

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to the *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited.

Volume 15, Number 5, May, 2010

ISSN 1531-7714

RateMyProfessors.com: Testing Assumptions about Student Use and Misuse

April Bleske-Rechek and Kelsey Michels
University of Wisconsin-Eau Claire

Since its inception in 1999, the RateMyProfessors.com (RMP.com) website has grown in popularity and, with that, notoriety. In this research we tested three assumptions about the website: (1) Students use RMP.com to either rant or rave; (2) Students who post on RMP.com are different from students who do not post; and (3) Students reward easiness by giving favorable quality ratings to easy instructors. We analyzed anonymous self-report data on use of RMP.com from 208 students at a regional public university and RMP.com ratings of 322 instructors at that university. Our findings suggest that (1) student motivations for posting on the website are wide ranging and moderate in tone; (2) few student characteristics differentiate those who post from those who do not post on the website; and (3) although easiness and quality are highly correlated, discipline differences in easiness but not in quality suggest that students can, and do, discriminate between easiness and quality. We concur with previous researchers (e.g., Otto, Sanford, & Ross, 2008) that, although the site is limited, RMP.com has more validity than generally assumed.

In 2008, Forbes Magazine joined the likes of U.S News and World Report by offering its first annual ranking of best colleges and universities in the United States. In 2009, Forbes Magazine argued that their evaluation system should be taken more seriously compared to others because their system focused less on reputation and money spent, and more on concerns directly facing students, such as whether courses would be interesting and rewarding (Steinberg, 2009). Given the focus of their evaluation system, they noted, 25% of their rankings were based on student evaluations of instructors taken from the website RateMyProfessors.com.

Forbes' use of RateMyProfessors.com data to rank U.S. colleges and universities demonstrates the degree to which the website is known and influencing how people think about higher education. It also raises a number of questions, including the following: Exactly what data are available on this site? Is student input on this site valid?

The Website

RateMyProfessors.com (RMP.com) was launched in 1999 as an outlet for students to rate and voice

commentary on their instructors. On the site, which students visit voluntarily, students use five-point likert type scales to rate their instructors' *easiness* ('How easy are the classes that this professor teaches?' 'Is it possible to get an A without too much work?'), *helpfulness* ('Is the teacher approachable and nice?' 'Is s/he willing to help you after class?'), and *clarity* ('How well does the teacher convey the class topics?' 'Is s/he clear in his presentation?' 'Is s/he organized and does s/he use class time effectively?'). The latter two scores, helpfulness and clarity, are averaged to provide a *quality* score for each instructor. Students also can rate instructors as "hot" (or not hot) by assigning them a chili pepper (or not). And, they can include open-ended responses about instructors. As of 2009, the site held over six million ratings on hundreds of thousands of instructors from over six thousand different universities. Although some instructors have only one or a couple of student posts, there are thousands on the site with 10 or more posts (Felton, Mitchell, & Stinson, 2004).

Empirical Analyses of RMP.com Posts

Systematic analyses of the ratings and comments on RMP.com have appeared only recently (Coladarci & Kornfield, 2007; Felton et al., 2004; Felton, Koper, Mitchell, & Stinson, 2008; Kindred & Mohammed, 2005; Riniolo, Johnson, Sherman, & Misso, 2006; Silva, Silva, Quinn, Draper, Cover, & Munoff, 2008). Despite the limited number of available analyses of RMP.com posts, there are consistent patterns in the findings. First, analyses of both rating scale data (Silva et al., 2008) and of the content and valence of open-ended responses (Kindred & Mohammed, 2005) show that students are more positive than negative in their postings. A greater percentage of open-ended responses are positive in valence than negative in valence, and mean ratings of instructors consistently fall above the midpoint of 3 (on a 1 to 5 scale). Second, there are positive associations among nearly all the rating scale items: Helpfulness and clarity are essentially redundant (Davison & Price, 2009; Otto et al., 2008) and the average of those (quality) is associated with easiness (Coladarci & Kornfield, 2007; Davison & Price, 2009; Felton et al., 2004; Felton et al., 2008). The association between ratings of instructor quality and instructor easiness is consistently strong and positive, and, as with the association between expected grade and instructor ratings on traditional student evaluations of instruction (see Marsh, 1984; Wachtel, 1998), the association in RMP.com data between quality and easiness is contentious because it can be interpreted in a variety of ways (see below). Overall, empirical analyses of RMP.com posts have documented patterns of findings that are strongly reminiscent of the findings on traditional teaching evaluations (Coladarci & Kornfield, 2007; Silva et al., 2008).

Empirical Analyses of Students' Use of RMP.com

Systematic analyses of students' use of RMP.com are sparse. Thus, there is very little available data on how often students visit the website, students' motivations for viewing ratings and posting ratings, and whether or not students who use the site – particularly to post ratings -- differ systematically from those who do not use the site. One recent analysis showed that the majority of students know about the website but less than a third of students have actually posted on it (Davison & Price, 2009). Findings from another study suggest that students generally visit and post with instructor competence and classroom experience in mind, and that students approach other students' comments with caution (Kindred & Mohammed, 2005).

However, that study was limited to a thematic analysis of comments from a small, select group of 22 students who were experienced with using or posting on the site.

Perhaps the limited amount of systematic research on RMP.com can explain competing voices in the available literature. On one hand, researchers have suggested that “findings are consistent with our expectations under the assumption that the ratings reflect(ed) student learning” (Otto et al., p. 364). Further, the corporate world is using these ratings to make decisions, as in the case of Forbes using RMP.com posts to rank colleges on educational quality. On the other hand, other researchers argue that “the information provided by the RMP website is not valid” (Davison & Price, 2009, p. 61) and that “high student opinion survey scores might well be viewed with suspicion rather than reverence” (Felton et al., 2004, p. 91).

Notably, the largest gaps in the empirical research pertain to knowledge of students' use of RMP.com. In our review of the existing scholarly and popular literature, we noticed three assumptions about students' use and misuse of the website that we propose underpin mixed evaluations of the site. Below, we offer evidence for the existence of these assumptions. We propose that testing these assumptions will help clarify whether RMP.com has useful information to offer instructors.

Assumptions about Student Use and Misuse of RMP.com

Assumption 1: Students use RMP.com to rant or rave. One assumption in the literature is that students use RMP.com to either rant or rave about instructors. For example, as noted by Felton et al. (2008): “The motives of students making these posts seem to range from a sincere desire to praise worthy performance to a desire to retaliate that, at its worst, is not much removed from the graffiti on the walls of restrooms” (p. 45). Similarly, Davison and Price (2009) state, “The onus is on the student to log in, register and take the time to post a rating on a particular instructor. This process lends itself to bias, with students who either loved or hated an instructor more likely to post.” (p. 52)

Research on traditional student evaluations of instruction shows clearly that students agree in their judgments of a given instructor (Aleamoni, 1987; Marsh & Roche, 1997). Thus, a given instructor's mean ratings on RMP.com should reflect student consensus about that instructor.¹ If the assumption that students use RMP.com to either rant or rave is correct, instructors'

mean ratings should be bimodal in distribution, with more ratings on both the low and the high ends. In addition, instructors who have particularly low or high ratings should have received a lot of ratings compared to those who fall in the middle. Finally, the assumption suggests that students' reports of why they have ever posted on RMP.com should reflect a desire to either champion or derogate an instructor.

Assumption 2: Students who post on RMP.com are different. A second assumption in the literature is that students who post on RMP.com are different from students who do not post. As noted by Posillico (2009): "However accurate or inaccurate the ratings may be, they are not representative of the whole class.... The main problem that both students and professors agree on is that you don't know who is posting on the web site. You don't know if they went to class, if they went to professor's office hours, if they did the homework, if they studied and overall what grade they got." Davison and Price (2009), in their analysis of the frequency with which students reflect on the easiness of a course, suggest, "students today are not interested in the learning process or the end product of knowledge...Websites like Rate My Professor will continue to cater to these (consumerist) demands." (pp. 61-62). These statements suggest that the students who go on to RMP.com are potentially a select group of jaded, grade-oriented students.

Assumption 3: Students reward easy instructors. A third assumption is that students are biased: they reward easy instructors with high quality ratings. For example, in reference to their documented associations among quality, easiness, and attractiveness ratings on the site, Felton et al. (2004) noted, "...these data raise the possibility that high-quality ratings may have more to do with an instructor's appearance and how easy he or she makes a course than with the quality of teaching" (p. 106). Davison and Price (2009) suggest, "The internal validity of the ratings is highly suspect...we argue that the limited questions on the RMP site are not robust measures of teaching effectiveness (p. 52)...the easier the course, the higher the overall score...Information provided by the RMP website is not valid." (p. 61)

As with the contentious issue of potential bias in traditional student evaluations of teaching (summarized concisely by Coladarci and Kornfield, 2009), there are multiple potential explanations for the association between easiness and quality on RMP.com, and these explanations are not mutually exclusive. First, it is possible that students reward lenient instructors (those

who give "easy As") with high quality ratings. Second, it is possible that high quality (effective) instructors make it easy to learn. Third, it is possible that interested and motivated students both enjoy the instructor's teaching and have an easier time learning (for varied views on the relative weight of these processes, see, e.g., Greenwald & Gillmore, 1997; Heckert, Latier, Ringwald-Burton, & Drazen, 2006; Marsh & Roche, 1997; McKeachie, 1997; Remedios & Lieberman, 2008).

Research on traditional student evaluations of instruction has provided various lines of evidence that effective instructors get high ratings because they make it easy to learn (Marsh, 1984). For example, in multi-section validity studies, in which different sections of students use the same textbook and exams but have different instructors, the instructors of students who perform better on exams receive more favorable ratings (for a review, see Cohen, 1981). Thus, research on traditional student evaluations of instruction suggests that students' ratings are, at least to some degree, valid indicators of instructor quality.

If traditional student evaluations have validity, then RMP.com ratings might, as well, because the same instructors are evaluated similarly on RMP.com and on traditional student evaluations of instruction. Instructors' RMP.com easiness ratings are strongly associated with their student evaluation workload/easiness ratings, and instructors' RMP.com quality ratings are strongly associated with their student evaluation ratings of overall effectiveness (Coladarci & Kornfield, 2007). In addition, ratings of clarity and helpfulness are negatively related to variability in easiness ratings, a pattern expected if RMP.com ratings reflect student learning as opposed to student bias (Otto et al., 2008). Notwithstanding these hints at validity in students posts, there also are consistent positive associations between attractiveness and both easiness and quality ratings on RMP.com (Felton et al., 2004; Felton et al., 2008; Riniolo et al., 2008). The link between instructor quality and attractiveness implicates bias in students' ratings, although it also is possible that instructor attractiveness is systematically tied to instructor personality (e.g., energy, confidence) or student willingness to attend to their instructors.

In summary, we propose that even if students do reward lenient instructors with high quality ratings, it is also possible that high quality (effective) instructors may – by virtue of being effective – make it easier for students to learn. Thus, we expect to find evidence that students do discriminate between easiness and quality.

The Current Research

The objective of the current research, then, was to test these three assumptions about student use and misuse of RMP.com. To test assumptions about the students who post on RMP.com and their motivations for posting, we surveyed 208 undergraduates about their use of RMP.com. To test the assumption that instructors are rewarded for easiness, we analyzed ratings about 322 instructors at that same university.

METHOD

Sample and Measures: Student Use of RMP.com

We collected self-report data from 208 students (155 women, 51 men, 2 unstated) attending the University of Wisconsin-Eau Claire, a regional public university of approximately 10,000 undergraduates (60% female, 40% male) and 525 instructional staff. Participants varied in year in school: 27% freshmen, 19% sophomores, 18% juniors, 25% seniors, and 12% fifth-year seniors or beyond. They also varied in major, with representation from the four major disciplines on campus: 24% Pre-Professional, (Education, Business, and Nursing), 22% Arts and Humanities (English, Foreign Language, History, Art, Religion), 6% Math and Natural Sciences (Biology, Math, Chemistry, Physics, Astronomy, and Computer Science), and 36% Social Sciences (Economics, Psychology, Geography, Political Science, Sociology, and Communication & Journalism). Participant GPAs ranged from 2.0 to 4.0 ($M = 3.23$, $SD = .39$).

In addition to supplying demographic information, participants completed two learning goal orientation inventories and two questions designed specifically to address the tradeoff between learning and grade orientations. Table 1 displays correlations among these learning and grade orientation scales, descriptive statistics for each scale, and inter-item reliability coefficients for each scale. Upon completion of the learning and grade orientation scales, participants responded to a number of items related to their use of RMP.com.

Achievement Goals. Students completed a scale to measure three achievement goal constructs. The scale includes six items to measure mastery, six items to measure performance-approach, and six items to measure performance-avoidance. (In our sample, we included only five of the six items for performance-avoidance because one item, “I wish my

Table 1 Correlations and Descriptive Statistics among Achievement Goals, Learning Goal Orientations, and Tradeoff Questions

Scale	1	2	3	4	5	6
Mastery	--					
Performance-Approach	.12	--				
Performance-Avoidance	-.22*	.43*	--			
Learning Orientation (LO)	-.55*	.08	-.29*	--		
Grade Orientation (GO)	-.44*	.14*	.50*	-.47*	--	
Tradeoffs	.39*	-.15*	-.37*	.46*	-.52*	--
No. items in scale	6	6	5	16	16	2
Mean	5.5	4.6	5.09	3.58	3.22	2.99
SD	.87	1.26	1.24	1.50	.46	.48
Cronbach α	.83	.88	.70	.74	.73	.65

Note. * $p < .05$. Achievement goal items were rated on a seven-point scale, LO and GO items were rated on a five-point scale, and tradeoff items were rated on a six-point scale. Higher tradeoff scores represent stronger orientation toward learning over grade.

courses weren't graded,” substantially reduced the inter-item reliability.) Sample mastery items include, “I desire to completely master the material presented in my courses” and “I want to learn as much as possible from this class.” Sample performance-approach items include, “It is important to me to be better than other students” and “I am motivated by the thought of outperforming my peers in this class.” Sample performance-avoidance items include, “My fear of performing poorly is often what motivates me” and “I worry about the possibility of getting a bad grade in this class” (Elliot & Church, 1997). Students rated their agreement or disagreement with each item on a seven-point likert type scale.

LOGO. Students completed the LOGO (Milton, Pollio, & Eison, 1986). This scale includes 16 items to measure learning orientation (LO) and 16 items to measure grade orientation (GO). Sample LO items include, “It is important for me to understand the content of my courses as thoroughly as possible” and “I find the process of learning new material fun.” Sample GO items include, “My goal in my courses is to get a better grade than most of the other students” and “I think grades provide me a good goal to work towards.” Students rated their agreement or disagreement with each item on a five-point likert type scale.

Tradeoffs. Students were presented with two different situations that had a tradeoff between grading and

learning. The first situation asked participants to choose between two sections of a course required for their degree. With Section A's instructor, the participant would learn a ton but be unlikely to earn an A; with Section B's instructor, the participant would learn next to nothing but be highly likely to earn an A. The second situation asked participants to choose between two elective courses. Course A was guaranteed to be very dull but also very easy. Course B was guaranteed to be very interesting but also very difficult. Students responded to the tradeoff questions on a six-point scale with Section/Course A on one end and Section/Course B on the other. For analyses, responses to the first item were reverse scored so that higher scores consistently represented a choice of "learning" over "grade." Because student responses to the two tradeoff items were consistent ($\alpha = .65$), we created an average of the two items, henceforth referred to as *tradeoffs*. As displayed in Table 1, students' selection of "learning" over "grade" on the tradeoff items was consistent with their standing on the learning and grade orientation scales and achievement goal orientation scales.

RateMyProfessors.com. Students responded to a number of questions regarding their use of the website. First, students reported the number of times they had viewed ratings on RMP.com and the number of times they had posted ratings on RMP.com. Second, we provided in alphabetical order a list of the seven pieces of information about an instructor that are available on RMP.com: clarity rating, easiness rating, helpfulness rating, hotness total (number of chili peppers), number of postings, open-ended comments, and quality rating. Participants ranked these items in order of how important they are to them when finding out about an instructor (1=most important, 7=least important). Third, for those students who had posted ratings on an instructor, we asked them to think about the most recent time they had posted. They then used a seven-point rating scale (*not at all* to *very much so*) to rate 18 different reasons for posting ratings on an instructor. Sample reasons included: "I thought the workload was too heavy," "I thought the course load was ridiculously easy," "I was angry with the instructor," and "I thought the instructor was an excellent teacher." The full list appears in the Appendix.

Sample and Measures: Instructor Ratings on RMP.com

We sought ratings from instructors from the same public university our student participants attended, the University of Wisconsin-Eau Claire. At the time of our

data collection in July 2008, there were 571 unique instructors with one or more postings (some instructors listed on the site had since left the university or retired). In accord with previous researchers' use of 10 postings as a standard of inclusion (Felton et al., 2004; Coladarci & Kornfield, 2007; see Footnote 1), we focused our analyses on 322 instructors who had more than 10 postings on RateMyProfessors.com. For each of those 322 instructors, we recorded the mean ratings for easiness, helpfulness, clarity, and quality. We also recorded each instructor's department, sex, number of posts, and number of semesters at the university since the fall of 1999 (RMP.com was launched in 1999). Instructor departments were divided into the same four disciplines as were the students: Arts & Humanities (34%), Pre-Professional (26%), Social Sciences (18%), and Math and Natural Sciences (22%). For each instructor, we calculated a weighted measure of posting frequency by dividing the number of *total posts* by the *number of semesters at the university* since fall of 1999. This weighted frequency translates into mean number of posts per semester. The measure is potentially limited in that RMP.com was not as well-known in 1999 as it subsequently became, but it does allow for basic discriminations between instructors who have received many ratings by virtue of being at the university for a long time and instructors who have received many ratings despite being at the university for only a short time.

RESULTS

Assumption 1: Students use RMP.com to rant or rave.

If the assumption that students use RMP.com to either rant or rave is correct, instructors' mean ratings should be bimodal in distribution, with more ratings on both the low and the high ends. Figure 1 shows that, contrary to this expectation (and in accord with previous findings; e.g., see Riniolo et al., 2006), ratings overall were more frequently positive than negative and not bimodal in distribution. Also contrary to the assumption, instructors rated as worse or better in quality were not rated any more frequently than others. As displayed in Figure 2, neither instructor quality nor instructor easiness was related to how frequently an instructor was rated, in either a linear ($ps > .44$) or nonlinear ($ps > .55$) fashion. Figure 3 emphasizes this lack of association between rating frequency and quality. With the sample of instructors split into thirds based on mean quality ratings, those who were rated as particularly low or high

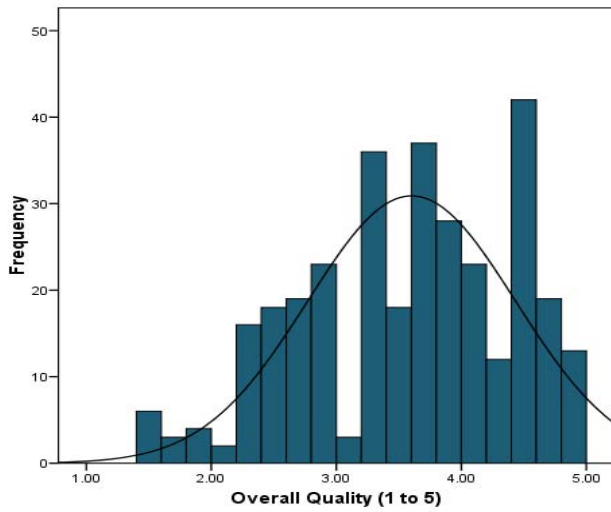


Figure 1. Distribution of quality ratings, for 322 instructors with 10 or more postings.

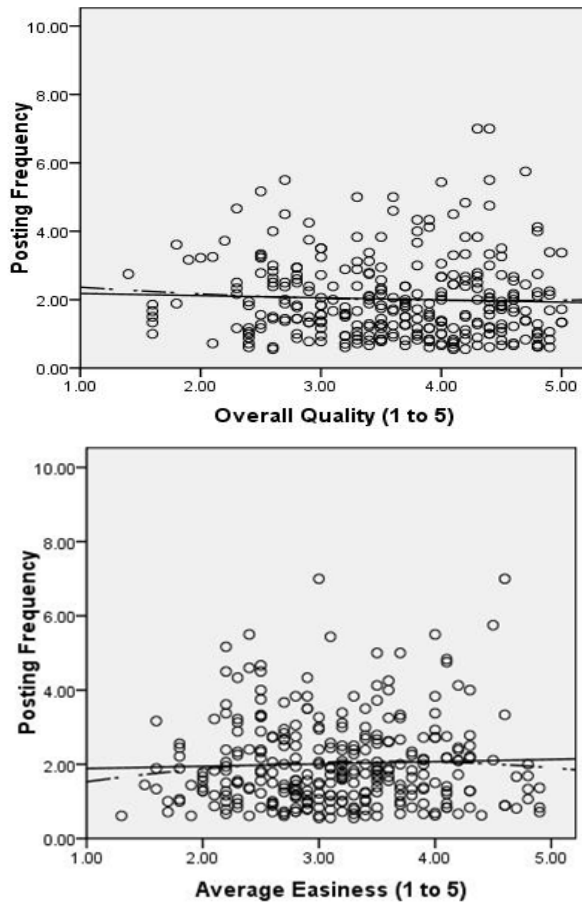


Figure 2. Scatter plot displays of the lack of association between weighted posting frequency and instructor easiness rating (upper panel), and the lack of association between weighted posting frequency and instructor quality rating (lower panel).

in quality received no more ratings than did those who fell in the middle, $F(2, 296) = .69, p = .50$, partial $\eta^2 = .005$. This same pattern was replicated when we split instructors into thirds based on mean easiness ratings, $F(2, 296) = .53, p = .59$, partial $\eta^2 = .004$.

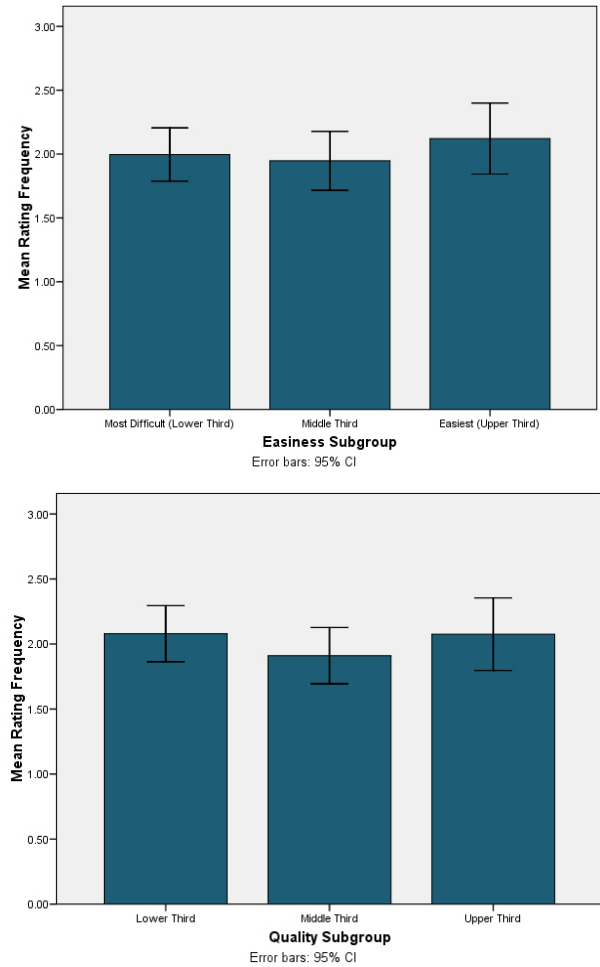


Figure 3. Weighted posting frequency as a function of instructors' placement in bottom, middle, or top third of the distribution of easiness ratings (upper panel) and quality ratings (lower panel).

Finally, if students post on RMP.com to either champion or derogate an instructor, those motives should be reflected in their ratings of the importance of those reasons for posting on the site. Table 1 provides the mean importance given by students to 18 reasons for posting. The results show no clear pattern of what motivates students to post on RMP.com, whether it be to rant or rave. Course difficulty and course easiness were ranked as some of the least important reasons for posting, with instructor helpfulness, fairness, and clarity of explanation receiving higher ratings. Our pattern of

findings reinforces Kindred and Mohammed's conclusion that,

"...While issues such as personality and appearance did enter into the postings, these were secondary motivators compared to more salient issues such as competence, knowledge, clarity, and helpfulness." (p. 11)

Assumption 2: Students who post differ from students who do not post on RMP.com.

A total of 84% of respondents (169 of 197) had visited RMP.com at least once and the majority of those had visited five or more times. A total of 23% of respondents (44 of 194) indicated that they had *posted* at least once on RMP.com, giving us 80% power to detect medium effects (Cohen, 1977) in comparisons of the two groups. For subsequent analyses we combined those who had posted only once ($n = 19$) with those who had posted two ($n = 8$), three ($n = 7$), four ($n = 3$), or five or more times ($n = 7$).

We conducted a number of analyses to test the assumption that students who had posted on RMP.com differ from those who had not. Our analyses revealed that students who had posted differed from those who had not posted in two ways. First, gender was marginally related to likelihood of posting, $\chi^2(1, N = 194) = 3.72, p = .05$, Cramer's $V = .14$. In our sample, 33% of men had posted, whereas only 19% of women had posted. Second, discipline was associated with having posted on the site, $\chi^2(4, N = 194) = 12.59, p = .01$, Cramer's $V = .26$. Twenty-five percent of students in the hard sciences had posted, as had 33% of those in the social sciences and 24% of those in pre-professional disciplines; in contrast, only 7% of students in the arts and humanities, and 10% of undeclared students, had posted. In summary, women and students in the arts and humanities were significantly under-represented among those who had posted on RMP.com.

In every other characteristic we measured, however, students who had posted on RMP.com did not differ from students who had never posted. First, the two groups did not differ in GPA (Posted: $M = 3.14, SD = .40$; Never posted: $M = 3.25, SD = .38$), $t(173) = 1.61, p = .11$. Second, the two groups did not differ in year in school, $\chi^2(4, N = 192) = 4.71, p = .32$. Third, the two

groups did not differ in their learning goal orientations. As illustrated in the three panels of Figure 4, independent samples t -tests showed that students who had posted on RMP.com did not differ from their non-posting counterparts on the LOGO, their achievement goals, or their responses to the tradeoff questions (p s ranged from .15 to .92).

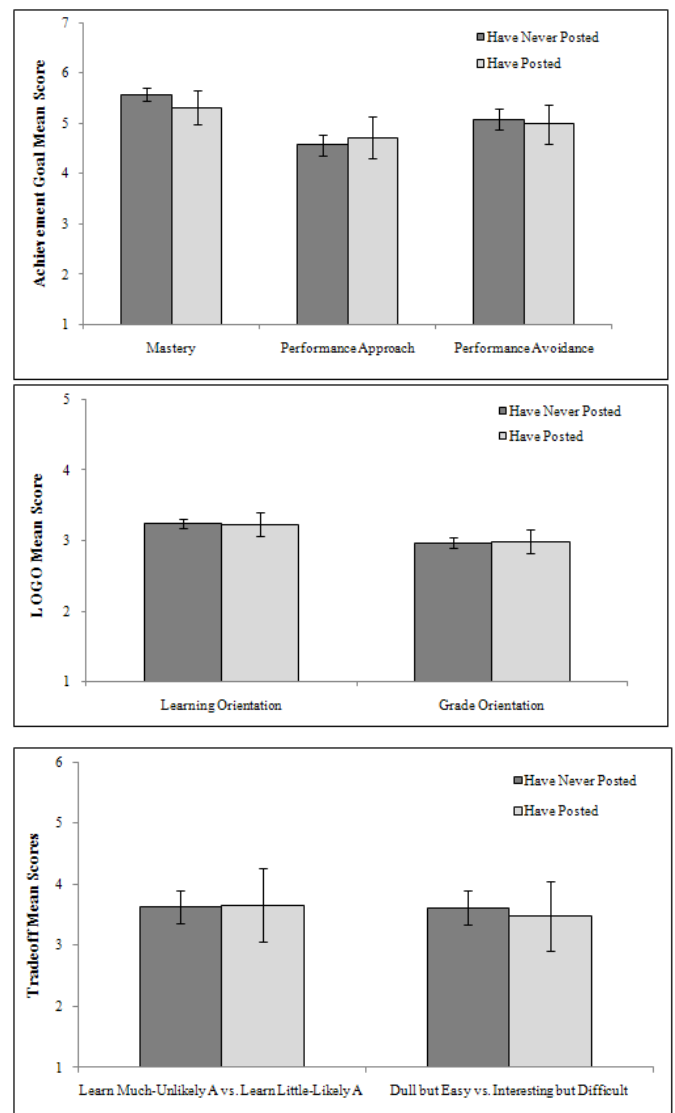


Figure 4. Comparison of students who had posted on RMP.com with students who had not posted on RMP.com: achievement goals (upper panel), learning and goal orientations (middle panel) and tradeoff questions (lower panel). Error bars represent 95% confidence intervals.

Assumption 3: Students reward easiness by giving favorable ratings to easy instructors.

As in previous studies of RMP.com ratings (Felton et al., 2004; Felton et al., 2008), instructors in our sample who were rated as high in quality were also rated as easy, $r(320) = .52, p < .001$. This association is displayed in Figure 5. In Figure 6, we display the match between instructors' placement into thirds for easiness ratings

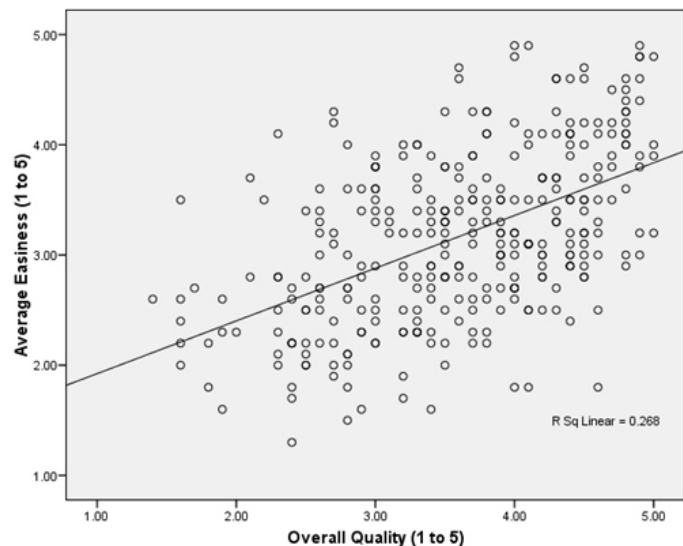


Figure 5. Scatterplot of the positive association between easiness and quality ratings for 322 instructors.

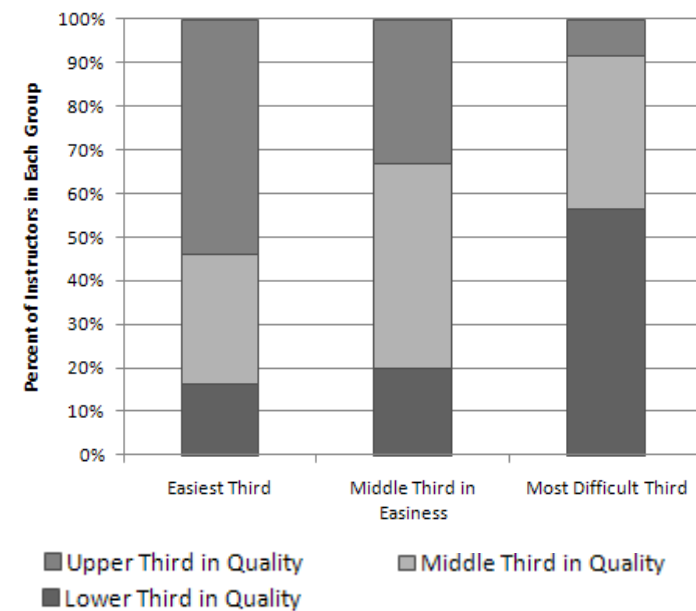


Figure 6. An alternate view of the association between easiness and quality: the percent of high, middle, and low quality instructors in each easiness subgroup.

and instructors' placement into thirds for quality ratings. The graph illustrates that, although the association between easiness and quality is strong, there are notable exceptions: 16% of the "easiest" instructors are also rated among the lowest third in quality; 8% of the "hardest" instructors are also rated as among the highest third in quality.

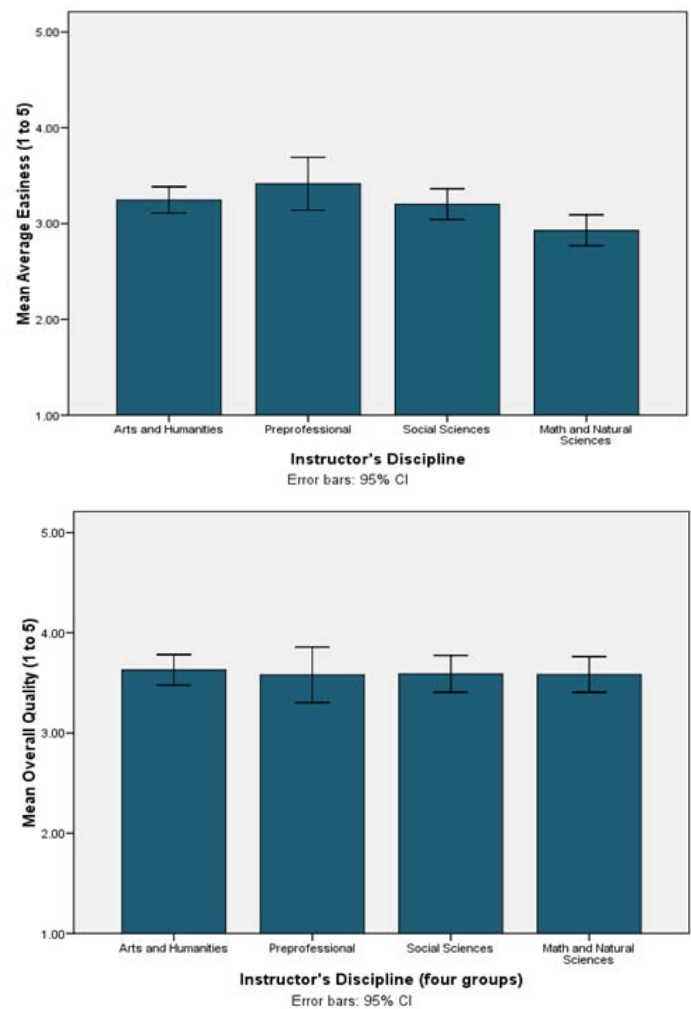


Figure 7. Instructor easiness ratings (upper panel) and quality ratings (lower panel) as a function of instructor discipline.

To test specifically whether students' perceptions of easiness and quality are wholly overlapping, we investigated students' perceptions of easiness and quality by discipline. Past research suggests that students rate math and science instructors as more difficult (Felton et al., 2008); in our sample, too, instructors from math and the natural sciences were rated as less easy (more difficult) than were instructors in the other disciplines, $F(3, 318) = 4.77, p = .003$, partial $\eta^2 = .04$ (math and natural sciences versus arts and humanities, $p = .02$;

math and natural sciences versus preprofessional, $p = .006$; math and natural sciences versus social sciences, $p = .10$). This association between easiness and discipline is displayed in the upper panel of Figure 7. As displayed in the lower panel of Figure 7, however, instructors from math and the natural sciences were rated as just as high in quality as were instructors in other disciplines, $F(3, 318) = .07$, $p = .98$, partial $\eta^2 = .00$. Together, our findings replicate the association between easiness and quality but clarify the possibility that if bias is operating, it is operating in addition to valid student perceptions of instructor quality.

DISCUSSION

We conducted this study to test three notions in the literature about RateMyProfessors.com: first, that students post on RMP.com to either rant or rave; second, that students who post on RMP.com are different from students who do not; and third, that students reward easy instructors with high quality ratings. Our findings put each of these notions in doubt.

Several patterns in our data suggest that students do not typically post on RMP.com to rant or rave. First, the distribution of instructor ratings followed a near-normal distribution, with ratings more on the positive than negative end and clearly not bimodal. Second, instructors of varying quality ratings did not differ in the frequency with which students rated them. Initially, one might be surprised at this because Davison and Price (2009) found a weak negative association ($r = -.13$) between quality and number of ratings. However, they did not weight the number of ratings by how many semesters each instructor in the sample had been at the given institution. That said, our findings should be interpreted with some caution for a different reason: Research has yet to document the extent to which mean ratings about a given instructor represent a consistent view from the students who are posting. We might assume that the mean rating for a given instructor represents consistent student opinion about that instructor, given that on traditional student evaluations students in the same course section show much agreement about their instructor (Marsh & Roche, 1997). However, it is possible that postings on RMP.com are different. For example, perhaps there are two students who rave for every one student who rants, resulting in a more positive than negative (but still moderate in tone) overall rating. Future research needs to assess inter-student rating reliability. The method we utilized in the current study, which involved focusing on student reports and instructor ratings from a single

university, could serve as a model not only for testing hypotheses about use and misuse of RMP.com, but also for answering questions related to inter-student reliability in response.

Students' self-reported reasons for posting about instructors give a third reason to question the assumption that students post to rant or rave. In their responses to various reasons for posting on an instructor, students rated *warning others about an instructor* or *communicating that an instructor was excellent* as the most important reasons; yet they also rated *interest in the course* and *instructor helpfulness* to be important. And, they rated the *course being too easy* or *the course being too hard* as less important reasons to post. Thus, there was no systematic trend for students' reports to involve posting to praise or defame a given instructor for reasons unrelated to the learning experience.

Our data also do not support the notion that students who post on RMP.com are different from students who do not post. In our sample, those who had posted on RMP.com were less likely to be female and in the arts and humanities (and more likely to be in the social sciences); beyond that, however, they differed little. The two groups of students held similar GPAs, were at similar points in their college career, and – by three separate indicators – were similar in their focus on grading versus learning. Thus, students who post are not more focused on “the grade” than are other students, as suggested by faculty with whom we have discussed RMP.com. These non-significant comparisons attest to the potential validity of the information available from students who post on RMP.com.

Finally, our data confirmed the possibility of bias in students' responses by replicating the positive link between ratings of instructor easiness and ratings of instructor quality. Our data also, however, clarified that it is misguided to jump to the conclusion that the association between easiness and quality is necessarily a product of just bias. In our sample, students (as expected) rated instructors in the natural sciences as more difficult than other instructors, but they did not rate them as any lower in quality. We also found that 16% of the “easiest” instructors also had the lowest quality ratings. These findings suggest that, at least for this sample of instructors, easiness did not make quality. We think that it is crucial for researchers to investigate all possible explanations for the positive association between easiness and quality, such as that quality instruction facilitates learning, or that a highly motivated

and interested student mindset leads to high ratings for both easiness and instructor quality.

CONCLUSION

Students of today are avid consumers living in a rapidly expanding information age. For better or worse, they expect to use the internet to find answers to their questions and to influence others. In an interesting parallel to our data showing that 84% of students have visited RMP.com and 23% have posted on RMP.com, a recent report on online consumer behavior showed that 84% of Americans say they attend to online customer evaluations in their decisions to purchase or product or service and 28% of Americans say they have posted their own feedback on the web about a product or service (Hosford, 2009). If education is a purchase item, and if instructor evaluation ratings and the course information that students want are not readily available, online rating sites such as RateMyProfessors.com are an obvious avenue for students to pursue. It seems unlikely, therefore, that these sites will fade away.

On a more positive note, the results of our analyses place assumptions about student misuse in question. Students who post do so for a variety of reasons and not just to complain or exclaim; they are similar academically to students who do not post; and patterns in their ratings suggest that easiness and quality are not synonymous to them. As noted by other researchers, we do not mean to imply that the system is perfect. Anyone browsing through RMP.com will find cases of students who clearly post with vengeance in mind. Other researchers have suggested that websites such as RMP.com take steps toward improving their rating system, such as limiting commentary to those who have clearly taken a course with a given instructor and adding additional questions that tap more elements of the learning experience (Davison & Price, 2009; Otto et al., 2008). We concur, and we hope that documenting the existing utility of the site's ratings will encourage researchers and other interested parties to help online rating sites implement changes that will enhance their utility.

REFERENCES

- Aleamoni, L. M. (1987). Student rating myths versus research facts. *Journal of Personnel Evaluation in Education*, 1, 111-119.
- Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, 51, 281-309.
- Coladarci, T., & Kornfield, I. (2007). RateMyProfessors.com versus formal in-class student evaluations of teaching. *Practical Assessment, Research & Evaluation*, 12, 1-15. Retrieved from: <http://pareonline.net/getvn.asp?v=12&n=6>
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.
- Davison, E., & Price, J. (2009). How do we rate? An evaluation of online student evaluations. *Assessment & Evaluation in Higher Education*, 34, 51-65.
- Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, 72, 218-232.
- Felton, J., Koper, P. T., Mitchell, J., & Stinson, M. (2008). Attractiveness, easiness, and other issues: Student evaluations of professors on ratemyprofessors.com. *Assessment & Evaluation in Higher Education*, 33, 45-61.
- Felton, J., Mitchell, J., & Stinson, M. (2004). Web-based student evaluations of professors: The relationships between perceived quality, easiness, and sexiness. *Assessment & Evaluation in Higher Education*, 29, 91-108.
- Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 52, 1209-1217.
- 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 learning? *College Student Journal*, 40, 588-596.
- Hosford, C. (2009, April 20). *Study: Online reviews significantly influence purchasing decisions*. Retrieved from: <http://www.btobonline.com/apps/pbcs.dll/article?AIID=20090420/FREE/904209991/1078>.
- Kindred, J., & Mohammed, S. (2005). "He will crush you like an academic ninja!": Exploring teacher ratings on Ratemyprofessors.com. *Journal of Computer-Mediated Communication*, 10(3), article 9; <http://jcmc.indiana.edu/vol10/issue3/kindred.html>
- Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76, 707-754.
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52, 1187-1197.
- McKeachie, W. J. (1990). Research on college teaching: The historical background. *Journal of Educational Psychology*, 82, 189-200.
- Milton, O., Pollio, H. R., & Eison, J. A. (1986). Learning for grades versus learning for its own sake. *Making sense of college grades* (pp. 124-149). San Francisco: Jossey-Bass Inc.

- Otto, J., Sanford, D. A., & Ross, D. N. (2008). Does ratemyprofessor.com really rate my professor? *Assessment & Evaluation in Higher Education*, 33, 355-368.
- Posillico, F. (2009, April 1). *Rate my professor's...changing scheduling decisions?*. Retrieved from <http://www.sbstatesman.com/2.873/rate-my-professor-s-changing-scheduling-decisions-1.35665> (Stony Brook, NY).
- Remedios, R., & Lieberman, D. A. (2008). I liked your course because you taught me well: The influence of grades, workload, expectations and goals on students' evaluations of teaching. *British Educational Research Journal*, 34, 91-115.
- 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, 19-35.
- Silva, K. M., Silva, F. J., Quinn, M. A., Draper, J. N., Cover, K. R., & Munoff, A. A. (2008). Rate my professor: Online evaluations of psychology instructors. *Teaching of Psychology*, 35, 71-80.
- Steinberg, J. (2009, August 6). *As Forbes sees it, West Point beats Princeton (and Harvard, too)*. Retrieved from <http://thechoice.blogs.nytimes.com/2009/08/06/west-point/>
- Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education*, 23, 191-213.

Citation

Bleske-Rechek, April & Michels, Kelsey (2010). RateMyProfessors.com: Testing Assumptions about Student Use and Misuse. *Practical Assessment, Research & Evaluation*, 15(5). Available online: <http://pareonline.net/getvn.asp?v=15&n=5>.

Note:

1. Coladarci and Kornfield (2007) investigated associations between RMP.com ratings and student evaluation ratings, for instructors with nine or fewer ratings on RMP.com compared to instructors with more than nine ratings. In further support of consistency among students in their perceptions of the same instructor, the associations between RMP.com ratings and student evaluation ratings were consistently strong and positive, both for instructors with nine or fewer ratings and for instructors with more than nine ratings. Cornfield and Koladarci suggested that the associations were even stronger for those with more than nine ratings on RMP.com, but the absolute differences between the correlation coefficients were small (e.g., .70 versus .67) and the researchers did not report statistics on whether the correlation coefficients were significantly different from each other. That RMP.com ratings and student evaluation ratings coincide is particularly notable in light of the fact that a given instructor's mean RMP.com ratings come from students in different courses taught by that instructor. Student consistency in their perceptions of instructors appears to be the explanation: Research on student evaluations has demonstrated that the same instructor's ratings from two different course correlate, on average, at .61, and that the same instructor's ratings from two different offerings of the same course correlate, on average, at .72 (Marsh, 1981).

Corresponding Author:

April Bleske-Rechek
Associate Professor of Psychology
University of Wisconsin-Eau Claire
105 Garfield Avenue
Eau Claire, WI 54702-4004
bleskeal [at] uwec.edu

Appendix

Students' Reasons for Posting on RateMyProfessors.com

	Mean	sd
⁶ I felt it was my duty to warn others about the instructor.	4.72	2.50
³ I thought the instructor was an excellent teacher.	4.53	2.53
¹⁰ I wanted to give others tips for doing well in the course.	4.49	2.11
² I thought the instructor was very helpful.	4.49	2.51
¹⁸ I thought the course was really interesting.	4.44	2.38
⁹ I thought the instructor was not at all helpful.	4.19	2.52
⁵ I thought the workload was too heavy.	3.70	2.08
¹⁴ I thought the instructor was unfair to me or others in the class.	3.67	2.64
¹² I found the instructor's explanations difficult to understand.	3.67	2.26
¹ I was angry with the instructor.	3.58	2.42
⁴ I thought the instructor was boring.	3.49	2.11
¹⁶ I felt like the class was a waste of time.	3.44	2.43
⁸ The instructor had bad ratings, but I thought s/he was a good instructor.	3.35	2.37
¹¹ I thought the course was too hard.	3.26	2.02
⁷ I just felt like posting.	3.00	2.00
¹⁵ The instructor had good ratings, but I thought s/he was a bad instructor	2.70	2.26
¹³ I thought the course was ridiculously easy.	2.60	1.90
¹⁷ I thought the course demanded very little of my time outside of class.	2.56	1.88

Note. Means are based on 43 students who completed the questionnaire (out of 44 students who reported that they had posted ratings on RateMyProfessors.com). Superscripted values represent the order in which items were presented on the questionnaire.