# Student Consensus on RateMyProfessors.com

April Bleske-Rechek, Amber Fritsch, and Brittany Henn
University of Wisconsin Eau Claire

## Background

• RateMyProfessors.com (RMP.com) is a widely used public forum on which students use five-point scales to rate their instructors on characteristics such as Clarity, Helpfulness, Overall Quality (Clarity & Helpfulness), and Easiness ("how much work it is to get a good grade").

• Despite RMP.com's widespread use and increasing influence,[1] there are few systematic analyses of the ratings on RMP.com, and pop and scholarly evaluations of the site are mixed.[2, 3, 4]

• Previous research out of our lab provided some evidence of the validity of ratings on the site: Students report rating instructors for a variety of pedagogical (rather than emotional) reasons, and students who post ratings do not differ from other students in their learning goals or grade orientation.[5]

• In the current study, we further test the assumption that students' postings on RateMyProfessors.com reflect students' emotions rather than objective appraisals of the instruction they have received.

• If students' postings reflect emotional responses to the instructor, then students should vary widely in their ratings about a given instructor. That is, each student's ratings should include varied sources of bias and error, and only by aggregating many students' responses should there be consensus about an instructor. According to this logic, the number of ratings should be negatively associated with degree of variance in the ratings – that is, it should take many ratings to get consensus around the mean.

• However, if students are not ranting and raving, but rather providing objective assessments of instructor pedagogy, then the number of ratings should not be related to variance in ratings – that is, it should not take many ratings to get consensus around the mean.

## Method

• We selected instructors from a single university who had 10 or more student ratings on RMP.com.

• First, we created a single data set for each of the instructors. Each instructor's data included quality, helpfulness, clarity, and easiness ratings of each student who had rated that instructor; we computed descriptive statistics ($M$, $SD$, and variance) for each of those variables.

• Then, we compiled an umbrella dataset that included all the instructors. For each instructor, we recorded their sex and discipline, and their means, standard deviations, and variances (as described above), as well as the number of student raters that had contributed to those summary statistics.

• After removing six outlier instructors who had more than 87 ratings, the final dataset included 367 instructors with anywhere between 10 and 86 ratings. All disciplines were well-represented.

## Results

### Instructors Vary in Perceived Quality and Easiness
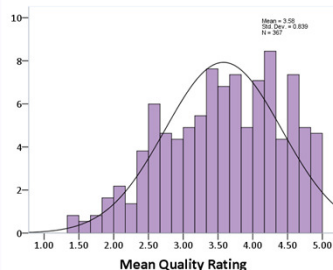


The histogram to the left shows the percent of instructors with each mean quality rating. The instructors under investigation here show a similar distribution of ratings as found in other research on RMP.com[6]: Instructors differ widely in how high or low in quality they are rated, but the distribution of means is slightly negatively skewed, such that, on average, instructors are rated more positively than negatively in terms of their quality.

The histogram to the right displays the percent of instructors with each mean easiness rating. Again, instructors differ widely in how low or high in easiness they are rated. Ratings follow a normal distribution.
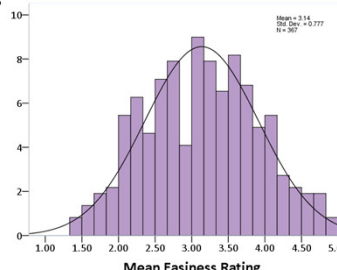
### Student Consensus About Instructors is of Similar Degree Across Number of Student Raters
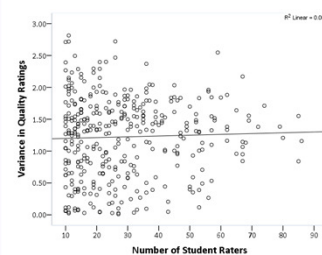


The scatter plot at left displays the association between number of quality ratings and degree of consensus in those ratings; the scatter at right displays the association between number of easiness ratings and degree of consensus in those ratings. Each dot represents a given instructor's number of ratings and degree of consensus in those ratings. As we expected and counter to negative assumptions about the reliability of student ratings on the site, degree of variance in a given instructor's ratings is not associated with how many students have rated them. In other words, instructors with 10 ratings show the same degree of consensus in their ratings as do instructors with 50 ratings. The scatter at left shows an $r(367) = .03$, $p = .52$; at right $r(367) = .09$, $p = .09$. If anything, the trend at right suggests that more variance in easiness ratings is tied to more raters, not fewer. Within each sex and within each discipline, number of ratings was not associated with degree of variance in instructor quality ratings, all $ps > .07$, values for $r$ ranged from -.18 to +.15.
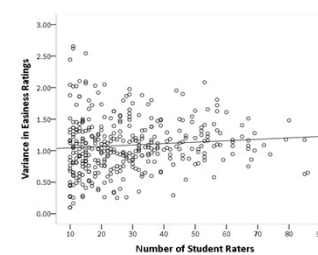
### Students Show the Most Consensus about Instructors with Really High or Low Mean Ratings
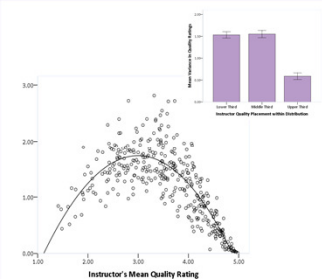


Although degree of variance in student ratings is not tied to how many students have provided ratings, degree of variance in student ratings about instructors is tied to the overall perception of their quality (shown at left, quadratic $F(2, 364) = 424.86$, $p < .001$, $R^2 = .70$) and easiness (shown at right, quadratic $F(2, 364) = 66.25$, $p < .001$, $R^2 = .27$). The effect is very robust for quality, and replicated by sex and discipline of instructor. Instructors with very high mean quality ratings showed very little variance in students' ratings (or strong consensus) – in some cases essentially no variance at all.
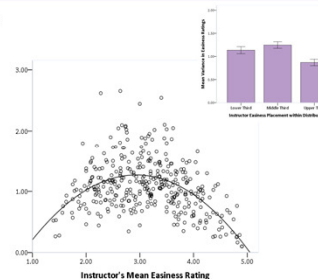
### Variance in Student Ratings follows Quality, not Easiness
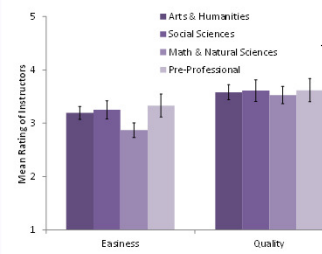


The bar graph at right displays a finding we have documented previously[5]: The 103 instructors in Math & Natural Sciences departments are rated as less easy than are instructors in each of the other disciplines (135 Arts & Humanities, 66 Social Sciences, and 63 Pre-professional), $F(3, 363) = 6.21$, $p < .001$, partial $\eta^2 = .05$, all pair-wise $ps \leq .01$. However, instructors in Math & Natural Sciences are rated similarly in quality, $F(3, 363) = 0.21$, $p = .89$, partial $\eta^2 = .002$. This finding implies that students are distinguishing between easiness and quality.

In further support of students as objective judges of instruction, the bar graph at right shows that variance in students' ratings of both easiness and quality do not differ by discipline. That is, students show similar consensus about their instructors, regardless of the type of discipline those instructors are in (easiness $F(3, 363) = 0.68$, $p = .56$, partial $\eta^2 = .006$; quality $F(3, 363) = 0.10$, $p = .96$, partial $\eta^2 = .001$).
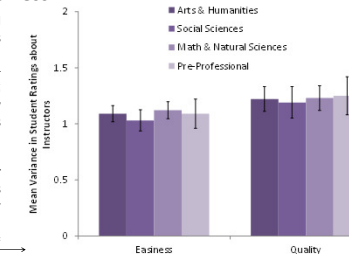
## Discussion

• Research suggests that a majority of students use RateMyProfessors.com to either view or post ratings about instructors at their institution.[5] However, scholars have suggested that student ratings are not valid – that students are biased, grade-oriented consumers.[3,4]

• In this study we have documented several effects that, taken together, suggest that students provide valid judgments of instruction: (1) the typical instructor is rated as higher in quality than in easiness; (2) student consensus about instructors can be achieved with as few as 10 ratings; (3) students show very strong consensus about those instructors who are perceived to be high in quality; and (4) although one might presume that disciplines differ in how susceptible they are to student subjectivity or bias, the degree of variance (and therefore, consensus) in students' ratings is similar across disciplines.

• In conclusion, we have shown that student consensus can be achieved with relatively few ratings. Our data place common assumptions about RateMyProfessors.com in question, and reinforce previous findings to suggest that, in the aggregate, students are providing objective ratings of quality of instruction.[2,5,7]

## References

1. 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/westpoint/

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

3. Davison, E., & Price, J. (2009). How do we rate? An evaluation of online student evaluations. *Assessment & Evaluation in Higher Education, 34*, 51-65.

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

5. Bleske-Rechek, A., & Michels, K. (2010). RateMyProfessors.com: Testing assumptions about students use and misuse. *Practical Assessment, Research & Evaluation, 15*, 1-12.

6. 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.

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

## Acknowledgements