

aepp.net

Association on Employment Practices and
Principles

**Proceedings of the
13th Annual International Conference 2005**
ISBN 1-878583-64-6

Conference Theme:
The Changing Face of Organizational Life

October 6-8, 2005
Baltimore, Maryland, USA

Edited by

Melissa A. Parris
University of Western Sydney

Margaret H. Vickers, PhD
University of Western Sydney

www.aepp.net

Does ratemyprofessor Rate My Professor?

James Otto, Towson University
Douglas Sanford, Towson University
Douglas Ross, Towson University

Abstract

This article explores and analyzes the relationships between learning and online student evaluations of university faculty. Online faculty ratings are a new, and quite different source of faculty evaluation data that has recently become available via the world wide web. Analysis of online ratings for 400 randomly selected instructors finds the pattern of relationships largely consistent with student learning. These results suggest that online ratings may have some use as evaluation measures of instructor performance for both hiring and promotion and tenure purposes.

Keywords: Online faculty ratings, promotion and tenure, evaluation

Introduction

Student faculty evaluations are an important measure of teaching effectiveness and are a consideration for promotion and tenure at many educational institutions. Extensive research has been done on student evaluations of faculty. These evaluations typically are based on forms that are filled out anonymously by students in a classroom using formal, well defined, and controlled processes (Centra, 2003; Martin, 1998; Read et al., 2001). We extend this research by exploring and analyzing a new, exciting, and quite different, source of faculty evaluation data that has recently become available on the world wide web; online faculty evaluation/rating sites. We investigate how well these online ratings are related to student learning. In particular, we investigate whether online ratings measure student learning or some other universal measure of professors, such as popularity. We also examine the possible use of online ratings in promotion, tenure and merit deliberations, as well as in hiring decisions on new faculty.

Online faculty rating sites include RateMyProfessor.com, PassCollege.com, ProfessorPerformance.com, RatingsOnline.com, and Reviewum.com (Stone, 2003; Foster, 2003). One may argue that data from these sites are characterized by bias, and are therefore not of value as a measure for faculty performance or student learning. However, when one examines the amount of traffic at these sites it becomes evident that online rating sites are popular with students and thus should probably be taken seriously by faculty. Very probably, some students use the data on these sites to develop expectations of their professors and to determine their schedules, which indirectly affects faculty teaching loads and student expectations. For example, RateMyProfessor.com has received over 3.8 million ratings on over 590,000 professors at better than 4800 schools (RateMyProfessor, August, 2005).

Little research has been done on the faculty rating data on these sites. Felton, Mitchell and Stinson (2004) analyzed professor overall quality ratings as a function of easiness and sexiness ratings. Otto et al. (2005) investigated the interrelationships between online rating variables. We extend this earlier research by examining whether online student ratings reflect learning, and are therefore a legitimate basis for evaluating professors

Online ratings may differ from traditional faculty evaluations received at the end of the semester. As summarized below, Otto et al. (2005) describe a number of issues related to bias, availability of ratings, and purpose of ratings.

- **Sample Representation.** This source of bias stems from the lack of controls in online ratings regarding who can rate a professor online. Students or non-students can post ratings even if they have not taken an instructor.
- **Sample Selection Bias.** This source of bias stems, by definition, from online raters being a self-selected sample. Online ratings may reflect not those of all students who have taken an instructor, but only a few who have strong feelings about the instructor.
- **Results Availability.** Online evaluations are instantly publicly available at any time to anyone with access to a browser.
- **Rating Criteria and Student Comments.** Comments and rating criteria are focused on helping students make informed decisions concerning faculty versus helping faculty improve their performance.

Students may respond to this rating information by choosing instructors who are best suited to their learning styles. Online ratings are available to instructors, too, and in response to bad online ratings, instructors may work to improve their performance. One can argue that the 'results availability' and 'rating criteria and student comments' issues could have virtuous effects on instructor performance and student learning. These virtuous effects are dependent on the validity of the ratings as a measure of student learning. There is a substantial literature on the extent to which traditional faculty evaluations are a valid measure of student learning (Harrison et al., 2004; March & Roche, 1997; 2000; McKeachie, 1997). These studies indicate that 'proving' bias or lack of bias in student evaluations of teaching is difficult, largely because of the challenge of reliably measuring student learning. However, research findings do consistently hold that learning is positively associated with instructor clarity and instructor helpfulness. Research findings also hold that course difficulty is not linearly associated with student learning. Centra (2003) elaborates on this finding. His study confirms a lack of linear relationship between course difficulty and student learning, but suggests a nonlinear relationship. Students may learn best not when the course is too easy or too difficult, but when the difficulty is between these extremes, or 'just right.'

Objective and Approach

Our objective is simply to look at the pattern of association between components of online ratings and test whether they are consistent with the pattern expected of valid measures of

student learning. We understand that online ratings are subject to possible bias, but also think that students may be providing their peers with accurate and dependable information. It is possible that online ratings of instructors, on average, are actually accurate and valid measures of student learning provided by the instructors.

Previous research indicates that the ratings components of instructor clarity in presentation of course material, instructor helpfulness as perceived by students, and course easiness will associate with student learning in the pattern shown on Figure 1. Our data source is from ratemyprofessors.com, and the definition of clarity, helpfulness, and easiness, was provided by the website. Our data source is from ratemyprofessors.com, and the definition of clarity, helpfulness, and easiness, was provided by the website.² Previous research demonstrates that both instructor clarity and instructor helpfulness are linearly positively correlated with learning, and we show this with the two positively sloped lines for clarity and helpfulness. These lines show that students who rate an instructor high on clarity and helpfulness will tend to have a high level of learning. Students who rate an instructor low on clarity and helpfulness will tend to have a low level of learning.

The association between learning and easiness found in the literature is less well-demonstrated. Most studies show that course difficulty (the opposite of easiness) is either insignificantly or positively associated with student learning (Marsh & Roche, 1997; 2000). However, Centra (2003) suggests that the relationship between easiness and student learning is non-linear. This finding is logically compelling, as instructors can impede learning by either being too easy, and not challenging students, or too difficult, and frustrating them. On Figure 1, we show this non-linear relationship with a non-linear inverse "U" relationship between student ratings for easiness and student learning. The optimal level of learning will be associated with instructors that have levels of easiness in-between the high and low ratings.

The relationships on Figure 1 imply that instructors who are perceived to be helpful to students and have clarity in their presentation of the course material will have increased learning in their classes. Therefore, if the online ratings reflect student learning, then we expect the following.

H1. Helpfulness and clarity will be positively correlated.

² **Easiness** - This is definitely the most controversial of the three rating categories, which is why it is NOT included in the "Overall Quality" rating. Although we do not necessarily condone it, it is certainly true that many students decide what class to take based on the difficulty of the teacher. When rating a teacher's easiness, ask yourself "How easy are the classes that this professor teaches? Is it possible to get an A without too much work?"

Helpfulness - This category rates the professor's helpfulness and approachability. Is the professor approachable and nice? Is the professor rude, arrogant, or just plain mean? Is the professor willing to help you after class?

Clarity - This is the most important of the three categories, at least to many people. How well does the professor convey the class topics? Is the professor clear in his presentation? Is the professor organized and does the professor use class time effectively?

Overall Quality - The Overall Quality rating is the average of a teacher's Helpfulness and Clarity ratings, and is what determines the type of "smiley face" that the Professor receives. Due to popular demand, a teacher's Easiness rating is NOT used when computing the Overall Quality rating, since an Easiness of 5 may actually mean the teacher is TOO easy.

Regarding easiness, we suggest that the variation in easiness will associate with learning, helpfulness, and clarity. At low levels of learning, helpfulness and clarity, easiness may be either low or high. On Figure 1, low levels of learning can have either high or low easiness, as indicated by the fact that a horizontal line from low learning intersects the easiness line twice. This pattern reflects the fact that instructors who do not promote learning would tend to be either too easy or too difficult. Similarly, with high levels of learning, easiness would tend to have middle values and a smaller variation. This pattern reflects the fact that instructors who promote learning tend to be helpful and clear, but neither too easy nor too difficult.

A similar pattern should hold for the relationship between easiness and helpfulness and clarity. At low levels of clarity and helpfulness, easiness may be either high or low, suggesting high variability. At high levels of clarity and helpfulness, easiness is expected to have a moderate value and a low variability.

H2. The level of variability in easiness, measured by the squared difference from its mean, will be inversely related to clarity and helpfulness.

We also expect that the measures for clarity, helpfulness, and easiness are not measures of the same underlying construct. Previous analysis of student learning shows that helpfulness and clarity associate with learning, while difficulty does not. Therefore, we expect that clarity and helpfulness will be associated with the latent factor of learning. Easiness will not, and should therefore represent another factor.

H3. Easiness, clarity, and helpfulness will not be unidimensional

Analysis and Results

Our unit of analysis for this research was the faculty member. From the RateMyProfessor.com website, we randomly selected 399 faculty from the 4077 educational institutions listed. Our variables are listed in Table I. Each variable was rated from 1 (low) to 5 (high), except for 'hotness' which was a running total of binary ratings (-1='not hot' or +1='hot'). For variability in easiness, we subtracted the value from the mean of easiness for the sample, then squared it. We divided the 'hotness' rating by the number of ratings for each faculty member to create a 'hotness ratio' for each faculty member. Perceived 'hotness' may be a measure of a type of universal appeal, and has been shown to affect overall ratings (Felton et al., 2004).

Hypothesis 1 is that helpfulness and clarity will be positively correlated. Average helpfulness and average clarity are strongly correlated (0.87), which is significant at the .000 level. Other analyses, including regression of average helpfulness on average clarity with control variables (not shown), give similar results. These findings support hypothesis 1 and confirm similar results originally reported in Otto et al. (2005).

Hypothesis 2 is that the variability in easiness will be inversely related to clarity and helpfulness. We tested this hypothesis with multiple regression, using clarity and helpfulness as the dependent variables and variability in easiness. Our resultant regression models have significant explanatory power as indicated by the R squared and F statistics. Examination of the residuals showed no evidence of non-spherical disturbances. The variance inflation factors for each variable were under 1.3, indicating that the results are not affected by multicollinearity.

Table 1 shows that average clarity and average helpfulness associated negatively with the variability in easiness, as hypothesized. There was also a negative correlation between variability in easiness and clarity/ helpfulness. These results support hypothesis 2.

Hypotheses 3 was tested with factor analysis using the Scree test (Cattell, 1978; Reise et al., 2000). Our analysis showed unidimensionality, which does not support Hypothesis 3.

Discussion and Conclusion

The analysis finds statistical support for hypotheses 1, and 2. The variables clarity, helpfulness, and easiness demonstrate patterns of association that are consistent with the assumption of learning. Not consistent with this assumption is the analysis indicating that clarity, helpfulness, and easiness are unidimensional. Past research has indicated that course difficulty, or workload, does not have an association with student learning (March & Roche, 1997; 2000; McKeachie, 1997).

Our analysis of online ratings from ratemyprofessor.com showed similarity to what we expected if the ratings were valid measures of student learning. Our analysis of a random sample of 399 ratings demonstrated that students' ratings of instructor clarity and helpfulness were strongly correlated. In addition, we found that the variability in easiness was inversely associated with clarity and helpfulness. These findings were consistent with our expectations under the validity assumption.

We did not expect to find that the ratings for clarity and helpfulness to be statistically unidimensional with instructor ratings for easiness. This finding is inconsistent with previous studies that show that course difficulty or workload is not statistically associated with student learning.

Overall, our results raise possible implications for instructor evaluation and perhaps promotion and tenure policy. First, online ratings may not be biased as a measure for student learning. While the ratings suffer from selection bias and sample representation, student responses may nonetheless reflect honest and true assessments of instructors. Our evidence is weak in this regard, but surprising given the high potential for bias. To the extent that online ratings can be demonstrated as unbiased, it may be appropriate to consider using this information to supplement decisions with respect to hiring and promotion, tenure, and merit decisions. Of course, such a change would have significant implications for university employment practices and policies.

References are available upon request

Table I: Regression Analysis Results

Variable	Dependent Variable	
	Unstandardized Coefficients	
	Average Clarity (Standard Error)	Average Helpfulness (Standard Error)
(Constant)	3.86** (.14)	3.92** (.14)
Variability in Easiness	-.20** (.05)	-.22** (.05)
Female Gender	-.21 (.13)	-.20 (.13)
Carnegie Classification - Master's	-.02 (.21)	-.02 (.22)
Carnegie Classification - Baccalaureate-General	-.21 (.21)	-.04 (.21)
Carnegie Classification - Associate's Colleges	.24 (.14)	.24 (.14)
Academic Area - English	-.28 (.20)	-.07 (.20)
Academic Area - Mathematics	-.70** (.25)	-.53* (.25)
Academic Area - Business	.25 (.25)	.13 (.26)
Academic Area - Computer Science	-.20 (.28)	-.21 (.29)
Hotness Ratio	1.06** (.18)	1.00** (.18)
Multiple Ratings	-.11 (.13)	-.05 (.13)
Adjusted R Square	0.145	0.131
F Statistic	7.14**	6.45**
N	399	399

* $p < .05$, ** $p < .01$

Figure 1. Relationships between Learning and Clarity, Helpfulness, and Easiness

