jr, G. W., Higgins, E., & Nardone, N. N. (2012). Just a harmless website? An experimental examination of RateMyProfessors.com's effect on student evaluations. *Assessment & Evaluation in Higher Education, 37*, 987–1002. doi:10.1080/02602938.2011.594497
- Liu, J., Hu, J., & Furutan, O. (2013). The influence of student perceived professors' "hotness" on expertise, motivation, learning outcomes, and course satisfaction. *Journal of Education for Business, 88*, 94–100. doi:10.1080/08832323.2011.652695
- MacNell, L., Driscoll, A., & Hunt, A. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education, 40*, 291–303. Retrieved from <http://link.springer.com/journal/10755>
- Marsh, H. W. (1984) Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and utility. *Journal of Educational Psychology, 76*, 707–754.
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective. *American Psychologist, 52*, 1187–1197. doi:10.1037/0003-066X.52.11.1187
- Mehng, S. A. (2015). *When is physical attractiveness not beneficial? Perceptions of warmth and competence, emotions, and job behaviors* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis, Minnesota.
- Mertler, C. A., & Vannatta, R. A. (2005). *Advanced and multivariate statistical methods* (3rd ed.). Glendale, CA: Pyrczak.
- Messner, M. (2000). White guy habitus in the classroom: Challenging the reproduction of privilege. *Men and Masculinities, 2*, 457–469. doi:10.1177/1097184X00002004005
-

- Micari, M., & Pazos, P. (2012). Connecting to the professor: Impact of the student–faculty relationship in a highly challenging course. *College Teaching*, *60*, 41–47. doi:10.1080/87567555.2011.627576
- Montell, G. (2006, September 27). The art of the bogus rating. *The Chronicle of Higher Education*. Retrieved from <http://chronicle.com/article/The-Arts-of-the-Bogus-Rating/46887/>
- Moritsch, B. G., & Suter, W. N. (1988). Correlates of halo error in teacher evaluation. *Educational Research Quarterly*, *12*, 29–34.
- National Center for Educational Statistics. (2016). Nontraditional undergraduates: Definition and data. Retrieved from <https://nces.ed.gov/pubs/web/97578e.asp>
- O'Toole, D. M., Spinelli, M. A., & Wetzel, J. N. (2000). The important learning dimensions in the school of business: A survey of students and faculty. *Journal of Education for Business*, *75*, 338–342. doi:10.1080/08832320009599037
- Otto, J., Sanford, D. J., & Ross, D. N. (2008). Does ratemyprofessor.com really rate my professor? *Assessment & Evaluation in Higher Education*, *33*, 1–16. doi:10.1080/02602930701293405
- Pannapacker, W. A. (2007). The inescapability of your past. *The Chronicle of Higher Education*, *53*, B11.
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*, 91–105. doi:10.1207/S15326985EP3702_4
- Petchers, M. K., & Chow, J. C. (1988). Sources of variations in students' evaluations of instruction in a graduate social work program. *Journal of Social Work Education*, *24*, 35–42. doi:10.1080/10437797.1988.10672095
- Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, *21*, 381–391.
- Rate My Professors (RMP). (2016a). About RateMyProfessors.com. Retrieved from <http://www.ratemyprofessors.com/About.jsp>
- Rate My Professors (RMP). (2016b). How do you tally your ratings? Retrieved from <http://www.ratemyprofessors.com/help.jsp#tally>
- Rate My Professors (RMP). (2016c). Why didn't the chili pepper I gave my professor show up? Retrieved from <http://www.ratemyprofessors.com/help.jsp#hotness>
- 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*, 137–152. doi:10.1037/a0019865
- Richmond, A. S., Berglund, M. B., Epelbaum, V. B., & Klein, E. M. (2015). $a + (b_1) \text{ professor–student rapport} + (b_2) \text{ humor} + (b_3) \text{ student engagement} = (\hat{Y}) \text{ student ratings of instructors}$. *Teaching of Psychology*, *42*, 119–125. doi:10.1177/0098628315569924
- 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. doi:10.3200/GENP.133.1.19-35
- Romano, S. T., & Bordieri, J. E. (1989). Physical attractiveness stereotypes and students' perceptions of college professors. *Psychological Reports*, *64*, 1099–1102. doi:10.2466/pr0.1989.64.3c.1099
-

- Rousseau, G. K., & Rogers, W. A. (1998). Computer usage patterns of university faculty members across the life span. *Computers in Human Behavior*, *14*, 417–428. doi:10.1016/S0747-5632(98)00014-4
- Rydell, R. J., McConnell, A. R., & Mackie, D. M. (2008). Consequences of discrepant explicit and implicit attitudes: Cognitive dissonance and increased information processing. *Journal of Experimental Social Psychology*, *44*, 1526–1532. doi:10.1016/j.jesp.2008.07.006
- Saeed, N., Yun, Y., & Sinnappan, S. (2009). Emerging web technologies in higher education: A case of incorporating blogs, podcasts and social bookmarks in a web programming course based on students' learning styles and technology preferences. *Journal of Educational Technology & Society*, *12*, 98–109.
- Scherr, F. C., & Scherr, S. S. (1990). Bias in student evaluations of teacher effectiveness. *Journal of Education for Business*, *65*, 356–358.
- Sebastian, R. J., & Bristow, D. (2008). Formal or informal? The impact of style of dress and forms of address on business students' perceptions of professors. *Journal of Education for Business*, *83*, 196–201. doi:10.3200/JOEB.83.4.196-201
- 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. doi:10.1080/00986280801978434
- Sinclair, L., & Kunda, Z. (2000). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. *Personality and Social Psychology Bulletin*, *26*, 1329–1342. doi:10.1177/0146167200263002
- Smith, C. T., & Nosek, B. A. (2011). Affective focus increases the concordance between implicit and explicit attitudes. *Social Psychology*, *42*, 300–313. doi:10.1027/1864-9335/a000072
- Sonntag, M. E., Bassett, J. F., & Snyder, T. (2009). An empirical test of the validity of student evaluations of teaching made on RateMyProfessors.com. *Assessment & Evaluation in Higher Education*, *34*, 499–504. doi:10.1080/02602930802079463
- Stonebraker, R., & Stone, G. (2015). Too old to teach? The effect of age on college and university professors. *Research in Higher Education*, *56*, 793–812. doi:10.1007/s11162-015-9374-y
- Stuber, J. M., Watson, A., Carle, A., & Staggs, K. (2009). Gender expectations and on-line evaluations of teaching: Evidence from RateMyProfessors.com. *Teaching in Higher Education*, *14*, 387–399. doi:10.1080/13562510903050137
- Theyson, K. C. (2015). Hot or not: The role of instructor quality and gender on the formation of positive illusions among students using RateMyProfessors.com. *Practical Research and Assessment*, *20*. Retrieved from <http://pareonline.net/getvn.asp?v=20&n=4>
- Timmerman, T. (2008). On the validity of RateMyProfessors.com. *Journal of Education for Business*, *84*, 55–61. doi:10.3200/JOEB.84.1.55-61
- Torok, S. E., McMorris, R. F., & Lin, W. (2004). Is humor an appreciated teaching tool? Perceptions of professors' teaching styles and use in humor. *College Teaching*, *52*, 14–20. doi:10.3200/CTCH.52.1.14-20
- Vodanovich, S. J., & Piotrowski, C. (2005). Faculty attitudes toward web-based instruction may not be enough: Limited use and obstacles to implementation. *Journal of Educational Technology Systems*, *33*, 309–318. doi:10.2190/V2N7-DMC4-2JWB-5Q88

Walsh, D. J., & Maffei, M. J. (1994). Never in a class by themselves: An examination of behaviors affecting the student–professor relationship. *Journal on Excellence in College Teaching*, 5, 23–49.

Wheeless, L. R. (1974). The effects of attitude, credibility, and homophily on selective exposure to information. *Speech Monographs*, 41, 329–338. Retrieved from <http://www.tandfonline.com/toc/rcmm19/41/4#.Vyzl7XoYE20>

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. doi:10.1037/0022-0663.80.1.118

Wilson, J., Beyer, D., & Monteiro, H. (2014). Professor age affects student ratings: Halo effect for younger teachers. *College Teaching*, 62, 20–24.

Wilson, R. (1998). New research casts doubt on value of student evaluations of professors. *Chronicle of Higher Education*, 44, A12–A14.

Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in Higher Education*, 12, 55–76.
