A COMPARISON OF FREE-FORM STUDENT EVALUATIONS ON RateMyProfessors.com AND A UNIVERSITY-BASED SYSTEM

Jesse M. Heines
Dept. of Computer Science
University of Massachusetts Lowell
Lowell, MA 01854 USA
1-978-934-3634

heines@cs.uml.edu

ABSTRACT
Student evaluations of teaching (SETs) have been around for decades. The debate on their validity and usefulness has raged just as long, with respected researchers on both sides of the table. The author is less concerned with numerical averages computed from SETs than with the substance of students' free-form responses to open-ended questions. This paper analyzes free-form responses posted on the popular website RateMyProfessors.com and compares them to analogous responses on a university web-based course evaluation system designed by the author. It examines differences between the two and the types of information that can be gleaned from such sources to improving teaching.

Categories and Subject Descriptors
J.1 [**Computer Applications**]: Administrative Data Processing – *education.*

General Terms
Management, Measurement, Standardization.

Keywords
Student Evaluations of Teaching, Course Evaluations, Web-based Course Surveys.

1. COAXING WATER FROM A STONE
Olivares [12] notes that "the use of student evaluations of teachers (SETs) in the U.S. is pervasive." Referencing Seldin [15], Olivares reports that "SET use in American institutions of higher education rose from 29% in 1973 to 86% in 1993." Despite their widespread use, Olivares [12] makes convincing arguments that these instruments are invalid measures of teacher effectiveness and that, at the very least, interpretation of their results must be tempered by a thorough understanding of students' interest in academic work, their reasons for taking a class, and their motivation for attending college in the first place [13].

SETs generally exist in the form of surveys that ask students to indicate whether the agree or disagree with statements such as

- The professor makes the course objectives clear.
- The professor presents material clearly.
- The professor is prepared for class.
- The professor comes to class on time.
- The professor encourages questions.
- The professor uses class time well.
- The professor is available during office hours.
- The professor grades exams and papers fairly.

d'Apollonia and Abrami [3] note that such items "reflect the characteristics that experts believe (a) can be judged accurately by students and (b) are important to teaching." The shortcoming, they continue, is that they do not fully assess instructional effectiveness, which "researchers define ... from a number of different perspectives." Marsh and Roche [11] agree, stating: "SETs are difficult to validate because no single criterion of effective teaching is sufficient."

On the other side of the table, numerous researchers believe that SETs correlate well with teaching effectiveness (see Felder [5] for a brief summary of these arguments with numerous citations). Those in this camp remain hopeful that SETs can be used to improve instruction, and Felder [6] has documented techniques for structuring SETs that he has found to help yield useful results.

Many professors believe that SETs are nothing more than a popularity contest, or simply a reflection of how "easy" a course is. Greenwald [7] quotes Snyder and Clair [16] as saying that "present evidence ... supports the notion that a teacher can get a 'good' rating simply by assigning 'good' grades." Greenwald goes on to quote Worthington and Wong [17]: "It is clear that ... an instructor [who] inflates grades ... will be much more likely to receive positive evaluations." Given this situation, it is not surprising that Olivares [12] argues: "there is ... little if any direct evidence to suggest that the widespread use of teacher ratings has resulted in more effective teachers or more learned students."

Regardless of whether one embraces or abhors SETs, the fact remains that many of us are contractually required to administer them. In addition, SETs are the *de facto* tool used to address the assessment requirement in the first criterion for computer science

program accreditation by the Accreditation Board for Engineering and Technology (ABET) [1]:

> The program has documented, measurable objectives, including expected outcomes for graduates. The program regularly assesses its progress against its objectives and uses the results of the assessments to identify program improvements and to modify the program's objectives.

Both of these situations exist at the author's and many other institutions, thus making the use of SETs quite well entrenched. The trick, therefore, is to make them useful for improving instruction.

The author is less concerned with how students rate him on the above statements than with whether students feel that they actually *learn* anything in his classes. When reviewing a draft of this paper, Olivares [14] commented: "[these] are not 'bad/poor' statements; rather, responses to these statements have not been shown to be related to learning." More importantly, responses to such statements provide no information on how teaching and learning can be *improved*. Getting that information, if it can be gotten at all, is akin to coaxing water from a stone. It must be carefully culled from students' free-form responses to open-ended questions. It is these responses that the author has found most valuable in his own learning about what all the way back in 1916 Dewey [4] referred to as "the art of instruction."

2. SOLICITING FREE-FORM RESPONSES

There are four problems with handwritten free-form responses on traditional paper-based course evaluation systems:

- They are often illegible.
- They are often unintelligible.
- They are seldom of substantial length.
- It is relatively easy to identify their authors from the handwriting, especially in small classes.

Three of these problems can be quite well addressed by changing the medium by which free-form responses are submitted. web-based systems require responses to be typed, virtually eliminating the first problem. One would think that this would also help the second problem, but the author's observation is that there is no improvement at all. One can't make a student write a long response if he doesn't want to, but at least Web forms can allow responses of virtually unlimited length if allowed to. Likewise, certain writing characteristics will always give a student's identity away to an astute observer, but at least identification by handwriting is eliminated.

The author developed a simple web-based system to solicit both discrete and free-form student evaluations in his own classes in the spring 2002 semester. When the system's existence became known by university administrators, they requested the author to expand it first for use by the author's department, then his college, and finally his entire university. Heines and Martin [9] provide extensive details of the system (including reactions to its use from the faculty union), but Figure 1 at least gives a feel for what it looks like to a student. (Note that this is a composite figure, not a straight screen capture.) The figure shows three of the six types of questions allowed by the system: agree-disagree, semantic differential, and free-form. Three others are also allowed: rating scale (1-5), multiple choice, and true-false (or yes-no) questions.

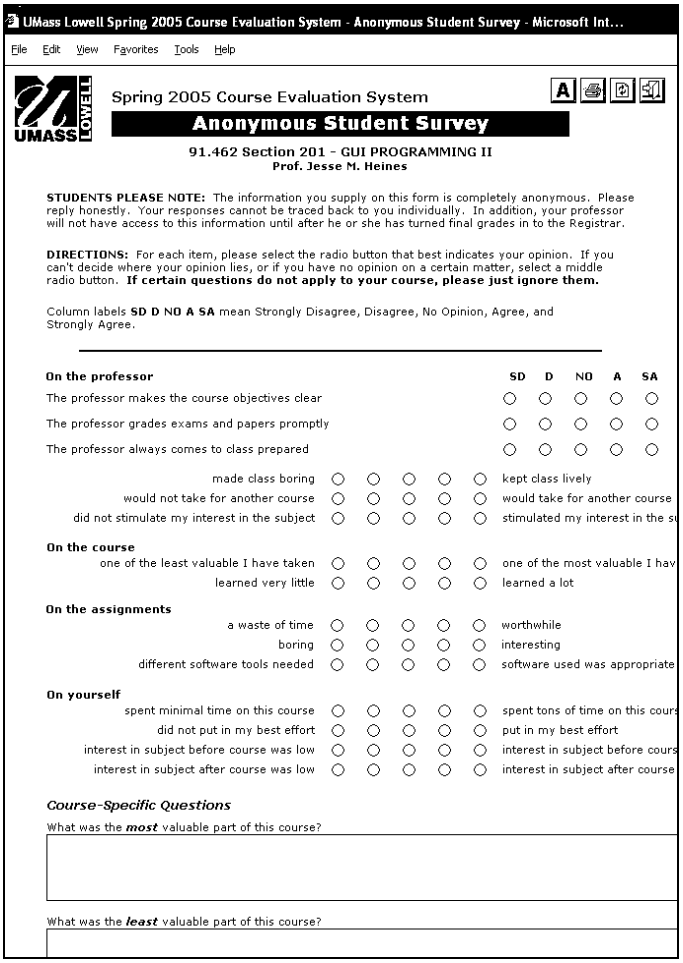

Figure 1. Representative student view of an evaluation form on the author's web-based system.
***Note:* This is a composite figure, not a straight screen capture.**

Comparing this form to the RateMyProfessors.com (RMP.com) form shown in Figure 2, two important differences are immediately apparent.

- RMP.com provides a fixed set of discrete questions, while the university system allows an unlimited number of such questions of six different types. Professors can even add their own questions specific to their courses.
- RMP.com limits its single free-form response to 350 characters, while the university system allows multiple free-form responses of virtually unlimited length.

Both systems can be used by students anytime, anywhere. RMP.com does not appear to have any provision for preventing students from "stuffing the ballot box," because one does not have to log in to submit a form. The university system, on the other hand, requires login to ensure that a single student can submit only one form per course while still maintaining anonymity [9].

Another major difference is that students are of course encouraged to use the university system, while use of RMP.com isn't "sold" at all (by the faculty). In some cases, classes are even dismissed early to provide students with time to go to the

computer labs and complete the evaluation forms. There are, therefore, major differences in the responding populations, with a relatively tiny percentage of students responding via RMP.com compared to the percentage responding via the university system.

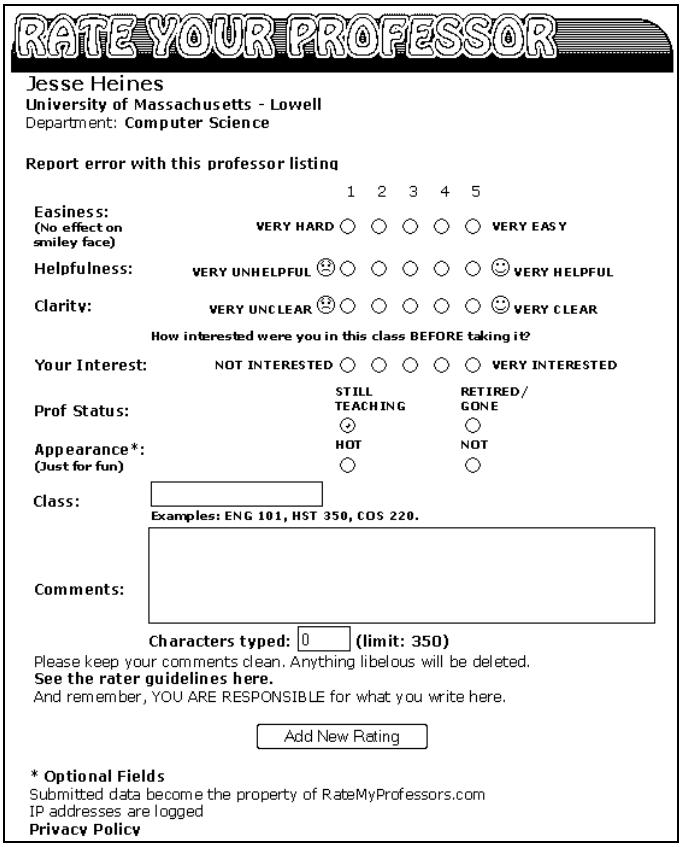

Figure 2. Student view of the evaluation form on RateMyProfessors.com.

3. RESPONSE RESULTS

3.1 Finding a Representative Sample

Evaluation data on all university systems the author is aware of is private, even on systems that are run by students. Therefore, it is nearly impossible to study a representative sample of evaluations. Furthermore, even if one had access to the data, it would be difficult to define a representative sample because, as discussed at length above, there is no clear-cut definition of quality teaching.

The author is therefore forced to use student responses on his own evaluations as a far-from-perfect sample. These responses, as one would imagine, are both positive and negative. Since the negative ones can actually be pretty hurtful, the author sought a way to compare his results and ratings to those of professors who were considered truly outstanding. Again he hit the same roadblocks: the lack of a quality teaching standard and access to data. But here RMP.com actually provided a solution!

Although no one has yet defined what constitutes quality teaching, Baylor University has established what can only be called the Nobel Prize of Teaching: the Robert Foster Cherry Award for Great Teaching. This award is "designed to honor great teachers, to stimulate discussion in the academy about the value of teaching, and to encourage departments and institutions to value their own great teachers" [2]. The prize is worth an unheard of sum for teaching awards: $215,000 to the recipient and another $35,000 to the winner's department! Nominees go through a rigorous evaluation process that includes interviews and presentations at Baylor, so it seems safe to assume that they must be pretty good teachers.

There have been 21 recipients of the Robert Foster Cherry Award since it was first awarded in 1991, and the website identifies 3 additional outstanding teachers as finalists for this year's award. The author was able to find public data on 11 of these 24 on RMP.com. Table 1 shows how the author stacked up against these exceptional teachers. (Note that in some years there were multiple recipients.)

Table 1. RateMyProfessors.com ratings of recipients of and recent finalists for the Robert Foster Cherry Award for Great Teaching and the author.

Professor	N	Ease	Help	Clar	OQ	Rank
2006 Finalist	5	3.0	4.8	4.8	4.8	1
2001 Winner	9	2.6	4.7	4.6	4.6	2
1992 Winner	42	4.4	4.6	4.4	4.5	3
2003 Winner	12	2.8	4.2	4.3	4.2	4
1998 Winner	4	3.8	4.5	4.0	4.2	4
2006 Finalist	14	3.3	4.6	3.7	4.1	6
The Author	16	3.7	3.9	4.2	4.1	6
2006 Finalist	31	2.1	3.7	3.9	3.8	8
1994 Winner	12	2.9	3.7	3.9	3.8	8
1995 Winner	17	3.8	3.8	3.9	3.8	8
1998 Winner	3	2.0	3.0	3.3	3.2	11
1992 Winner	9	2.9	2.1	4.0	3.1	12
Means	**15**	**3.1**	**4.0**	**4.1**	**4.0**	

Legend *(quoted text is from the RMP website):*

- **N** = number of ratings posted on RMP
- **Ease** = "Average Easiness" = "How easy are the classes that this professor teaches? Is it possible to get an A without too much work?"
- **Help** = "Average Helpfulness" = "Is the professor approachable and nice? Is the professor rude, arrogant, or just plain mean? Is the professor willing to help you after class?"
- **Clar** = "Average Clarity" = "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?"
- **OQ** = "Overall Quality" = average of all Helpfulness and Clarity ratings (not an average of the two averages, thus the apparent rounding errors are not really errors).
- **Rank** = rank in this table.

Interestingly, *Easiness* is not included in a professor's "Overall Quality" rating, since the creators of RMP.com claim that "an Easiness of 5 may actually mean the teacher is TOO easy" [*emphasis in the original*].

Given this data, the author feels a little more comfortable about revealing his negative ratings! And given the prior discussion about the correlation between easy grading and students' perception of overall quality, it is interesting to note that the Ease and OQ columns exhibit a Pearson correlation of 0.39. While not

significant due to the small sample size (n=12, p=0.325, Spearman's *rho*=0.307), the value of this correlation is consistent with that of analogous correlations reported in the literature [8].

3.2 Characteristics of Free-Form Responses on RateMyProfessors.com

RMP.com's single free-form question tends to generate complaints of little value to a professor who wants to improve. Of 135 free-form responses for the 11 people listed in Table 1, the author would categorize 96 of them as positive, 18 as neutral, and 21 as negative. While 16 of the 21 negative responses included some reason *why* they were negative, these were mostly personal, relating to the personality of the professor rather than material covered in the course. Here are a few examples of complete responses (with grammar and spelling corrected) for some of the people listed in Table 1:

- The WORST professor ever. Poor paper feedback, makes you feel like a failure. PLEASE!!! RE-THINK taking any of his classes.
- Too moody. Is unnecessarily rude and insulting when responding to emails from students.
- My worst experience thus far.
- He is by far the worst teacher. Class is so simple, and he asks really weird questions that basically asks you to just give a summary of the book. He tries too hard to act cool. All he does is flirt with girls to pass the time. It can be really creepy.

Of course, the professors who got these comments got positives ones as well. For example, the same professor about whom the last comment above was made also received:

- THE BEST TEACHER AT [*college name removed by the author*]!!!!!!!!!!!!! Take everything you can with him. He is wonderful, helpful, clear, organized, intelligent, passionate, funny, and so sweet! TAKE HIM NOW! [*All exclamation points are in the original.*]

So there's obviously a difference in opinion here, or perhaps the professor posted this message about himself! Actually, it is important to point out that that is indeed possible on RMP.com!

3.3 Characteristics of Free-Form Responses on the University System

As noted earlier, free-form responses on RMP.com are limited to 350 characters. In almost all cases, therefore, free-form responses on the university system were longer, more substantive, more specific to particular class activities, and included both positive and negative comments, making them more constructive. The nine free-form evaluations from one graduate class averaged 95 words each, and one was 320 words long! Thus, it is clear that students provided much longer responses on the university system than on RMP.com.

Free-form responses on the university system's free-form questions were also more focused, because the author was able to ask targeted questions:

- What was the *most* valuable part of this course?
- What was the *least* valuable part of this course?
- What topics should we have spent *more* time on?
- What topics should we have spent *less* time on?
- Please write any additional comments or suggestions you have (either positive or negative) that you think would help improve this class. Thank you.

As one would expect, these questions yielded many more specific suggestions for improvement rather than just complaints about things the students didn't like. For example, in response to the question about what we should have spent more time on, students in the author's senior undergraduate capstone project wrote:

- The design aspect of our project, seeing that it was a capstone project of considerable size. [More] software design strategies would have made the process easier.
- Time was spent right. Interested in more of the web stuff
- More complex topics in Java, e.g., creating custom controls.

Graduate students are even more explicit and exacting:

- Class material is excellent, and the professor certainly teaches it well. However, he can be abrasive at times, which makes interaction difficult. Still, he is very fair.
- I never had any bad feelings about any interaction with any professor during my program in this department. But 1 or 2 times, I felt [that this professor] was little bit harsh in his reply to me. It was just a request for an extension of the assignment, so [he] could have [replied with a] simple answer YES or NO. Professor is a very good person, but talking to some other students I feel sometimes he loses temper on small things.

As on RMP.com, the author could provide counterexamples in which students said he was "easy to approach," but that is not the point. These responses are precise and thoughtful. They not only provide a clear indication of a problem, but they zero in on the aspects of the author's interactions with students that he needs to work on.

Two final examples of free-form responses relate to a web-based classroom response system of the author's own design called the "Engaged Classroom" that he used in class. Students wrote:

- I feel the Engaged Classroom Project has some nice aspects to it, but it really needs to be fine tuned to take up less class time, because I couldn't see this being used in a 1 hour class. Having the laptops there has serious potential for some hands on coding. I find that if I type the code I'll remember it much more than if I'm shown it... this may not be entirely appropriate for some of the material taught in the class since some things are accentuating the concepts rather than the literal code (which I think is great), but I think it could and should be used when appropriate.
- The Engaged Classroom is a good start but needs to be improved. Pre-set questions may help and MORE questions may help as well. Maybe you can incorporate mini-non-graded quizzes as an interesting part so that people are more challenged but aren't pressured if they make mistakes. The other part that was nice was being able to visit the course website while in class to see the class notes. It is definitely a great venture and there are so many more things that can and should be added.

It is clear from these examples that responses on the university system are not only longer, but are far more substantive and constructive than those on RMP.com. Even when not focused by course-specific questions, that is, when simply asked, "Please write any additional comments or suggestions you have (either positive or negative) that you think would help improve this class," students tend to write more thoughtful responses and offer more specific suggestions for improvement.

4. IMPROVING INSTRUCTION

The author is surely not alone in the mixed responses he receives from students on SETs, and we can all take some solace in seeing that our ratings on RMP.com compare OK with those of winners of the Robert Foster Cherry Award for Great Teaching. The main issue, however, is that the comments posted on RMP.com provide little substance that one can use to improve teaching.

By contrast, responses on the university system are not only longer, more explicit, and more substantive, they more often provide concrete suggestions for improving both courses and instructors. They are representative of a larger cross-section of the student population, and one can have far more confidence that no student has submitted multiple responses for a single course.

By refining the question strategy to provide even more specific student responses, the author hopes that the system can be further improved. The goal remains to make it a system that improves what Dewey [4] has called "the art of instruction" and that fosters what Olivares [13] would identify as "more effective teachers [and] more learned students."

5. ACKNOWLEDGMENTS

Recent graduate students Saurabh Singh and Shilpa Hegde provided invaluable assistance in building the web-based course evaluation system designed by the author. Graduate student Martin Schedlbauer assisted with statistical data analysis. Fred Martin of the University of Massachusetts Lowell and Orlando Olivares of Bridgewater State University reviewed drafts of this paper and provided detailed suggestions for revision. Olivares also helped shape the author's thinking through discussions of these issues and sharing a preprint of his upcoming publication [13].

6. REFERENCES

[1] ABET Computing Accreditation Commission. *Criteria for Accrediting Computing Programs*. ABET, Inc., Baltimore, MD, 2004, 1

[2] Baylor University. *Robert Foster Cherry Award for Great Teaching*. http://www.baylor.edu/cherry_awards/, 2005.

[3] d'Apollonia, S., and Abrami, P. Navigating student ratings of instruction. *American Psychologist, 52,* 11 (Nov. 1997), 1198-1208.

[4] Dewey, J. *Democracy and Education*. The Macmillan Company, 1916.

[5] Felder, R. What do they know anyway? *Chemical Engineering Education, 26*, 3 (Summer 1992), 134-135.

[6] Felder, R. What do they know anyway? 2. Making evaluations effective. *Chemical Engineering Education, 27*, 1 (Winter 1993), 28-29.

[7] Greenwald, A. Validity concerns and usefulness of student ratings of instruction. *American Psychologist, 52*, 11 (Nov. 1997), 1182-1186.

[8] Greenwald, A., and Gillmore,G. Grading Leniency Is a Removable Contaminant of Student Ratings. *American Psychologist, 52*, 11 (Nov. 1997), 1209-1217.

[9] Heines, J., and Martin, D. Development and deployment of a web-based course evaluation system: trying to satisfy the faculty, the students, the administration, and the union. *Proceedings of the First International Conference on Web Information Systems & Technologies (WebIST)* (Miami, FL, May 26-28, 2005). Inst. for Systems & Technologies of Information, Control & Communication, Setúbal, Portugal, 441-448.

[10] Hernández, L.O., Wetherby, K, and Pegah, M. Dancing with the devil: faculty assessment process transformed with web technology. *Proceedings of the 32nd Annual ACM SIG UCCS Conference on User Services* (Baltimore, MD, October 10-13, 2004), 60-64.

[11] Marsh, H., and Roche, L. Making students' evaluations of teaching effectiveness effective. *American Psychologist, 52*, 11 (Nov. 1997), 1187-1197.

[12] Olivares, O. Student evaluations of teachers: intended and unintended social consequences. *Journal of Excellence in College Teaching, 15*, 3 (2004), 105-127.

[13] Olivares, O. Student evaluations of teachers as goal-directed behavior in a sociocultural context. *Journal of Excellence in College Teaching*, in press.

[14] Olivares, O. Personal correspondence. August 30, 2005.

[15] Seldin, P. How colleges evaluate professors: 1983 versus 1993. *American Association for Higher Education Bulletin, 46*, 2 (October 1993), 6-8, 12.

[16] Synder, C., and Clair, M. Effects of expected and obtained grades on teacher evaluation and attribution of performance. *Journal of Educational Psychology, 68* (1976), 75-82.

[17] Worthington, A., and Wong, P. Effects of earned and assigned grades on student evaluations of an instructor. *Journal of Educational Psychology, 71* (1979), 764-775.