10-2012

My Professor is Hot! Correlates of RateMyProfessors.com Ratings for Criminal Justice and Criminology Faculty Members

Richard R. Johnson

Angela D. Crews
Marshall University, crewsa@marshall.edu

My Professor is Hot! Correlates of RateMyProfessors.com ratings for criminal justice and criminology faculty members.

Abstract

"RateMyProfessors.com" ratings of the easiness, helpfulness, clarity, overall quality, and "hotness" of 407 criminal justice and criminology faculty members from across the United States were collected. Data were analyzed to determine what faculty characteristics determined these ratings. Experience working in the criminal justice field predicted higher ratings, while years of teaching experience was predictive of lower ratings. After controlling for instructors easiness and "hotness" ratings, the instructors' ascribed characteristics (such as race and sex) explained the greatest proportion of variance in clarity, helpfulness, and overall quality scores. Professional characteristics, such as years of experience, publication rate, and possession of a doctorate were less influential on Ratemyprofessors.com scores.

Key Words

Ratemyprofessors.com, teaching, online, internet, evaluations, quality, faculty, criminal justice

Over the past decade, internet websites where students can rate their instructors at both grade school and college levels have proliferated (Pfeiffer, 2006). The most popular of these websites at the post-secondary level is RateMyProfessors.com (RMP). Those who operate these for-profit websites (which make money by selling advertising banner space on their websites) contend that they serve students by providing them information so that they can make better choices when they select courses (Associated Press, 2003). These operators also suggest that if institutions of higher education were more forthcoming with their own data on student evaluations of teaching, these very popular websites would not be in such demand.

Academicians, however, have voiced many complaints about these websites. The primary complaint is the lack of quality controls. These websites have no safeguards to prevent persons who are not the instructors' former students from posting ratings of instructors (Otto, Stanford, & Ross, 2008; Timmerman, 2008). Documented cases, for example, describe disgruntled faculty members exacting vengeance on colleagues by posing as students and posting disparaging rankings and comments about those colleagues (Carnevale, 2006). Nothing prevents instructors from posting phony, over-inflating ratings of themselves or of their colleagues on these websites (Montell, 2006). Furthermore, only limited safeguards prevent an individual from registering multiple postings about a single instructor (Montell, 2006). Traditional student evaluations of teaching administered by educational institutions, while far from perfect, do not usually suffer from these weaknesses (Otto et al., 2008; Timmerman, 2008). Serious questions exist, therefore, about the validity of the instructor ratings posted on these websites.

This potential lack of validity would not be as important if these online instructor ratings were used purely for entertainment. Unfortunately, however, empirical evidence suggests that many students actually rely upon these ratings when deciding on a course, course section, or

instructor to take (Kindred & Mohammed, 2005). Forbes Magazine uses RMP instructor scores to calculate its annual ranking of the nation's best colleges, with RMP scores comprising twenty-five percent of each school's total score (Ewalt, 2010). Evidence also suggests that college administrators "unofficially" use these ratings to evaluate faculty members (Montell, 2006; Pannapacker, 2007), and that faculty search committees "unofficially" use them to help determine the quality of job candidates (Montell, 2006; Pannapacker, 2007). Given that these unofficial websites have been used "unofficially" to make very official decisions about the rankings of universities and the careers of faculty members, potential bias in these online ratings becomes more frightening.

Using a sample of full-time criminal justice instructors, the present study extended the current literature on these online instructor evaluations in three important ways. First, it determined what instructor professional characteristics are correlated with RMP scores to determine if these characteristics influenced ratings. Second, it determined what instructor ascribed characteristics are correlated with RMP scores to determine if these characteristics influenced ratings. Third, it expanded the literature by specifically investigating correlates of online evaluations for instructors within the academic discipline of criminal justice and criminology in the United States.

Literature Review

RateMyProfessors.com

RMP is the most popular ratings website for students to rate instructors (Associated Press, 2003). Since 1999, when RMP went online, the site has received more than 11 million ratings of more than 1 million instructors from more than 6,000 technical schools, colleges, and universities (RMP, 2010). On the website, respondents can anonymously rate instructors on

easiness, helpfulness, clarity, and "hotness" (i.e., physical attractiveness). The respondents also may make lightly censored comments about the instructor in a provided text block (Felton, Koper, Mitchell, & Stinson, 2008; Kindred & Mohammed, 2005). The arithmetic mean scores of the instructor are publicly displayed under the instructor's name and school affiliation. Individuals making posts on the website clearly are self-selected by the willingness to make anonymous public comments about individual instructors. The motives of those making these posts seem to range from a desire to praise instructors they liked to an open retaliation against instructors they did not like (Felton et al., 2008; Kindred & Mohammed, 2005).

Furthermore, posting evaluations about instructors is not restricted to former students. A jilted lover, former spouse, or vindictive colleague could easily pose as a student and post negative ratings about the instructor. In fact, cases of this type of slander have been documented (Aslet, 2006; Montell, 2006; Pannapacker, 2007). Just as easily, the instructor, or a kind friend could pose as a student and post ratings and comments that exaggerate the quality of the instructor. Evidence suggests that this practice is common as well (Montell, 2006). Additionally, only limited safeguards prevent an individual from registering multiple postings about a single instructor to artificially increase or decrease the instructor's scores (Otto et al., 2008; Timmerman, 2008).

The validity of these online instructor evaluations, therefore, is seriously in question. Recent empirical research from academic disciplines outside criminal justice and criminology have found that RMP ratings of instructor helpfulness, clarity, and overall quality were primarily a function of their easiness and "hotness" ratings, not student learning or actual instructor performance (Coladarci & Kornfield, 2007; Felton, Mitchell, & Stinson, 2004; Otto et al., 2008). Analyzing the RMP ratings of 6,852 instructors from 369 colleges and universities in the United

States and Canada, Felton and associates (2008) found that the easiness and "hotness" scores of the instructors strongly predicted the helpfulness, clarity, and overall quality scores of the instructors. Using a random sample of 399 instructors from 373 institutions in the United States, Otto et al. (2008) also found that RMP helpfulness and clarity scores were highly correlated with each other, suggesting their validity in measuring instructor quality. Additionally, they reported that instructor "hotness" was positively correlated with overall quality, but easiness was not. Unlike the previous studies, these researchers found an inverse relationship between instructor easiness and overall quality.

Comparisons to Traditional Student Evaluations

In an effort to determine whether internet ratings were similar to traditional student evaluations, Coladarci and Kornfield (2007) compared the RMP ratings of 426 University of Maine faculty members with their official student evaluations of teaching scores calculated by the university. Results indicated that RMP scores for easiness, helpfulness, clarity, and overall quality were highly correlated with the scores for similar items on their institution's official student evaluation of teaching surveys. Nevertheless, instructor easiness and "hotness" held the greatest predictive strength for instruction quality scores in both the RMP ratings and the official university evaluation scores.

This strong predictive relationship between the instructor's low level of academic rigor (as measured by "easiness"), the instructor's physical attractiveness (as measured by "hotness"), and the quality of instruction ratings recorded on RMP has been replicated in discipline specific studies. Lawson & Stephenson (2005) examined RMP scores of 295 economics faculty members and determined that, even after controlling for the sex of the faculty member, the instructor's ease and "hotness" were strong predictors of the instructor's overall quality rating. Using a

sample of business professors from five universities, Timmerman (2008) found that easiness and attractiveness best predicted overall quality scores, but that the RMP scores were also highly correlated with similar measurements on official university student evaluations of teaching.

Weaknesses of Traditional Student Evaluations

Literature on research related to traditional evaluations is extensive (Wilson, 1998), and a full review is beyond the scope of this paper. These types of evaluations, however, have been criticized because they may base assessments of instructors on things other than actual instructional quality (i.e., grading leniency, course difficulty, instructor physical and personality traits). RMP scores' reliability with traditional student evaluations of teaching scores, then, may not necessarily attest to their validity. A large body of literature has highlighted the weaknesses and inherent biases that affect these traditional evaluations (for reviews see Feldman, 1983, 1987, 1993; Marsh & Roche, 1997; Wilson, 1998).

Traditional student evaluations of teaching can be biased against instructors who impose high academic standards, persons of color, female instructors, older instructors, and physically unattractive instructors. Studies of traditional student evaluations of teaching consistently suggest a strong, negative correlation between course difficulty and student evaluation scores (Chacko, 1983; Hoffman, 1983; Meredith, 1982; Stratton, Myers, & King, 1994), however one study has suggested this relationship might be curvilinear (Centra, 2003), with instructors who are *extremely* lax also being rated poorly.

The literature on the relationship between instructor race and student evaluations of teaching has produced mixed results. While some studies suggested that instructor race is not correlated with student evaluations of teaching (Feldman, 1993), others indicated that nonwhite instructors were rated lower than were whites (Feldman, 1993; Shapiro, 1990; Stack, 2000). Sex

bias has been far more consistent, with female instructors generally rated lower than males on helpfulness, availability to students, and overall teaching effectiveness (Feldman, 1993; Lueck, 1993; Miller & Chamberlin, 2000). Instructor's years of teaching experience (also a proxy measure of instructor age) is inversely related to student evaluations of teaching scores (Feldman, 1983), and physically attractive instructors are generally rated higher (Felton, Mitchell, & Stinson, 2004; Pike, 1999).

Professional characteristics of faculty members, such as level of scholarly productivity, possession of a doctorate, and professional field experience may also be influential. Scholarly productivity has had an impact on instructor evaluation scores when productivity is measured in a specific manner. Reviews and meta-analyses of studies measuring the link between faculty research productivity and student evaluations of teaching have revealed a consistent, positive correlation between the instructor's publication rate for peer-reviewed research articles and student evaluation scores (Feldman, 1987; Gomez-Mejia & Balkin, 1992; Marsh, 1987). When publication rate measures included books, monographs, and non-peer reviewed articles, no correlation between publication productivity and student teaching evaluations was found (Gomez-Mejia & Balkin, 1992; Marsh, 1987).

Stack (2000) suggested that instructors possessing a doctorate should have higher student evaluation scores than instructors without a terminal degree because their higher degrees provide them with more extensive knowledge about their field, as well as formal training and tutored experience in methods of teaching. The empirical evidence on this hypothesis, however, is equivocal (Finegan, 1998; Sonner, 2000). Finally, a small body of literature suggests that instructors with professional field experience in the academic discipline they are teaching tend to receive higher student evaluation scores than instructors lacking such professional field

experience (Schrink, Roy, & Ransburg, 1999; Sonner, 2000; Zahn & Schramm, 1992). This may be especially true in academic disciplines of a more applied nature, such as criminal justice.

The present research thus aimed to study faculty members in criminal justice and criminology programs and answer three specific research questions. 1) What faculty characteristics correlate with RMP ratings? 2) What influence do faculty characteristics have on RMP scores after controlling for the easiness (lack of rigor) and "hotness" (physical attractiveness) ratings of the instructor? 3) Does the number of raters influence RMP scores, given that a small sample of self-selected individuals (disgruntled students, exceptionally pleased students, malicious and kind colleagues, friends, family, and ex-lovers) post evaluations on RMP?

Method

To explore these three questions, a sample of full-time faculty members who were rated on RMP was created. Data on the professional and ascribed characteristics of the instructors were gathered. These characteristics were then examined for empirical relationships to the five specific RMP scores. Further analysis then controlled for level of academic rigor and physical attractiveness by treating the RMP easiness and "hotness" scores as independent variables and adding them to the multivariate models estimating the remaining three RMP scores.

Sample

Scores from the RMP website and information collected from institutional websites were used to create a sample of full-time criminal justice and criminology faculty members. Data collection began by randomly selecting 10 states from which to draw cases (Arizona, Connecticut, Illinois, Michigan, Missouri, New Jersey, New Mexico, North Carolina, Oklahoma, and Wisconsin). All four-year colleges and universities within each of these states were searched

in RMP for all rated, full-time criminal justice & criminology faculty members. (Part-time faculty members and community college institutions were excluded because of the relative unavailability of the information needed to construct the independent variables). When a rated faculty member was located, a cross check was made with the website of the respective university to confirm that the faculty member was employed full-time in a criminal justice or criminology program within the institution. If this was indeed true, the faculty member was added to the sample as a case. Initially, 466 cases comprised the sample.

Data were gathered on the faculty member's ascribed and professional characteristics. These data were most easily obtained from faculty photographs, biographical notes, and curriculum vitae displayed on their department websites. A few institutions, however, provided only the briefest information on the faculty member, such as only a name and email address. In such cases, online searches were conducted to locate a photo and/or more biographical data on the faculty member to obtain data on key independent variables. In 59 cases, information about the faculty member on one or more independent variables could not be located. These incomplete cases were eliminated from the sample, resulting in a final sample size of 407, which represents about 77% of all the full-time criminal justice & criminology faculty members in criminal justice & criminology departments within the ten selected states (N=529).

Measures

Each instructor's RMP scores in the five categories (easiness, helpfulness, clarity, "hotness", and overall quality) functioned as the primary dependent variables. For easiness, helpfulness, and clarity, individual raters score the professor on a scale from 1 to 5, with 1 being the lowest score (i.e., "hardest," "least helpful," "least clear,") and 5 being the highest (i.e., "easiest," "most helpful," "most clear"). To evaluate an instructor's easiness, raters were asked,

"How easy are the classes that this professor teaches? Is it possible to get an A without too much work?" To evaluate helpfulness, raters were asked, "Is the teacher approachable and nice? Is he rude, arrogant, or just plain mean? Is he willing to help you after class?" To evaluate clarity, raters were asked, "How well does the teacher convey the class topics? Is he clear in his presentation? Is he organized and does he use class time effectively?"

"Hotness" was a dichotomous variable and raters answered the question "Is your professor hot?" with either a "hot" or "not." Instructors who are rated "hot" show a red chili pepper icon on their page, alongside their ratings. Each "hot" rating an instructor receives is given +1 by RMP and each "not" rating is given -1, so an instructor who receives an equal number of "hot" and "not" ratings will have a total of "0" and will not merit a chili pepper. Finally, the measure of overall quality is calculated by the RMP website, not the respondents, by averaging the instructor's scores on helpfulness and clarity.

For the four ordinal scale ratings (easiness, helpfulness, clarity and overall quality), the mean score of all the ratings (to two places after the decimal) is displayed on the website, transforming the ratings from ordinal to ratio-level measures. "Hotness" was measured dichotomously, with the presence of a chili pepper scored as 1 (yes, hot) and the absence of a chili pepper scored 0 (no, not hot).

Drawing from the research on traditional forms of student course evaluations of instructors, measures were created for the ascribed and professional characteristics of the instructors. Instructor race was dichotomously coded as white (1) or non-white (0) and was determined by reviewing the photos of the instructors. Attributing the race of the instructor in this manner may have resulted in some measurement error due to misidentification; however, online raters may have equally made the same types of errors in their perceptions of an

instructor's race or ethnicity. Instructor sex was also determined by reviewing the faculty member's photo and first name.

Terminal degree was measured dichotomously as whether or not the instructor had obtained a doctorate (excluding juris doctorates). This information was gathered from faculty bios and vitas on their institutions' websites. Average annual publication rate for the present study was measured as the total number of peer-reviewed journal articles authored or co-authored by the instructor, divided by the number of years the instructor had held his or her highest degree. In most cases, online vitas were available for this information. When vitas were not accessible, information regarding publications was obtained through searches of the National Criminal Justice Reference Service abstracts database, and the year the highest degree was obtained was retrieved from the Dissertation Abstracts database, which holds records for all doctoral dissertations and masters theses in the United States in recent decades.

Professional work experience as a practitioner was defined dichotomously as whether or not the instructor had ever been employed as a law enforcement officer, probation / parole officer, practicing prosecuting or defense attorney, or had ever been employed in any capacity in a correctional institution. If the instructor's online bio or vita mentioned any such practitioner experience, then the instructor was dichotomously measured as having had prior criminal justice system practitioner experience. The instructor's years of full-time teaching experience was calculated as the number of years since the start of the first full-time academic position listed in the instructor's bio or vita. If this information was missing for an instructor, the number of years since the completion of the instructor's highest degree was substituted.

A measure of the number of raters who rated the instructor on the RMP website was also included. The degree of sampling error decreases as the number of respondents in the sample

increases (Rosenthal, 2001), thus reducing the influence of a few atypical evaluations (such as an overly negative rating by a single disgruntled student or vindictive colleague, or an overly positive rating submitted by the instructor or a friendly colleague). It could be anticipated that ratings based on a smaller number of respondents would tend to be significantly more or less positive than ratings based on larger numbers of respondents.

Finally, although literature is lacking on type of institution, expectations of students attending a major research university may differ from students attending smaller universities and liberal arts colleges. The same may also be said for faculty members who have rated themselves or their colleagues on the website. Clearly, the requirements for tenure differ across these various types of institutions (Manger, 1997), so the emphasis on teaching probably also varies across institutional types. A dichotomous measure was created that differentiated among universities classified by the Carnegie-Mellon Foundation as Research I institutions and universities in other classifications.

Procedure

The procedure followed in this study began with a descriptive analysis of the sample's univariate statistics. Next, the bivariate relationships among the variables were explored. Multiple regression analyses were then conducted using each of the five RMP ratings in turn as dependent variables. Finally, multiple regression analyses were again conducted, this time using the RMP easiness and "hotness" ratings as independent variable proxy measures of academic rigor and physical attractiveness.

Results

Descriptive Analysis

Table 1 reveals the descriptive statistics for the sample. The sample of 407 instructors was primarily white (93%), male (72%), and possessed a doctorate (75%). Fewer than half (49%) had practitioner experience, and the sample had nearly 14 years of teaching experience on average. Most were not teaching at Research I universities (64%). The sample had published an average of less than one peer-reviewed article per year, and 21 percent of the sample had not published at all. Most instructors in the sample were rated only by a few respondents. While the number of raters ranged from 1 to 67, 25 percent of the sample had only 1 rater, and 50% had less than 10 raters. Although some prior research considered only professors who had more than 25 ratings (e.g., Riniolo, Johnson, Sherman & Misso, 2006), we were interested in exploring whether the number of raters was empirically associated with any of the outcome variables. Two variables, average annual publications per year and number of raters, were heavily skewed (Shapiro-Wilk's test for normality at $p < .001$), requiring transformations into their natural logs to permit use in linear models (Miles & Shevlin, 2001).

INSERT TABLE 1 ABOUT HERE

The mean rating given to the sample on easiness was 3.29, 3.71 on helpfulness, 3.66 on clarity, and 3.70 on overall quality. The means all differed by no more than 0.32 points, and the standard deviations varied by no more than 0.12 points. A Shapiro-Wilk's test for normality was conducted on each of these four ratings variables, none of which reached statistical significance at the .05 level, suggesting that these four dependent variables were fairly-normally distributed around their respective means (Shapiro & Wilk, 1965). The designation of "hotness" was earned only by 17% of the sample, which was consistent with prior RMP research (Riniolo et al., 2006).

Bivariate Analysis

The first step in the inferential analysis involved the calculation of the bivariate relationships between the variables in the study. Table 2 displays the bivariate correlation matrix and indicates that all of the dependent variables, with the exception of "hotness",' were highly correlated with each other (Pearson's r > .50). Although easiness, helpfulness, clarity, and overall quality were strongly and positively correlated with each other, "hotness" was not correlated with easiness, and only weakly positively correlated with helpfulness, clarity, and overall quality.

INSERT TABLE 2 ABOUT HERE

Easiness was positively correlated with criminal justice practitioner experience, but negatively correlated with possession of a doctorate, logged number of publications, and years of teaching experience. Helpfulness was positively correlated with practitioner experience, but negatively correlated with possession of a doctorate and years of teaching experience. Clarity was positively correlated with being white, male, practitioner experience, and number of raters, but negatively correlated with a doctorate and years of teaching experience. Overall quality was positively correlated with being white and practitioner experience, but negatively correlated with a doctorate and years of teaching experience. Finally, none of the independent variables was correlated with "hotness".

Multiple Regression Analyses

The next step involved regressing each of the independent variables on the dependent variables easiness, helpfulness, clarity, overall quality, and "hotness" in multivariate analyses. Ordinary least squares (OLS) regression was used to estimate the multivariate linear models for

easiness, helpfulness, clarity, and overall quality because they were ratio-level measures that were fairly-normally distributed. Because "hotness" was dichotomous, binary logistic regression was used for this model estimate. Table 3 reveals the results of these tests.

INSERT TABLE 3 ABOUT HERE

Model 1 predicted the instructors' RMP easiness score, and only three independent variables were statistically significant. Practitioner experience was associated with a higher rating for easiness, as was employment by a Research I university. Teaching experience was negatively associated with easiness, with more experience resulting in a lower easiness score. The second model predicted the helpfulness scores and this time four independent variables were statistically significant. Logged publication rate and practitioner experience raised ratings for helpfulness. A doctorate and years of teaching experience reduced scores on helpfulness.

The third model predicted clarity scores, and seven independent variables were significantly significant. White, male, practitioner experience, logged publications, and number of raters was associated with higher clarity scores. A doctorate and teaching experience were associated with lower clarity scores. The fourth model predicted overall quality scores and produced five statistically significant relationships. Practitioner experience, logged publication rate, and being white increased overall quality ratings. Teaching experience and a doctorate resulted in lower ratings of overall quality. Finally, in the logistic regression model predicting "hotness" revealed no statistically significance relationships with the independent variables.

Across all five models, the most consistent predictors of RMP ratings were practitioner experience and years of teaching experience. Practitioner experience was associated with higher

ratings of easiness, helpfulness, clarity, and overall quality. Years of teaching experience, was associated with lower ratings on easiness, helpfulness, clarity, and overall quality. The least consistent predictors in the models were sex, whether the instructor was at a Research I university, and number of raters, each of which were only statistically significant predictors in one model. None of these five models demonstrated great explanatory power either. At best, the models explained 20% of the variance in each dependent variable. At worst, they explained no more than 3% of the variance.

Next, multivariate models were created to predict the helpfulness, clarity, and overall quality scores, but adding the easiness and "hotness" measures as independent variables. Literature pertaining to student evaluations of teaching has consistently suggested a negative correlation between course difficulty and student evaluation scores, and that this is one of the strongest predictors of instructor evaluation scores (Chacko, 1983; Hoffman, 1983; Meredith, 1982; Stratton, Myers, & King, 1994). We reasoned, therefore, that it would be important to add a measure of academic rigor for each instructor as an independent variable. We assumed that each instructor's easiness rating could serve as a proxy measure for the academic rigor the instructor maintained in her courses.

A large body of literature documents the positive impact of attractiveness on student evaluation scores, with attractive instructors receiving better evaluations (Bokek-Cohen & Davidowitz, 2008; Felton et al., 2004; Hamermesh & Parker, 2005; Kindred & Mohammed, 2005; Pike, 1999). We therefore also used each instructor's "hotness" rating as a proxy measure of physical attractiveness.

Table 4 displays the results of three OLS regression models involving the helpfulness, clarity, and overall quality rankings as dependent variables. The model R-squared values

indicated that the inclusion of the easiness and "hotness" measures dramatically increased the explanatory power of the models. All three of the models more than doubled in explanatory power with the inclusion of easiness and "hotness" as independent variables. In all three models displayed in Table 4, the easiness rating was the strongest predictor, having standardized coefficients ranging from .47 to .57, more than twice as large as the next strongest in each model.

INSERT TABLE 4 ABOUT HERE

After controlling for instructor perceived easiness, several other predictors proved to now be statistically significant (albeit much weaker) predictors of helpfulness, clarity, and overall quality ratings. Logged publication rate was a statistically significant and a positive predictor of helpfulness, clarity, and overall quality. "Hotness" was associated with higher ratings in all three models, as was being white. Even after controlling for easiness, instructors with criminal justice system practitioner experience still generally received higher ratings in all three models.

Years of teaching experience no longer influenced the instructor's helpfulness rating; however, more teaching experience still predicted lower scores of clarity and overall quality. Male instructors had higher ratings than females for clarity, but not helpfulness or overall quality. Instructors with a doctorate ranked lower in helpfulness and overall quality than those without a doctorate. More raters now tended to increase ratings on clarity and overall quality. Finally, employment by a research university no longer proved a significant predictor.

The evidence suggests that instructor helpfulness, clarity, and overall quality are influenced by instructor easiness and personal characteristics that, in a prejudice-free world, should not matter. In separate OLS regression analyses not reported here in tabular form, models

including only instructor easiness, “hotness”, race, sex, and experience (potentially a proxy measure for instructor age) produced R-square values of .39 for helpfulness, .36 for clarity, and .43 for overall quality. By comparison, models containing the remaining predictors of practitioner experience, terminal degree, publication rate, a research university, and number of raters only produced R-square measures of .12 for helpfulness, .15 for clarity, and .15 for overall quality.

As with all studies, this one was not without limitations, and these limitations must be remembered when generalizing about these findings. First, there were no controls for the span of time RMP reviewers posted reviews. An instructor could have had a rough start to her teaching career, but over years of teaching, her scores improved. As RMP scores are mean values, this change in score would not have been detected in the present study. Second, the present study did not control for the possible influences of instructor teaching load, class size, or type of course (graduate versus undergraduate, or methods and statistics versus substantive courses). Any of these factors may potentially influence RMP scores. Third, since instructors without any RMP ratings were not included in the sample, we cannot comment on how these instructors may have differed from those who actually had at least one RMP rating.

Discussion

The findings of this study lead to five major conclusions. First, instructor RMP scores on clarity, helpfulness, and overall quality appear to be heavily influenced by the easiness of the instructor. The perception that the instructor gave easy grades was the strongest predictor of helpfulness, clarity, and overall quality in the multiple regression models. The influence of instructor easiness on traditional student evaluation scores (Felton et al., 2004; Heckert, Latier, Ringwald-Burton, & Drazen, 2006; Pike, 1999), and RMP scores (Coladarci & Kornfield, 2007;

Otto et al., 2008) has also been documented in previous studies. Obviously, students tend to attribute many positive attributes to instructors who provide them with higher grades and lower stress. Furthermore, faculty members that are rated as easiest tend to be male, have practitioner experience, and are early in their academic career. Several reasons may explain why instructors with these qualities are perceived as easy.

Workplace gender studies have documented that women often perceive that they have to work harder than men do in male-dominated occupations in order to be accepted by their coworkers (Acker & Feuerverger, 1996). Perhaps female instructors, therefore, perceive that they need to incorporate a higher level of academic rigor in the classroom in order to be accepted as equals by their male colleagues. Another possibility is that gender biases held by the student raters, especially the gender stereotype that women should be more nurturing, cause students (and anyone making a bogus rating) to perceive female instructors as harder when they enforce the same academic standards as male faculty members. A third possibility, considering the evidence that many of the RMP raters are bogus (self-ratings and ratings by family, friends, colleagues, and enemies), perhaps these bogus raters want to help overcome perceived student biases about the lack of professionalism and rigor of female instructors, or the lack of warmth of male instructors.

Possessing prior professional work experience in the criminal justice system also corresponded with higher ratings of easiness. Students may perceive instructors with prior practitioner experience as easier because they have "real life" experiences from which to draw upon to illustrate the material in the text, easing the students' ability to absorb the material. Often these "war stories" are entertaining, making the classroom experience more pleasurable. A second possibility, considering the evidence on bogus ratings, is that colleagues with pure

academic backgrounds make bogus ratings about these instructors, assuming that their use of "war stories" amounts to a low level of academic rigor. Concerns about the reliance on telling "war stories" and the low academic standards of vocational-oriented criminal justice programs (often referred to as "cop shops") have been a topic of debate within the discipline for decades (Morn, 1995). A third possibility is that many instructors who are prior practitioners are in their second career, causing some to take a "retirement job" perspective toward their academic employment, doing a minimal amount of work and taking full advantage of the flexible schedule academia affords.

The finding that easiness was negatively associated with the instructor's teaching experience may be the result of the tenure system. Instructors, especially those in tenure-track positions, may be extra wary of receiving low student evaluation scores. Low scores on student course evaluations, and student complaints, can reduce the likelihood a lecturer's contract will be renewed, and reduce the likelihood of tenure for an assistant professor. It is possible, therefore, that instructors in the first few years of their career will grade lightly in an attempt to appease students, avoid complaints, and obtain high student evaluation scores. It is also likely that the research and service demands on pre-tenure assistant professors serve to discourage pre-tenure instructors from giving out many assignments in order to reduce their own grading workload.

The second major conclusion from the findings of this study is that the instructors' ascribed characteristics matter. Even after controlling for the perceived easiness of the instructor, instructors were still more likely to be rated higher in clarity, helpfulness, and overall quality if they were male, white, and 'hot.' These ascribed characteristics explained almost twice as much variance in RMP ratings than did the professional characteristics that should matter, such as publication rate or possession of a doctorate. After controlling for easiness, male instructors were

rated significantly higher than female instructors were on clarity. There is evidence from the study of traditional student course evaluations that student raters often demonstrate bias against female instructors (Feldman, 1993; Lueck, 1993), perceiving male instructors as more competent and professional. Evidence was found here to suggest that this same bias exists in RMP ratings.

A stronger bias appears to exist for the race of the instructor. The evidence from traditional course evaluations research has demonstrated rater bias against instructors who are persons of color (Feldman, 1993; Shapiro, 1990; Stack, 2000). After controlling for instructor easiness, white instructors received significantly higher ratings than non-white instructors did in helpfulness, clarity, and overall quality. The RMP ratings appear to display across the board bias against instructors who are persons of color that cannot be overcome by the instructor, even if he or she were to provide easy grades.

Physical attractiveness, in the form of a "hotness" rating, also significantly influenced RMP scores. After controlling for level of easiness, instructors who were rated as "hot" generally received higher scores for clarity, helpfulness, and overall quality. Likely, this is due to the well documented attribution errors many people make leading to assumptions that physically attractive people are more competent (Jackson, Hunter & Hodge, 1995), trustworthy (Ohanian, 1990), and intelligent (Langlois et al., 2000). This provides another example of bias within RMP ratings.

The third major conclusion is that professional characteristics have only a small influence on RMP ratings. Just as with traditional student evaluations of instructors (Feldman, 1993; Shapiro, 1990; Stack, 2000), instructor professional credentials (such as possession of a doctorate, scholarly publication rate, and years of teaching experience) produced weaker associations to RMP scores than did the instructors' ascribed characteristics. Whether or not the

instructor possessed a doctorate only weakly predicted helpfulness and overall quality scores, and did so in the opposite direction one would expect. Having a doctorate actually tended to *decrease* scores for helpfulness and overall quality. Perhaps faculty members with doctorates are perceived more intimidating than those without.

Teaching experience was weakly associated with clarity and overall quality, and not associated at all with helpfulness. Perhaps instructors who have been teaching longer, especially those who are tenured, experienced job burnout and put less effort into explaining the course material. Another possibility is that instructors who have dealt with their subject material for long periods become so comfortable with the jargon and concepts of their field, they tend to forget their students are not at their level of knowledge.

Logged publication rate, however, increased scores for helpfulness, clarity, and overall quality. One might expect that a more active research agenda would correspond with less of an emphasis on teaching, but this does not appear to be the case with RMP scores. Even after controlling for instructor easiness, those with higher publication rates were perceived as more clear and helpful. Just as with faculty members who possess criminal justice practitioner experience, instructors who are active researchers in their field may have more "real life" experiences to draw upon in the classroom. While an instructor may never have been a correctional officer, she may have been an active researcher who has interviewed hundreds of prison inmates.

The fourth major conclusion is that the number of raters has very little influence on RMP scores. This was a most interesting finding in light of probability theory and the error associated with smaller samples. Based on probability theory, one would expect that smaller numbers of raters would be heavily influenced by the ratings of the few bogus raters. The fact that the results

did not change measurably as the sample size of raters increased, presents two possibilities. First, this may be because the mean number of raters was generally small, as the case with the largest number of raters only had 67 raters. A second possibility is that the distribution of negative and positive bogus ratings is random, with each instructor having an equal chance of receiving a negative or positive rating from a bogus rater.

As RMP and similar websites were created in part because official student course evaluations are not made public, perhaps there needs to be more transparency in the official, university-sanctioned evaluations. Institutions that use standardized evaluations that have been tested for validity and reliability, such as the Student Instructional Report II (Centra, 1998), could make instructor scores public. We assume that many faculty members and unions would likely oppose such a move, but faculty members are already being publicly rated on RMP and other websites. If universities released official course evaluation information, at least there would be the assurance that those completing the evaluations were actually students enrolled in the course, not a vindictive colleague. There would also be greater control over the evaluation instrument actually used.

Nevertheless, this article added more evidence to the argument that faculty members' teaching quality should be rated on things other than student evaluations. Using course pretests and posttests provide evidence of learning key course material. Peer observation in the classroom is another method that may be used to measure true teaching performance. Reviewing syllabi for course rigor with regard to course readings and assignments is an evaluation method that can limit the biases of instructors' ascribed characteristics. Whatever techniques a committee or administrator may consider, the evidence here suggests that RMP ratings are more a measure of who is easy, white, and male, than who is a good teacher.

References

Acker, S., & Feuerverger, G. (1996). Doing good and feeling bad: the work of women university teachers. *Cambridge Journal of Education, 26*, 401-422.

Aslet, M. (2006, June 28). Breaking up. *The Chronicle of Higher Education*. Retrieved September 9, 2010 from http://chronicle.com/article/Breaking-Up/46834/

Associated Press (2003, February 19). Web warnings: Sites extol, slant professors. *CNN*. Retrieved August 11, 2010, from http://www.cnn.com/2003/EDUCATION/02/l7/rating.professors.ap/

Bokek-Cohen, Y. & Davidowitz, N. (2008). Beauty in the classroom: Are students influenced by professors' appearance? *Jewish Educational Leadership, 6*(3). Available online: http://www.lookstein.org/online_journal.php?id=198.

Carnevale, D. (2006, March 10). University of Saskatchewan fires tenured professor accused of maligning colleagues on ratemyprofessors web site. *The Chronicle of Higher Education*, 52(27), A28.

Centra, J. A. (1998). *Development of the Student Instructional Report II*. Princeton, NJ: Educational Testing Service.

Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education, 44*, 495-518.

Chacko, T. I. (1983). Student ratings of instruction: A function of grading standards. *Educational Research Quarterly, 8*, 19-25.

Coladarci, T., & Kornfield, I. (2007). RateMyProfessors.com versus formal, in-class student evaluations of teaching. *Practical Assessment, Research, and Evaluation, 12(6)*, 1-15.

Ewalt, D. M. (2010, August 11). Special report: America's best colleges. Forbes. Retrieved August 11, 2010 from http://www.forbes.com/2010/08/01/best-colleges-university-ratings-rankings-opinions-best-colleges-10-intro_print.html

Feldman, K. A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education, 21*, 45-116.

Feldman, K. A. (1987). Research productivity and scholarly accomplishments of college teachers as related to their instructional effectiveness: A review and exploration. *Research in Higher Education, 26*, 227-298.

Feldman, K. A. (1993). College students' view of male and female college teachers: Evidence from students' evaluations of their classroom teachers. *Research in Higher Education, 34*, 151-211.

Felton, J. F., Koper, P. T., Mitchell, J. F., & Stinson, M. F. (2008). Attractiveness, easiness, and other issues: Student evaluations of professors on Ratemyprofessors.com. *Assessment and Evaluation in Higher Education, 33*, 45-61.

Felton, J. F., Mitchell, J. F., & Stinson, M. F. (2004). Web-based student evaluations of professors: The relations between perceived quality, easiness, and sexiness. *Assessment and Evaluation in Higher Education, 29*, 91-108.

Finegan, T. A. (1998). Do introductory economics students learn more if their instructor has a Ph.D.? *American Economist, 42*, 34-44.

Gomez-Mejia, L. & Balkin, D. (1992). Determinants of faculty pay: An agency theory perspective. *Academy of Management Journal, 35*, 921-953.

Hamermesh, D.S. & Parker, A.M. (2005). Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review, 24*, 369-376.

Heckert, T.M., Latier, A., Ringwald-Burton, A. & Drazen, C. (2006). Relations among student effort, perceived class difficulty appropriateness, and student evaluations of teaching: Is it possible to "buy" better evaluations through lenient grading? *College Student Journal*, *40*(3), 588-596.

Hoffman, R. A. (1983). Grade inflation and student evaluation of college courses. *Educational and Psychological Research, 3*, 151-160.

Jackson, L.A., Hunter, J.E. & Hodge, C.N. (1995). Physical attractiveness and intellectual competence: A meta-analytic review. *Social Psychology Quarterly*, *58*(2), 108-122.

Kindred, R., & Mohammed S. (2005). He will crush you like an academic ninja: Exploring teacher ratings on ratemyprofessors.com. *Journal of Computer-Mediated Communication, 10*, 9-20.

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*(3), 390-421.

Lawson, R. A., & Stephenson, E. F. (2005). Easiness, attractiveness, and faculty evaluations: Evidence from ratemyprofessors.com. *Atlantic Economic Journal, 33*, 485-486.

Lueck, T. L. (1993). The interaction effects of gender on teaching evaluations. *Journalism Educator, 48*, 46-54.

Manger, D. (1997, September 5). Report says standards used to evaluate research should also be used for teaching and service. *The Chronicle of Higher Education, 43(1)*, A18-A19.

Marsh, H. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11*, 253-388.

Marsh, H. & Roche, L. (1997). Making students' evaluations of teaching effectiveness effective: the critical issues of validity, bias, and utility. *American Psychologist, 52*(11), 1187-1197.

Meredith, G. M. (1982). Grade-related attitude correlates of instruction / course satisfaction among college students. *Psychological Reports, 50*, 1142.

Miles, J., & Shevlin, M. (2001). *Applying regression and correlation*. Thousand Oaks, CA: Sage.

Montell, G. (2006, September 27). The art of the bogus rating. *The Chronicle of Higher Education*. Retrieved September 10, 2010 from http://chronicle.com/article/The-Arts-of-the-Bogus-Rating/46887/

Morn, F. (1995). *Academic politics and the history of criminal justice education*. Westport, CT: Greenwood.

Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising, 19*(3), 39-52.

Otto, J., Stanford, D. A., & Ross, D. N. (2008). Does ratemyprofessor.com really rate my professor? *Assessment and Evaluation in Higher Education, 33*, 355-368.

Pannapacker, W. A. (2007, April 20). The inescapability of your past. *The Chronicle of Higher Education, 53(33),* B11.

Pfeiffer, S. (2006, September 20). Rating sites flourish behind a veil if anonymity. *The Boston Globe*. Retrieved September 9, 2010 from http://www.boston.com/business/technology/articles/2009/09/20/ratings_sites_flouirsh_behind_a_viel_of_anonymity/.

Pike, G. R. (1999). The constant error of the halo in educational outcomes research. *Research in Higher Education, 40*, 61-86.

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? *The Journal of General Psychology*, *133*(1), 19-35.

Rosenthal, J. A. (2001). *Statistics and data interpretation*. Belmont, CA: Wadsworth-Thomson.

Schrink, J. L., Roy, S., & Ransburg, J. (1999). Perceptions of alumni concerning their level of preparation for criminal justice careers: A pilot study. *Journal of Criminal Justice Education, 10(1),* 75-85.

Shapiro, E. G. (1990). Effect of instructor and class characteristics on students' class evaluations. *Research in Higher Education, 31*, 135-148.

Shapiro, S. S. & Wilk. M. B. (1965). An analysis of variance test for normality. *Biometrika,* 52, 591-599.

Sonner, B. S. (2000). A is for adjunct: Examining grade inflation in higher education. *Journal of Education for Business, 76*, 5-8.

Stack, S. (2000). Part-time faculty status and student evaluation of teaching: An analysis of criminal justice classes. *Journal of Criminal Justice Education, 11*, 251-265.

Stratton, R. W., Myers, S. C., & King, R. H. (1994). Faculty behavior, grades, and student evaluations. *Journal of Economic Education, 25*, 5-15.

Timmerman, T. (2008). On the validity of RateMyProfessor.com. *Journal of Education for Business, 84*, 55-61.

Wilson, R. (1998). New research casts doubt on value of student evaluations of professors. *Chronicle of Higher Education*, January 16, A12.

Table 1. Variable Descriptive Statistics (N = 407)

Variable	Min	Max	Mean	SD
Dependent variables				
Easiness rating	1.0	5.0	3.29	0.85
Helpfulness rating	1.0	5.0	3.71	0.90
Clarity rating	1.0	5.0	3.66	0.97
Overall quality rating	1.0	5.0	3.70	0.90
Hotness rating	0.0	1.0	0.17	0.37
Independent variables				
White	0.0	1.0	0.93	0.26
Male	0.0	1.0	0.72	0.45
Terminal degree	0.0	1.0	0.75	0.43
Mean publications per year (logged)	-6.91	1.81	-1.92	2.74
Practitioner experience	0.0	1.0	0.49	0.50
Years of teaching experience	1	37	13.70	9.44
Research I institution	0.0	1.0	0.36	0.48
Number of raters (logged)	0	4.2	1.93	1.35

Table 2. Variable Bivariate Correlations

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Easiness rating	1.0 *												
2. Helpfulness rating	.60 *	1.0 *											
3. Clarity rating	.53 *	.89 *	1.0 *										
4. Overall quality rating	.62 *	.95 *	.96 *	1.0 *									
5. Hotness rating	.09	.16 *	.16 *	.16 *	1.0 *								
6. White	-.01	.09	.22 *	.15 *	.03	1.0 *							
7. Male	.07	.06	.11 *	.09	.06	-.11 *	1.0 *						
8. Terminal degree	-.19 *	-.20 *	-.20 *	-.20 *	-.02	-.10	.03	1.0 *					
9. Pubs per year (logged)	-.15 *	-.01	.00	.00	.07	-.05	.12 *	.56 *	1.0 *				
10. Practitioner experience	.28 *	.30 *	.36 *	.35 *	.05	.14 *	.16 *	-.50 *	-.32 *	1.0 *			
11. Years of teaching experience	-.16 *	-.14 *	-.13 *	-.16 *	.03	.13 *	.24 *	.22 *	.08	-.19 *	1.0 *		
12. Research I institution	-.04	-.06	-.08	-.08	-.01	.02	.03	.31 *	.47 *	-.30 *	.20 *	1.0 *	
13. Number of raters (logged)	-.00	.04	.11 *	.07	-.04	.02	-.01	-.09	-.12 *	.14 *	-.01	-.01	1.0 *

Significance level: * $p < .05$.

Table 3. Regression models of ratings

Variables	Easiness Beta (SE)	Helpfulness Beta (SE)	Clarity Beta (SE)	Overall Quality Beta (SE)	Hotness Coefficient (SE)
White	-.03	.07	.21***	.13**	.30
	(.16)	(.17)	(.17)	(.16)	(.57)
Male	.07	.03	.11*	.07	.25
	(.10)	(.10)	(.11)	(.10)	(.33)
Terminal degree	-.04	-.15*	-.10*	-.11*	-.32
	(.12)	(.13)	(.13)	(.13)	(.40)
Mean publications	-.08	.19**	.18**	.18**	.13
per year (logged)	(.02)	(.02)	(.02)	(.02)	(.08)
Practitioner	.24***	.25***	.28***	.29***	.26
experience	(.10)	(.10)	(.11)	(.10)	(.33)
Years of teaching	-.12**	-.12*	-.14**	-.15**	-.05
experience	(.01)	(.01)	(.01)	(.01)	(.19)
Research I	.09*	.01	-.01	-.01	-.10
institution	(.10)	(.10)	(.10)	(.10)	(.33)
Number of raters	-.05	.01	.08*	.04	-.07
(logged)	(.03)	(.03)	(.03)	(.03)	(.10)
F-test	5.80***	7.66***	12.75***	10.68***	
R^2	.10	.13	.20	.18	.03

Significance level: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4. OLS regression with easiness and hotness

Variables	Helpfulness Beta (SE)	Clarity Beta (SE)	Overall rating Beta (SE)
White	.08*	.22***	.15***
	(.14)	(.15)	(.13)
Male	-.01	.08*	.03
	(.08)	(.09)	(.08)
Terminal degree	-.13*	-.08	-.08*
	(.11)	(.11)	(.10)
Mean publications per year	.23***	.21*	.22***
(logged)	(.02)	(.02)	(.02)
Practitioner experience	.11**	.16**	.15**
	(.09)	(.09)	(.08)
Years of teaching experience	-.05	-.09*	-.08*
	(.00)	(.00)	(.00)
Research I institution	-.04	-.05	-.06
	(.08)	(.09)	(.08)
Number of raters (logged)	.04	.10*	.07*
	(.03)	(.03)	(.03)
Easiness	.56***	.47***	.57***
	(.04)	(.05)	(.04)
Hotness	.09*	.09*	.09*
	(.09)	(.10)	(.09)
F-test	30.33***	28.41***	36.42***
R^2	.43	.42	.48

Significance level: * $p < .05$, ** $p < .01$, *** $p < .001$.