

# Hot or Not: Do Professors Perceived as Physically Attractive Receive Higher Student Evaluations?

TODD C. RINILOLO  
KATHERINE C. JOHNSON  
TRACY R. SHERMAN  
JULIE A. MISSO

*Department of Psychology  
Medaille College*

---

**ABSTRACT.** Previous research investigating the influence of perceived physical attractiveness on student evaluations of college professors has been limited to a handful of studies. In this study, the authors used naturally occurring data obtained from the publicly available Web site [www.ratemyprofessors.com](http://www.ratemyprofessors.com). The data suggested that professors perceived as attractive received higher student evaluations when compared with those of a nonattractive control group (matched for department and gender). Results were consistent across 4 separate universities. Professors perceived as attractive received student evaluations about 0.8 of a point higher on a 5-point scale. Exploratory analyses indicated benefits of perceived attractiveness for both male and female professors. Although this study has all the limitations of naturalistic research, it adds a study with ecological validity to the limited literature.

**Key Words:** naturalistic research, physical attractiveness, student evaluations, teacher characteristics

---

**OUR PURPOSE IN THE PRESENT STUDY** was to investigate whether college professors perceived as physically attractive received higher student evaluations compared with colleagues that were perceived as nonattractive. To begin investigating this topic, we reviewed some relevant literature.

First, research results indicate that a variety of factors influence student evaluations of college professors (for pertinent reviews, see Cashin & Downey, 1992; Greenwald & Gillmore, 1997; Marsh & Roche, 1997; McKeachie, 1997). For

---

*The authors thank Janice Laben and Raymond Bailey. They also acknowledge the helpful comments by the action editor, F. Richard Ferraro, and three anonymous reviewers.*

*Address correspondence to Todd C. Riniolo, Department of Psychology, Medaille College, 18 Agassiz Circle, Buffalo, NY 14214; [triniolo@medaille.edu](mailto:triniolo@medaille.edu) (e-mail).*

example, in the Dr. Fox studies (Naftulin, Ware, & Donnelly, 1973; Ware & Williams, 1975), a professional actor (whose character was called Dr. Fox) was videotaped using high and low levels of expressiveness. When students viewed the taped lectures, Dr. Fox received higher evaluations when using the expressive style. Likewise, Williams and Ceci (1997), in a naturalistic study using an experienced professor who taught identical courses in the fall and spring semesters (a large section of Developmental Psychology), found an enthusiastic teaching style (while presenting the same course content) resulted in much higher student evaluations. Radmacher and Martin (2001), using hierarchical regression with a wide range of variables (professor's age and extraversion traits; student's current grades, gender, enrollment status, ACT scores, and age), found that professors' extraversion was the strongest predictor of midterm student evaluations of teaching effectiveness. However, students' enrollment status, current course grade, and age were also positively correlated with midterm student evaluations.

In addition, Freeman (1994), using three written descriptions of hypothetical professors (feminine, masculine, androgynous), found that students preferred both male and female professors who possess androgynous characteristics. Also, positive personality characteristics (e.g., caring, enthusiasm, sense of humor) were associated with higher student evaluations when undergraduates were asked to describe their best college professor (Basow, 2000; Waters, Kemp, & Pucci, 1988). Research results also indicate that perceived learning, prior interest in the subject (Marsh & Roche, 2000), students' expectations of grades (Greenwald & Gillmore, 1997; Millea & Grimes, 2002), nonverbal behavior (Ambady & Rosenthal, 1993), course workload (Marsh & Roche, 1997), and student motivation (Cashin & Downey, 1992) also influence student evaluations. In summary, the current evidence suggests that an array of factors, not just the quality of the course, impact student evaluations of their college professor.

Second, research results indicate that being perceived as physically attractive is associated with a wide range of positive outcomes (Bloch & Richins, 1992; Hosoda, Stone-Romero, & Coats, 2003; Langlois et al., 2000). For example, individuals perceived as attractive are more likely to receive help from strangers than are persons perceived as unattractive (Benson, Karabenick, & Lerner, 1976). In both simulated (Mazzella & Feingold, 1994) and real judicial trials (Stewart, 1980), defendants perceived as attractive were more likely to receive a more lenient punishment if found guilty of a crime. Researchers have also demonstrated that persons perceived as attractive (a) were viewed as more socially competent (Eagly, Ashmore, Makhijani, & Longo, 1991), (b) were viewed as having greater academic potential by teachers (Ritts, Patterson, & Tubbs, 1992), (c) were more persuasive communicators (Chaiken, 1979), and (d) were preferred by voters in political elections (Budeshheim & DePaola, 1994; Sigelman, Thomas, Sigelman, & Ribich, 1986). Researchers have shown that individuals perceived as attractive receive higher incomes than co-workers perceived as unattractive (Frieze, Olson, & Russell, 1991; Hamermesh & Biddle, 1994). Hosoda et al.'s (2003) meta-analysis

demonstrated that individuals perceived as attractive obtain better outcomes for a variety of job-related issues (e.g., hiring, promotion, performance evaluation). Although some factors, such as concern for others and integrity, have not demonstrated an influence of perceived attractiveness (Eagly et al.), the overall literature has indicated a wide variety of positive outcomes. In fact, Myers (2005) summarized this literature as, "Good looks are a great asset" (p. 432).

It is important to note that an individual's physical attractiveness is not an objective variable like heart rate or weight that can be measured with precise accuracy. Although there is general agreement about who is attractive both within and between cultures (Langlois et al., 2000), evaluating physical attractiveness is partially a subjective judgment (Eagly et al., 1991; Monin, 2003). Thus, individual raters can perceive and evaluate physical attractiveness somewhat differently. Furthermore, "there seems to be no agreed-upon criteria for defining physical attractiveness in attractiveness research" (Hosoda et al., 2003, p. 457). For the present study, the classification of "attractive" and "nonattractive" groups (i.e., professors) based on the perceptions of the majority of raters (i.e., students).

Moreover, rating attractiveness is not solely influenced by the physical appearance of the target (e.g., the professor) and individual preferences of the perceiver (e.g., the student), but additional influences can contribute. For example, the target's personality characteristics (Gross & Crofton, 1977), similarity of attitudes between perceiver and target (Klentz, Beaman, Mapelli, & Ullrich, 1987), the perceived familiarity of the target (Monin, 2003), the perceiver's sense of self (Horton, 2003), and the dating status and commitment to a partner in close relationships of the perceiver (D. J. Johnson & Rusbult, 1989; Simpson, Gangestad, & Lerma, 1990) all influence perceptions of attractiveness. Furthermore, in the majority of the physical attractiveness literature, researchers have relied upon first impressions. However, differences may exist between initial impressions compared with those following repeated exposures when the perceiver has additional information about the target (Eagly et al., 1991; Hosoda et al., 2003).

Research also indicates that other factors, such as the gender of the perceiver and the clothing of the target, also influence the perception of physical attractiveness (Abbey, Cozzarelli, McLaughlin, & Harnish, 1987; Buckley, 1983; Workman & Orr, 1996). For example, Williamson and Hewitt (1986) found that males perceived female models as more attractive in sexually alluring clothing, whereas women rated the female models as more attractive in neutral attire. Likewise, in studies investigating sexual harassment (K. P. Johnson & Workman, 1992) and acquaintance rape (Cassidy & Hurrell, 1995; Workman & Freeburg, 1999; Workman & Orr), researchers have also shown that clothing (e.g., skirt length) and gender of the perceiver can influence the perceptions of a target (i.e., the victim). Also, marketing researchers have demonstrated that adornments (e.g., makeup, hairstyle, jewelry) can also alter perceptions of physical attractiveness (Bloch & Richins, 1992; Mack & Rainey, 1990). In summary, the evaluation of who is perceived as

physically attractive is not simply an objective variable, but is partially a subjective judgment that can be influenced by multiple inputs.

Currently, we are aware of only four studies (Ambady & Rosenthal, 1993; Buck & Tiene, 1989; Goebel & Cashen, 1979; Hammermesh & Parker, in press) in which researchers have attempted to investigate the influence of perceived physical attractiveness on student evaluations of college professors. In an initial study, Goebel and Cashen used 10 college freshmen to classify black-and-white photographs as attractive or unattractive. Twenty different freshmen subsequently judged presumed teaching effectiveness from the photographs. Results showed the photos judged as attractive received higher ratings. Buck and Tiene modified Goebel and Cashen's study with 42 undergraduate seniors by attaching a written statement about teaching philosophy (authoritarian or humanistic) to photos judged as attractive or unattractive. Results indicated no main effects of attractiveness on perceived competence, but interaction effects indicated that attractive female authoritarian photos received higher ratings compared with those of male (both attractive and unattractive) and unattractive female authoritarian photos. However, results of both studies have limited generalizability because the authors relied upon presumed (as opposed to real) student evaluations.

In the first study investigating the influence of perceived physical attractiveness using actual student evaluations, Ambady and Rosenthal (1993) primarily focused on the influence of nonverbal behavior. The authors asked two female undergraduates to rate attractiveness (5-point scale) from a still video clip of 13 graduate teaching fellows (6 women) who were teaching sections for undergraduate courses. The authors subsequently correlated perceived attractiveness ratings with real end-of-semester evaluations (comprised of the mean ratings from the students in the section). Perceived attractiveness was not statistically related to student evaluations ( $r = .32, ns$ ), perhaps because of the low statistical power associated with the small sample size ( $n = 13$ ). The results of that study are not only limited because of the small sample size of the teaching fellows, but have limited generalizability because perceived attractiveness was judged by only two female raters.

The most comprehensive investigation of this topic was recently performed by Hamermesh and Parker (in press). Six undergraduate students (3 women) rated the perceived attractiveness of 94 professors by using photographs posted on departmental Web sites. Physical attractiveness ratings (10-point Likert scale) were compared with the professors' real end-of-semester evaluations (number of students who completed student evaluations ranged from 5 to 380). Regression analysis indicated a strong influence of perceived attractiveness on student evaluations. Professors rated as attractive were more likely to receive higher evaluations. Subsequent analysis indicated that the influence of perceived attractiveness was stronger for male as compared with female professors. However, the results of that study were limited by the small number of students who rated perceived attractiveness.

In summary, the aforementioned literature on perceived physical attractiveness and student evaluations of college professors not only is restricted to a handful of studies, but is limited with respect to ecological (i.e., real-world) validity in several important ways. First, previous researchers did not use both the rankings of professors' perceived attractiveness and evaluations by students who were enrolled in the course. In addition, researchers who used real end-of-semester evaluations relied on very small numbers of students to rate attractiveness. Relying on ratings from a handful of undergraduates who were not enrolled in the course is potentially problematic and may have limited generalizability because evaluating attractiveness is partially a subjective judgment with multiple inputs (Eagly et al., 1991; Monin, 2003). Furthermore, previous researchers have relied upon perceptions of still images, which may differ from perceptions obtained by face-to-face interaction (Eagly et al.). As Buck and Tiene (1989) have noted, relying on a still image measures an initial impression. However, student evaluations are typically given at the end of the college semester after students have been repeatedly exposed to the professor. Thus, the perceiver may have different inputs for rating attractiveness between initial impressions and repeated exposures over time (Eagly et al.), a limitation that extends to most research on attractiveness (Hosoda et al., 2003).

Our purpose in the present study was to add to the limited literature by examining perceived physical attractiveness and student evaluations in a naturally occurring database of concurrent ratings. The universities shown in Table 1 have large numbers of student evaluations (as of June 1, 2004, ranging from 20,131 to 36,312) on the Internet Web site [www.ratemyprofessors.com](http://www.ratemyprofessors.com). A public Web site designed for students, [www.ratemyprofessors.com](http://www.ratemyprofessors.com) posts anonymous and voluntary evaluations of college professors (the Web site is not university sponsored). Although our study has all the limitations of any research using naturalistic data (e.g., a potentially biased sample, lack of experimental control, the potential of multiple ratings), it adds to the literature a study with ecological validity. In this study, we compared professors rated as attractive with a nonattractive control group matched for department and gender. Furthermore, we performed multiple replications to determine if the results were statistically reliable (Riniolo & Schmidt, 2000). On the basis of the literature demonstrating that attractiveness is associated with many positive outcomes, we predicted that professors perceived as physically attractive would receive higher evaluations compared with colleagues perceived as nonattractive.

## Method

### *Participants*

In this study, we used student evaluations of professors from the Web site [www.ratemyprofessors.com](http://www.ratemyprofessors.com). We obtained evaluations on June 1, 2004, using the four schools with the most ratings (see Table 1). We selected the most rated

**TABLE 1. Descriptive Statistics From www.ratemyprofessors.com**

Category	Grand Valley State University <sup>a</sup>				University of Delaware <sup>b</sup>			
	<i>n</i>	%	<i>M</i>	<i>SD</i>	<i>n</i>	%	<i>M</i>	<i>SD</i>
All professors	1,714				2,000			
Total ratings	6,312				27,756			
All professors <sup>c</sup>	522				331			
Total ratings	25,590				15,599			
Student evaluations			3.56	0.82			3.50	0.86
Attractive <sup>e</sup>	80				45			
% attractive professors		15				14		
Student evaluations			4.22	0.46			4.11	0.74
No. of ratings			49.6	20.8			39.2	14.7
Hotness total/ no. of ratings			.25	0.17			.25	0.16
Nonattractive <sup>c</sup>	442				286			
Student evaluations			3.44	0.82			3.41	0.84
No of ratings			48.9	23.8			48.4	27.8

<sup>a</sup>Allendale, MI. <sup>b</sup>Newark, DE. <sup>c</sup>San Diego, CA. <sup>d</sup>Harrisonburg, VA. <sup>e</sup>≥ 25 ratings.

schools because (a) large numbers produced more precise estimates and have greater statistical power than smaller samples do (Cohen, 1988), (b) the large number of evaluations indicates that the Web site is well known and actively being used by students, and (c) replication is the best method for determining whether results are statistically reliable (Riniolo & Schmidt, 2000). Also, because just a single student rating will include a professor in the database, we limited data for subsequent statistical analysis to professors that had received at least 25 student evaluations. Researchers have demonstrated high reliability between class-average responses with at least 25 ratings (Marsh & Roche, 1997).

Descriptive information about the full pool of professors is provided in Table 1. Subsequent statistical comparisons between attractive and nonattractive groups (see description of matched analysis in a later section) included 156 professors (50 women, 32%) from Grand Valley State University, 90 professors (48 women, 53%) from the University of Delaware, 106 professors (32 women, 30%) from San Diego State University, and 48 professors (14 women, 29%) from James Madison University. Table 2 shows the number of departments represented.

San Diego State University <sup>c</sup>				James Madison University <sup>d</sup>			
<i>n</i>	%	<i>M</i>	<i>SD</i>	<i>n</i>	%	<i>M</i>	<i>SD</i>
2,205				1,214			
26,921				20,131			
285				275			
12,175				11,621			
		3.48	0.91			3.5	0.89
56				30			
	20				11		
		4.14	0.56			4.39	0.42
		47.5	28.1			38.0	13.9
		.34	0.23			0.27	0.15
229				245			
		3.32	0.90			3.41	0.88
		41.5	16.9			42.6	16.3

Given the naturalistic context of this study, limitations are evident, especially with regard to the ratings from which we took our data. In particular, the ratings made at the Web site are anonymous. Table 3 shows relevant information about undergraduate student populations of the four universities used in this study (i.e., the assumed populations). All have similar numbers of male and female students (range = 58%–60% female), but differences exist, such as percentage of minority and out-of-state students and scores on standardized tests (i.e., ACT and SAT).

### Measures

On the Web site, raters calculate a professor's overall quality on a 5-point Likert-type scale (5 indicating the highest rating) by averaging how the instructor scores on helpfulness and clarity. Definitions of helpfulness and clarity are provided if raters "click" to receive further information. Helpfulness is defined on the Web site as: "This category rates the professor's helpfulness and approachability. Is the professor approachable and nice? Is the professor rude, arrogant, or just plain mean? Is the professor willing to help you after class?" Clarity is defined as: "How

TABLE 2. Departments Represented on www.ratemyprofessors.com

Department	Grand Valley State University		University of Delaware		San Diego State University		James Madison University	
	Men <sup>a</sup>	Women <sup>b</sup>	Men <sup>a</sup>	Women <sup>b</sup>	Men <sup>a</sup>	Women <sup>b</sup>	Men <sup>a</sup>	Women <sup>b</sup>
Accounting	1	1	1	0	1	0	N/A	N/A
Anthropology	1	0			0	1	1 <sup>a</sup>	0 <sup>b</sup>
Biology					2	1	1	2
Business	2	2	1	1			1	0
Chemistry	1	0					1	0
Communication	1	0			2	1	1	1
Computer Science	2	0						
Criminal Justice	3	1	0	1				
Economics	1	0	3	2	5	0	1	0
Education	1	2	1	1	0	1		
English	6	3	5	5	7	4	2	0

Fine Arts	2	2	0	2	0	1
Geography	0	1	4	0	1	0
Geology	6	0	4	0	1	0
History	1	0	4	1	1	0
Hospitality	2	3	1	8	2	1
Languages	3	2	1	1	4	1
Math	3	2	1	1	1	1
Music	2	0	0	0	4	0
Philosophy	3	3	0	1	4	0
Political Science	3	0	2	2	3	0
Psychology	6	3	1	0	2	2
Science	5	0	1	0	1	0
Social Science	0	1	0	1	0	0
Sociology	1	1	0	1	0	1
Theology					1	0
Writing					1	1

<sup>a</sup>Number of attractive/nonattractive male matches (each match represents two professors). <sup>b</sup>Number of attractive/nonattractive female matches (each match represents two professors).

TABLE 3. Undergraduate Information of Student Populations

Characteristic	Grand Valley State University <sup>a</sup>	University of Delaware <sup>b</sup>	San Diego State University <sup>c</sup>	James Madison University <sup>d</sup>
Students ( <i>n</i> )	17,807	17,200	27,345	14,991
Full-time (%)	84	86	79	96
Women (%)	60	58	58	60
Applicants admitted (%)	73	42	50	62
Out of state (%)	4	58	7	30
Minority (%)	10.6	12.3	40.7	10.2
Largest minority	African American	African American	Hispanic American	Asian American or Pacific Islander
Live on campus (%)	29	50	48	40
Full-time freshman retention for 2002 (%)	78	90	82	92
ACT scores >24 (%)	46	71	39	NA
SAT verbal >600 (%)	NA	42	18	35
SAT Math >600 (%)	NA	54	26	39
Full-time faculty (%)	68	81	60	72
Student/faculty ratio	17:1	13:1	19:1	17:1

Source. *Four-year colleges*, 35th ed, by Peterson's, Princeton, NJ, 2004.

<sup>a</sup>Allendale, MI. <sup>b</sup>Newark, DE. <sup>c</sup>San Diego, CA. <sup>d</sup>Harrisonburg, VA.

well does the professor convey the class topics? Is the professor clear in his presentation? Is the professor organized and does the professor use class time effectively?" Perceived attractiveness (note that photos of professors do not appear on the Web site) is calculated from an optional appearance question. The Web site notes that this question is "just for fun," asking students whether their professor is "hot" or "not." The Web site calculates a hot, or not hot, rating (defined as a *hotness total*). Those with an equal or negative balance are assigned a zero rating (i.e., nonattractive), whereas positive balances are displayed (i.e., attractive). For group comparisons, those professors with a positive hotness total were classified as attractive (i.e., a majority who answered this optional question on the Web site perceived the professor as physically attractive). This attractiveness marker resulted in about 15% of the professors (with at least 25 ratings) being classified as attractive (see Table 1 for percentages at each university). Unfortunately, the Web site does not provide how many total hot or not ratings were cast. Table 1 shows the average hotness total divided by the total number of overall ratings for the professors perceived as attractive.

### *Procedure*

We sorted data by number of ratings and included professors with at least 25 ratings in the sample. We divided the sample into attractive and nonattractive groups on the basis of the student ratings. We subsequently matched the sample for gender and department on the basis of attractive and nonattractive controls. In instances in which more than one potential control existed (i.e., a nonattractive professor of the same gender and department), we randomly selected the control from all potential matches. We only included instances of matched professors in the sample used for subsequent matched data analyses. To control for inflation of error rates, we limited the a priori planned analysis to the matched analysis just described.

## **Results**

Descriptive statistics for all attractive and nonattractive professors, prior to matching, are provided in Table 1. After matching, independent *t* tests revealed statistically significant differences between groups because attractive professors had consistently higher evaluations compared with nonattractive controls (see Table 4). We subsequently converted the *t* values into Cohen's *d*, a measure of effect size in which 0.2 indicates a small difference between groups, 0.5 indicates a medium difference, and 0.8 indicates a large difference (Cohen, 1988). Results indicated a large effect size difference between groups (see Table 4).

We performed several exploratory analyses. First, we conducted separate analyses for male and female professors using the same participants in Table 4. Results showed that both attractive men and attractive women scored higher evaluations

**TABLE 4. Student Evaluations After Controlling for Department and Gender**

School	Attractive		Nonattractive control		<i>df</i>	<i>t</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Grand Valley State	4.22	.46	3.39	.81	154	7.83**	1.25
University of Delaware	4.11	.74	3.44	.86	88	3.94**	0.83
San Diego State	4.13	.57	3.32	.82	104	5.93**	1.15
James Madison	4.41	.43	3.36	.86	46	5.35**	1.54

\*\* $p < .001$ .

when compared directly against same-gender nonattractive controls (see Table 5). Second, we identified no statistical differences when comparing attractive men with attractive women at the same university (as shown in Table 2, departments varied between genders). Third, to investigate the relation between number of ratings and student evaluations, we computed correlations (Pearson's  $\rho$ ) using all professors with at least 25 ratings. We performed these analyses to investigate whether professors with low or high student evaluations were more likely to motivate a greater number of ratings. As shown in Figure 1, we found statistically significant results in one of the four schools. However, as shown by  $R^2$ , the variance was small (range = 0%–1.9%) even in the isolated instance of a statistical difference.

### Discussion

Our purpose in this study was to investigate perceived physical attractiveness and student evaluations of college professors by using data obtained from the Web site [www.ratemyprofessors.com](http://www.ratemyprofessors.com) (i.e., a naturally occurring database of concurrent ratings). Results indicated that professors perceived as attractive received higher student evaluations than did nonattractive controls that were matched for both department and gender. In real numbers, professors perceived as attractive scored about 0.8 of a point higher on a 5-point scale (see Table 4). We interpret this difference as a practically meaningful result because professors perceived as attractive move from slightly higher than average on the 5-point scale (i.e., an okay professor) to above-average ratings (i.e., a good professor). With institutionally sponsored student evaluations, moving into the above-average category is often the difference on such important decisions for professors as promotion, tenure, and salary increases (Millea & Grimes, 2002; Williams & Ceci, 1997). Furthermore, results from this study

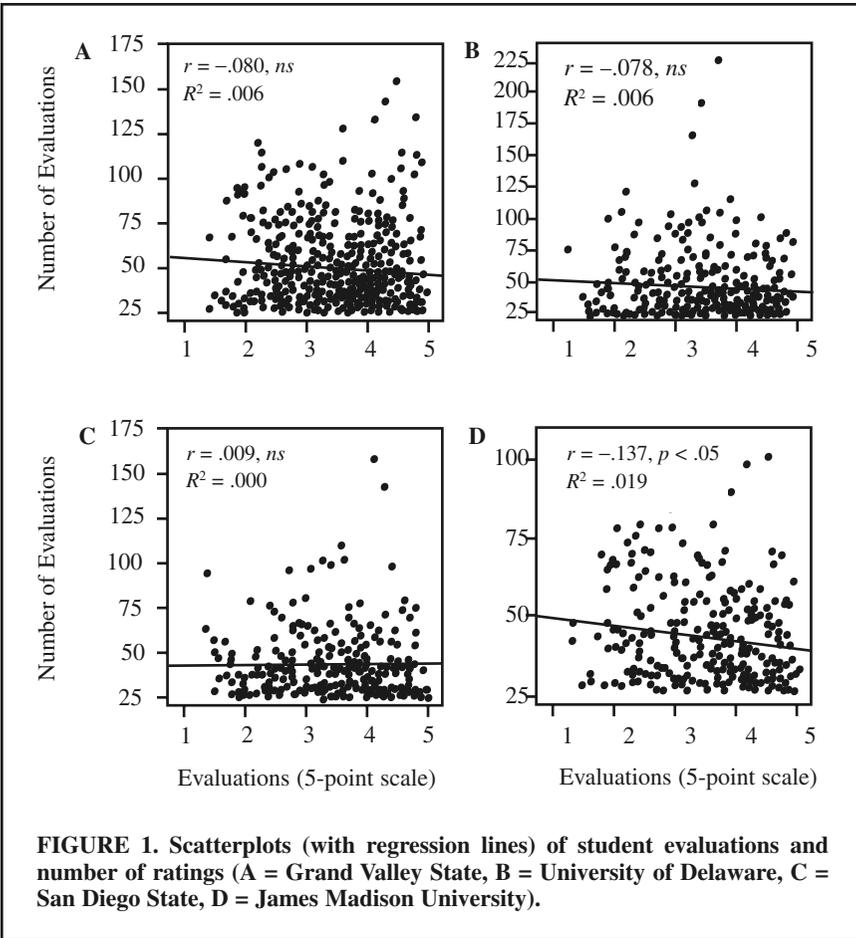
**TABLE 5. Student Evaluations Controlled for Department and Separated by Gender**

School	Attractive		Nonattractive control		<i>df</i>	<i>t</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
<i>Male professor</i>							
Grand Valley State	4.18	0.48	3.44	0.82	104	5.73**	1.11
University of Delaware	4.08	0.86	3.54	0.72	40	2.20*	0.68
San Diego State	4.14	0.63	3.20	0.88	72	5.27**	1.23
James Madison	4.33	0.44	3.35	0.95	32	3.85**	1.32
<i>Female professor</i>							
Grand Valley State	4.28	0.44	3.29	0.78	48	5.51**	1.56
University of Delaware	4.14	0.63	3.36	0.98	46	3.29*	0.95
San Diego State	4.11	0.40	3.59	0.59	30	2.90*	1.03
James Madison	4.61	0.36	3.39	0.65	12	4.36**	2.33

\* $p < .05$ . \*\* $p < .001$ .

were not an isolated finding, but were consistent across four separate universities. Perhaps the most interesting aspect of the results is the range of student evaluations. Ratings for professors perceived as nonattractive ranged from very low to extremely high (i.e., the full spectrum of student evaluations). However, ratings for professors perceived as attractive rarely dropped below an average score (only 6 out of 211 scored below an average rating of 3 on a 5-point scale).

Also, our overall results are consistent with a recent experimental investigation performed by Hamermesh and Parker (in press) in which the authors found a strong influence of perceived attractiveness on real end-of-semester evaluations. Klahr and Simon (2001) have advocated complementary approaches (both experimental and naturalistic) in the process of scientific discovery to provide convergent evidence. However, in contrast to the Hamermesh and Parker study in which the authors found a larger impact of perceived attractiveness for male professors, we found no evidence of gender differences because both male and female professors perceived as attractive received relatively equivalent ratings. Further research into the potential impact of gender differences and perceived attractiveness on student evaluations is warranted to determine the discrepancy between results.



Of course, there are many potential limitations that could affect the overall validity of this study because we obtained data from a naturally occurring database. The most significant limitations are the lack of knowledge of the participants providing the ratings of professors and the potential for multiple ratings. There is no way to verify who provided the ratings because anyone could potentially contribute to the data. Likewise, the potential for multiple ratings is problematic because a single rater could artificially inflate or deflate a professor's overall rating. Despite the anonymous input, the basic characteristics of the ratings can be described. First, professors (with at least 25 ratings) had an average student evaluation of about 3.5 (see Table 1) indicating that most professors sampled were rated above average. Second, Figure 1 shows that the ratings are widely dispersed and not just clustered at the extremes on the 5-point student evaluation scale, indicating a wide distribution of input that is not solely targeted at

evaluating professors rated as very poor (i.e., motivated to “slam” professors) and outstanding (i.e., motivated to praise professors) or both. Because students have a long history of disseminating and sharing information about professors (Williams & Ceci, 1997), it may be that the ratings, as indicated by the large number of inputs at these institutions (see Table 1), are most often used by students to communicate information.

It is important to note that although this study indicates that professors perceived as physically attractive receive higher student evaluations, our results should not be viewed in any way as intended to establish a causal link. Our results merely (a) add to the sparse current literature, (b) evaluate the potential for a practical data set analysis in contributing to the literature, (c) provide a complementary approach with ecological validity, and (d) lead to further research questions that can be evaluated using more rigorous experimental designs (as opposed to the naturalistic data collection used in this study).

It would be interesting for future research to determine the consistency of the initial perceptions of attractiveness (e.g., first day of classes) with attractiveness ratings at the end of the semester. As previously mentioned, multiple inputs beyond the actual physical appearance of the target and individual preferences of the perceiver contribute to the evaluation of physical attractiveness, which may vary with time (Eagly et al., 1991; Hosoda et al., 2003). Also, different levels of initial attractiveness (nonattractive, somewhat nonattractive, neutral, attractive, very attractive) may be more or less stable across time. Future researchers should attempt to establish the stability of the perception of attractiveness across time in the college classroom. Future researchers should also investigate whether evaluations of professors performed by peers, department chairs, and deans are also influenced by perceived attractiveness. Finally, future researchers should attempt to establish the validity of the student ratings at [www.ratemyprofessors.com](http://www.ratemyprofessors.com) compared with those of expert evaluators and real in-class student evaluations.

## REFERENCES

- Abbey, A., Cozzarelli, C., McLaughlin, K., & Harnish, R. J. (1987). The effects of clothing and dyad sex composition on perceptions of sexual intent: Do women and men evaluate these cues differently? *Journal of Applied Social Psychology, 17*, 108–126.
- 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.
- Basow, S. A. (2000). Best and worst professors: Gender patterns in students' choices. *Sex Roles, 43*, 407–417.
- Benson, P. L., Karabenick, S. A., & Lerner, R. M. (1976). Pretty pleases: The effects of physical attractiveness, race, and sex on receiving help. *Journal of Experimental Social Psychology, 12*, 409–415.
- Bloch, P. H., & Richins, M. L. (1992). You look “mahvelous”: The pursuit of beauty and the marketing concept. *Psychology & Marketing, 9*, 3–15.
- Buck, S., & Tiene, D. (1989). The impact of physical attractiveness, gender and teaching

- philosophy on teacher evaluations. *Journal of Educational Research*, 82, 172–177.
- Buckley, H. M. (1983). Perception of physical attractiveness as manipulated by dress: Subject versus independent judges. *The Journal of Psychology*, 114, 243–248.
- Budesheim, T. L., & DePaola, S. J. (1994). Beauty or the beast? The effects of appearance, personality, and issue information on evaluations of political candidates. *Personality and Social Psychology Bulletin*, 20, 339–348.
- Cashin, W. E., & Downey, R. G. (1992). Using global student rating items for summative evaluation. *Journal of Educational Psychology*, 84, 563–572.
- Cassidy, L., & Hurrell, R. M. (1995). The influence of victim's attire on adolescents' judgments of date rape. *Adolescence*, 30, 319–323.
- Chaiken, S. (1979). Communicator physical attractiveness and persuasion. *Journal of Personality and Social Psychology*, 37, 1387–1397.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- 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.
- Freeman, H. R. (1994). Student evaluations of college instructors: Effects of type of course taught, instructor gender and gender role, and student gender. *Journal of Educational Psychology*, 86, 627–630.
- Frieze, I. H., Olson, J. E., & Russell, J. (1991). Attractiveness and income for men and women in management. *Journal of Applied Social Psychology*, 21, 1039–1057.
- Goebel, B., & Cashen, V. (1979). Age, sex and attractiveness as factors in student ratings of teachers: A developmental study. *Journal of Educational Psychology*, 71, 646–653.
- Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 52, 1209–1217.
- Gross, A. E., & Crofton, C. (1977). What is good is beautiful. *Sociometry*, 40, 85–90.
- Hamermesh, D. S., & Biddle, J. E. (1994). Beauty and the labor market. *American Economic Review*, 84, 1174–1194.
- Hamermesh, D. S., & Parker, A. M. (in press). Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review*.
- Horton, R. S. (2003). Similarity and attractiveness in social perception: Differentiating between biases for the self and the beautiful. *Self & Identity*, 2, 137–152.
- Hosoda, M., Stone-Romero, E. F., & Coats, G. (2003). The effects of physical attractiveness on job-related outcomes: A meta-analysis of experimental studies. *Personnel Psychology*, 56, 431–462.
- Johnson, D. J., & Rusbult, C. E. (1989). Resisting temptation: Devaluation of alternative partners as a means of maintaining commitment in close relationships. *Journal of Personality and Social Psychology*, 57, 967–980.
- Johnson, K. P., & Workman, J. E. (1992). Clothing and attributions concerning sexual harassment. *Home Economics Research Journal*, 21, 160–172.
- Klahr, D., & Simon, H. A. (2001). What have psychologists (and others) discovered about the process of scientific discovery? *Current Directions in Psychological Science*, 10, 75–79.
- Klentz, B., Beaman, A. L., Mapelli, S. D., & Ullrich, J. R. (1987). Perceived physical attractiveness of supporters and nonsupporters of the women's movement: An attitude-similarity-mediated error (AS-ME). *Personality and Social Psychology Bulletin*, 13, 513–523.
- Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, 126, 390–423.
- Mack, D., & Rainey, D. (1990). Female applicants' grooming and personnel selection. *Journal of Social Behavior and Personality*, 5, 399–407.

- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist, 52*, 1187–1197.
- Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology, 92*, 202–228.
- Mazzella, R., & Feingold, A. (1994). The effects of physical attractiveness, race, socio-economic status and gender of defendants and victims on judgments of mock jurors: A meta-analysis. *Journal of Applied Social Psychology, 24*, 1315–1344.
- McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist, 52*, 1218–1225.
- Millea, M., & Grimes, P. W. (2002). Grade expectations and student evaluation of teaching. *College Student Journal, 36*, 582–590.
- Monin, B. (2003). The warm glow heuristic: When liking leads to familiarity. *Journal of Personality and Social Psychology, 85*, 1035–1048.
- Myers, D. G. (2005). *Social psychology* (8th ed.). New York: McGraw-Hill.
- Naftulin, D., Ware, J., & Donnelly, F. (1973). The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education, 48*, 630–635.
- Peterson's. (2004). *Four-year colleges* (35th ed.). Princeton, NJ: Author.
- Radmacher, S. A., & Martin, D. J. (2001). Identifying significant predictors of student evaluations of faculty through hierarchical regression analysis. *The Journal of Psychology, 135*, 259–268.
- Riniolo, T. C., & Schmidt, L. A. (2000). Searching for reliable relationships with statistics packages: An empirical example of the potential problems. *The Journal of Psychology, 134*, 143–151.
- Ritts, V., Patterson, M. L., & Tubbs, M. E. (1992). Expectations, impressions, and judgments of physically attractive students: A review. *Review of Educational Research, 62*, 413–426.
- Sigelman, C. K., Thomas, D. B., Sigelman, L., & Ribich, F. D. (1986). Gender, physical attractiveness, and electability: An experimental investigation of voter biases. *Journal of Applied Social Psychology, 16*, 229–248.
- Simpson, J. A., Gangestad, S. W., & Lerma, M. (1990). Perception of physical attractiveness: Mechanisms involved in the maintenance of romantic relationships. *Journal of Personality and Social Psychology, 59*, 1192–1201.
- Stewart, J. E. (1980). Defendant's attractiveness as a factor in the outcome of criminal trials: An observational study. *Journal of Applied Social Psychology, 10*, 348–361.
- Ware, J., & Williams, R. (1975). The Dr. Fox effect: A study of lecturer effectiveness and ratings of instruction. *Journal of Medical Education, 40*, 149–156.
- Waters, M., Kemp, E., & Pucci, A. (1988). High and low faculty evaluations: Descriptions by students. *Teaching of Psychology, 15*, 203–204.
- Williams, W. M., & Ceci, S. J. (1997). "How'm I doing?" *Change, 29*, 13–24.
- Williamson, S., & Hewitt, J. (1986). Attire, sexual allure, and attractiveness. *Perceptual & Motor Skills, 63*, 981–982.
- Workman, J. E., & Freeburg, E. W. (1999). An examination of date rape, victim dress, and perceiver variables within the context of attribution theory. *Sex Roles, 41*, 261–277.
- Workman, J. E., & Orr, R. L. (1996). Clothing, sex of subject, and rape myth acceptance as factors affecting attributions about an incident of acquaintance rape. *Clothing & Textiles Research Journal, 14*, 276–284.

*Manuscript received August 8, 2004*

*Revisions accepted for publication March 8, 2005*

Copyright of *Journal of General Psychology* is the property of Heldref Publications and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.