JOURNAL OF CHEMICAL EDUCATION

Article

pubs.acs.org/jchemeduc

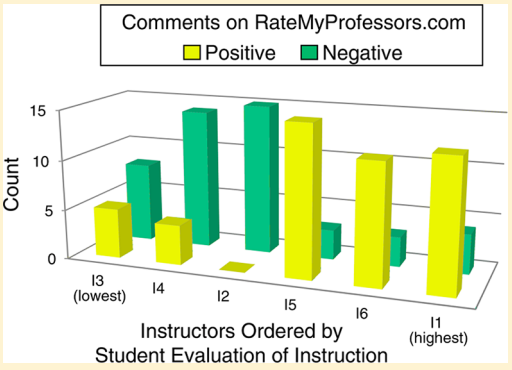
Use of RateMyProfessors.com Data as a Supplemental Tool for the Assessment of General Chemistry Instruction

Adam Bergin, Kevan Sharp, Todd A. Gatlin, Adrian Villalta-Cerdas, Austin Gower, and Santiago Sandi-Urena*

Department of Chemistry, University of South Florida, Tampa, Florida 33620, United States

Supporting Information

ABSTRACT: Commercial online instructor evaluations have gained traction in influencing students' decisions on professor and course selections at universities. RateMyProfessors.com (RMP) is the most popular of such evaluation tools and houses a wealth of information from the students' viewpoint. The purpose of this study was to determine whether RMP data could be used to analyze and inform general chemistry instruction at a particular institution. The entire RMP database for the general chemistry program was sampled to produce a subset of 60 random RMP entries from six instructors. Each entry was composed of ratings in several areas and open-ended comments. The quantitative RMP data were consistent with measures from the institutional Student Evaluation of Instruction forms corresponding to the same instructors. In addition, a survey investigating RMP use patterns demonstrated that general chemistry students who contributed to RMP were not significantly different from the rest of the cohort across seven academic and demographic comparison criteria. The RMP qualitative information was analyzed using an inductive approach from which seven categories emerged as important to students' learning environment. This analytical model allows for categorizing students' statements in a systematic and meaningful manner to extract valuable supplemental information usable for program evaluation.



KEYWORDS: *General Public, First-Year Undergraduate/General, Chemical Education Research, Internet/Web-Based Learning, Professional Development, Undergraduate Research, Testing/Assessment*

FEATURE: Chemical Education Research

INTRODUCTION

In the Spring of 2010, we started the CHEmical Education Research–Undergraduate Program at USF, CHEER–UP@USF!, to actively encourage creative and free thinking in undergraduate students while stressing the importance of academic and research rigor. The program strives to develop a research agenda constructed from undergraduates' perspectives and informed by their experiences as key participants in their own education. In turn, findings may inform and impact undergraduate chemistry instruction. The exploration of how students use online faculty rating sites and their perceived usefulness were among the ideas generated in consultation with the first group of undergraduates. This study is a product of that interest.

With more than 13 million entries for more than 1.7 million instructors from over 7500 institutions in the United States, Canada, and the U.K., RateMyProfessors.com (RMP) claims to be the highest traffic site of its kind.[1] RMP visitors may rate their instructors using a five-point scale on three items: easiness, helpfulness, and clarity. The average of the latter two is called the overall quality. Two additional items complete an entry: "Interest level prior to attending class"; and quality of the "Textbook used". Additionally, visitors may leave open-ended comments that are consistent with site guidelines; otherwise, the RMP site moderation team may remove them.[1] More detailed descriptions of RMP can be found elsewhere.[2,3]

Despite the controversy that surrounds its validity, RMP has gained traction in influencing students' decisions and the way that students, instructors, and administrators alike think of higher education and assessment of instruction. For instance, in its annual ranking of America's Top Colleges, *Forbes Magazine* uses RMP as one of three metrics for student satisfaction and weights it at 17.5% of a school's rating.[4] *Forbes* argues that a growing body of research suggests that RMP should be taken seriously, given its consistency with measures such as institutional evaluations of instruction.[5−7] Recently, *The Princeton Review* partnered with RMP to identify the "best 300 professors" in the United States.[8] Reactions to this report may be varied but certainly it has caught the attention of those involved with higher education. Nevertheless, legitimate criticism is not absent from literature. Davidson and colleagues[9] assert that "the information provided by the RMP Web site is not valid" and allude to lack of external validity as a "huge problem". In our experience, instructors are often quick to

Published: December 31, 2012

ACS Publications © 2012 American Chemical Society and Division of Chemical Education, Inc.

dismiss RMP based on personal assumptions that contributors simply want to reward easiness. However, researchers have suggested that these assumptions are not always well founded.[2] Instructors also often argue RMP is biased because its contributors are self-selected and their contributions are not monitored.[9] Nonetheless, research has not identified significant academic or demographic differences between student contributors and noncontributors.[2]

Our intention is neither reconciling these two opposing stances nor providing supporting evidence for either stance. Our views resonate with Cashin's suggestion to consider student ratings as data and not as evaluations.[10] Thus the responsibility rests upon the evaluators to incorporate multiple sources of data to advance a proper judgment of instruction effectiveness. Our approach is somewhat different from what has characterized previous scholarly scrutiny of RMP. We do not seek to probe the validity of the site or judge the ratings for individual instructors in an attempt to generalize findings or establish comparisons among instructors or institutions. Instead we intend to determine the appropriateness of using information from a set of RMP data as a complementary source of evidence when conducting assessment of instruction at the program level. This evidence is essentially different from other sources such as institutional evaluations or direct observations of instruction because it does not stem from an institutional initiative. Our focus on a program allows aggregation of data across multiple instructors and semesters that introduce a level of variation that may reduce the effect of bias. Although the initial unit of data gathering is at the instructor level, this aggregation provides a snapshot of the program that evaluators can use in their analysis. Furthermore, most of the criticism about RMP has centered on the numerical evaluations, the ratings,[7] whereas our ultimate goal is to use the information in the open-ended comments. The purpose of this study was to determine whether RMP data could be used to analyze and inform general chemistry instruction at a particular institution. To accomplish this goal, we addressed three guiding questions:

1. Are students who contribute to RMP different from the general chemistry cohort, and, if so, in what ways?
2. Are RMP data consistent with institutional student evaluations of instruction?
3. Can we coherently categorize the qualitative entries in RMP?

We viewed the first two questions as a means of gathering evidence for validity of the data set. The third question was aimed at developing a procedure to systematically analyze the data set and facilitate extraction of meaningful information to aid in evaluating the program. Others may then use a similar approach to evaluate the usefulness of their RMP data as a component of their program assessment.

DATA SOURCES

RMP Patterns of Use Survey

This study took place at a large, research-intensive, southeastern U.S. university. The initial question was whether or not students who use the site to post comments differ systematically from the rest of the cohort. For this comparison, we used six criteria gathered as part of a longer survey that investigates RMP patterns of use: gender; major; year status; grade-point average (GPA); course load; and previous chemistry grade. Convenience sampling, a nonprobability technique based on accessibility of participants, furnished information from 276 General Chemistry 2 students (49% of initial enrollment, Fall of 2011). A subset of 132 participants completed the survey during lecture and shortly after the first of three midterm exams. The remaining 144 were surveyed in the laboratory during the last week of classes. No students had dropped out of the laboratory sections chosen. We used hard copies of the anonymous survey (see the Supporting Information). Data collection followed IRB guidelines; no compensation was offered for participation.

RMP Qualitative Data

Although the information mined is in the public domain, we decided to maintain instructors' anonymity throughout the study. Sampling considered general chemistry instructors who were active in the program during 2008, 2009, and 2010 and who had received at least ten comments. To provide another level of anonymity, we randomly selected a subset of six instructors from the eight that met the initial criteria. A researcher assigned codes (I1−I6), and neutralized gender designations (e.g., possessive pronouns) by changing all to feminine forms. In addition, references that could potentially lead to the identification of instructors were removed without altering the nature of the comments. The second sampling stage involved random number generation to select 10 comments per instructor, thereby producing the final sample: 60 RMP entries.

RMP Quantitative Data

RMP uses a five-point scale (1 is lowest, 5 is highest) for its ratings, and reports historical aggregates for instructors; that is, the single values for easiness, helpfulness, clarity, and quality condense ratings received throughout instructors' entire RMP history. Because individual ratings for each entry are available, we calculated averages using only the 10 comments sampled for each instructor rather than using the historical average. Our rationale was that this gives a more accurate picture of the evaluation of instructors within the time window of the study. Comparison with the historical aggregates showed they were not different.

Student Evaluation of Instruction Data

Numerical data for the student evaluation of instruction (SEI) are publicly available in the institution's Web site. Information is aggregated and posted by instructor, course, and semester. Hard copies of the SEI forms are administered toward the end of the term at instructors' discretion. The evaluation consists of eight items (E1−E8) and uses a five-point scale (see the Supporting Information). Additionally, students may expand their comments in writing on the back of the form; however, no written instructions mention this. Only numerical data are processed and saved by the institution. Forms are returned to instructors and written comments are only available to them. We collected the numerical SEI online data for the instructors and time period corresponding to the RMP data set. Because the SEI ratings were reported by course and semester for the individual instructors, we aggregated results over the time period of the study to produce unique weighted mean values for each instructor for each SEI item. The purpose of this aggregation is to make a comparison with RMP numerical data that we aggregated for the same time period.

Table 1. Comparison of Contributors and Noncontributors, RateMyProfessors.com Patterns of Use Survey

Criteria	Values (N)	Contributors, %	Noncontributors, %	χ^2 (p) Values
Gender	Male (111)	44	41	0.15 (0.7)
	Female (159)	56	59	
Major	Biology or related (87)	29	33	0.16 (1.0)
	Chemistry or related (29	12	10	
	Biomedical sciences (86)	32	32	
	Other or undecided (68)	27	25	
Status (Year in School)	Second year (148)	50	58	4.50 (0.1)
	Third year (82)	47	30	
	Fourth year (28)	3	12	
GPA (Four-Point Scale)	4.00−3.75 (29)	15	10	6.20 (0.2)
	3.74−3.50 (63)	37	22	
	3.49−3.25 (62)	24	23	
	3.24−3.00 (61)	12	24	
	2.99 or below (52)	12	21	
Course Load (Number of Credits)	1−10 (24)	6	9	0.45 (0.9)
	11−13 (95)	35	35	
	14−15 (113)	44	42	
	16−18 (38)	15	14	
Most Recent Chemistry grade	A+ to A− (76)	26	28	3.50 (0.2)
	B+ to B− (141)	65	50	
	C+ and below (53)	9	22	

STUDY DESIGN AND PROCEDURES

Comparison of RMP Contributors and Noncontributors

We used the RMP patterns of use survey to separate participants into those who left RMP ratings and comments, the contributors, and those who did not, the noncontributors. Subsequently, we used categorical statistics to compare the response distributions for these two groups along the six criteria introduced above. We discarded six surveys that were missing information or were completed inappropriately.

Text Analysis and Coding: Categorization and Analysis of the RMP Entries

Text analysis was conducted using the individual comments as unit of analysis. In our inductive approach, three researchers read the entries for one instructor separately and with the purpose of exploring the text and identifying and extracting thematic patterns.[11] This open coding was based on identifying significant statements or textual units of meaning that then produced categories.[12] This approach required individual researchers to work back and forth between their proposed categories and the original text. Subsequently, we compared and contrasted the categories extracted in a process of consensus meaning-making based on arguing for substantive significance.[11] The categories were defined and characterized to warrant clear and common understanding of their meaning. A new set of entries for another instructor was then analyzed and the process repeated. Although each time the up-to-then agreed upon categories served as an a priori analytic framework, we modified them when continuous comparison with the data suggested it was appropriate. New categories were added as they emerged. This analysis was completed for all six instructors after which we conducted an additional revision of the entire set of entries until a comprehensive set of categories was established.

Consistency between RMP Data and SEI data

Correlation analyses were used to compare numerical data from RMP (Overall Rating, Easiness, Helpfulness, and Clarity) with data from the SEI (items E1−E8). Our main interest was the correspondence between RMP Overall Rating and E8, Overall Assessment of Instructor. Subsequently, we assessed the consistency between the RMP comments—the qualitative component—and the SEI. To this end, we transformed RMP qualitative data into quantitative data by assigning valence to individual significant statements. We conducted a count per instructor for the text analysis categories that allowed valence, that is, positive or negative attributions were possible. This tally resulted in a frequency table with counts of positive and negative valence for each category per instructor. In principle, consistency would be supported by higher SEI overall rankings being associated with higher positive valence ratios (more positives than negatives).

RESULTS AND DISCUSSION

RMP Contributors and Noncontributors Comparison

We found that 12% of the survey respondents reported having made RMP contributions to rate their chemistry instructors; they are the contributors. Other authors have reported higher general posting rates;[2,7] however, their reports do not refer exclusively to posting behavior for specific science courses. The subset that visited the site but contributed no postings, 80%, is similar to values in the literature.[7] Only 5% knew about RMP but never visited, and 3% did not know about the site. We combined the latter three groups into a single one: non-contributors. The results in Table 1 suggest that students who contributed on RMP were not significantly different from the others. Whereas the RMP use pattern survey was applied in 2011, the entries sampled correspond to 2008−2010. However, there were no significant changes in the program or characteristics of the student body in the period 2008−2011. Therefore, we assumed that the contribution trends were not different at these two different times. A considerable 64% of the sample acknowledged they reviewed ratings for chemistry professors. We excluded first-year students (n = 8) from the computations because of the small count. Moreover, students taking General Chemistry 2 their first college semester are

dx.doi.org/10.1021/ed300277n | J. Chem. Educ. 2013, 90, 289−295

Table 2. RateMyProfessors.com Analysis Categories

	Categories	Subcategories	Example Statements
1	Nature of Course: Assessment and grading	Assessment	"She doesn't write the tests—the Chem. department does."
		Grading	"CHM II is really easy grading wise."
2	Nature of Subject: Inherent difficulty or ease of the subject		"Gen Chem I is hard in general."
3	Student Responsibility: Perceived expectations to succeed		"You just have to keep up, do hw [homework], get your clicker points, study for the tests."
4	Quality of Instructor: Holistic assessment of the professor		"She is a pretty great teacher for introducing chemistry." "Professor is entirely incapable of teaching undergraduates."
5	Instruction: Effectiveness, methods, and quality of instruction	Description	"She LITERALLY reads from the book every class."
		Quality	"She explains things in a way you can understand them." "Expect to force yourself to learn on your own as she doesn't help you learn the material here."
6	Suggestions: Forthright recommendations to students		"Get help as soon as you have problems and form a study group."
7	Instructor Traits: Personality, attitude, and interaction with students	Personality	"She is a very nice person."
		Affective element	"She will do all she can to help you succeed on her end." "She will yell at you for being late in front of the entire lecture hall."

atypical and have no prior college chemistry experience they can rate.

Categorization and Analysis of RMP Entries

Our evaluation of RMP postings produced seven categories, given in Table 2. We contend that as these categories emerged from unconstrained open-ended comments, they encompass the aspects students truly deemed relevant. A brief description follows:

1. Nature of the Course includes salient aspects that in students' views characterize the course. It was broken down into two subcategories: assessment and grading. In the subcategory Assessment, students inform about and comment on the evaluation procedures (e.g., homework, and whether or not attendance was mandatory). In the subcategory Grading, statements pertain to factors directly linked to grades (e.g., how tests or quizzes are graded).
2. Nature of the Subject describes comments that address the perceived inherent workload and intellectual demands students associate with chemistry as an object of study.
3. Student Responsibility includes statements that reflect perceived expectations of the student enrolled in chemistry to perform well. These expectations are not the restatement of the official expectations, but are grounded in students' experiences.
4. Quality of Instructor considers only statements that holistically and categorically assess the performance of the instructor and does not include references to personality and appearance or methods and didactics.
5. Instruction is a category collecting statements emphasizing the nature of the course instruction: effectiveness, methods, and quality. The first subcategory deals merely with descriptive aspects, Description, while the second subcategory addresses the quality of instruction, Quality.
6. Suggestions include entries that clearly have the purpose of influencing readers' decisions or behavior. Comments in this category share the uniqueness of being directed to the site's users—other students—and not about the instructor, instruction, or the course itself.
7. Instructor Traits, subcategorized as Personality and Affective Element, collect comments about the faculty members and their engagement with the course; it excludes any remarks related to the overall instructional efficacy. We identified a few instances that pertained to general qualities; however, most statements (72%) linked personal characteristics to the learning experience or focused on instructor−student interactions.

Textual analysis of the 60 entries produced 190 significant statements or units of meaning. This corresponds to a mean value of 3.2 by contributor with a mode of 3 (42% of the entries). Table 3 shows the distribution of these significant statements among the emerged categories. Statements that were unrelated to the recurrent themes and that appeared only in one or two entries were excluded from the analysis. This is the case, for example, of students responding to others' postings: "Instead of criticizing her teaching, you morons should learn how to THINK for yourselves, and not expect everything to be spoon-fed to you."

Two of the categories, Nature of Subject and Student Responsibilities, have relatively low frequency (Table 3). For a statement to fall under Nature of Subject we required an explicit reference to chemistry and not surprisingly, they were invariable in the tone of "hard" and "demanding". Different studies show that chemistry is perceived as abstract, difficult, and above all disconnected from relevant problems in daily life.[13] Knowing the impact of self-efficacy and predisposition on academic performance, evidence of a negative subject reputation should encourage instructors to address this aspect during instruction and promote a friendlier view of chemistry. We speculate that the low occurrence observed for this category may stem from an undistinguishing use of the terms course and subject by students. Even if only in a low frequency, the uncovering of this category stresses the need for specificity in the evaluation of instruction. From the students' perspective, framing the evaluation of a chemistry class may be very different from evaluating a course in a different subject. Although at this particular institution instructors have the option of amending the SEI, the general chemistry program uses it in its generic form therefore not accounting for differences in subjects. Similarly, Student Responsibility showed a relatively low frequency (∼5%). We consider this category to be intimately related to Suggestions; however the voices speaking through the comments in these dimensions are slightly different. Student Responsibility comments pose a more rhetorical or reflective stance while the language in Suggestions

dx.doi.org/10.1021/ed300277n | J. Chem. Educ. 2013, 90, 289−295

Table 3. Frequencies of RateMyProfessors.com Significant Statements by Category and Student Evaluation of Instruction Correspondence

Categories	Frequencies: Count (Total = 190)	Frequencies: Percentage	Associated SEI Items[a,b]	
Nature of Subject	5	2.6	—	None
Student Responsibility	9	4.7	—	None
Nature of Course	22	11.5	E1	Description of the course objectives and assignments
Suggestions	29	15.3	—	None
Quality of Instructor	30	15.8	E4	Availability to assist students in and out of class
			E7	Facilitation of learning
			E8	Overall assessment of instructor
Instructor Traits	43	22.6	E4	Availability to assist students in and out of class
			E5	Respect and concern for students
			E6	Stimulation of interest in the course
Instruction	52	27.4	E2	Communication of ideas and information
			E7	Facilitation of learning
			E8	Overall assessment of instructor

[a]SEI: Student Evaluation of Instruction. [b]E3, Expression of expectations for performance, did not correspond with any of the emergent categories.

is unequivocal: The commenter directly addresses other students. The Suggestions made by contributors may be very informative in the sense that they reveal the views students hold about instruction and instructor. For instance, from the following statements: "Go to class for clicker questions, attend small sessions" and "Do the homework, study the old exams, and you'll be ok", one may infer that in terms of learning, these students value more the supplemental activities than attending class, which does not necessarily come across as a learning opportunity but as a mere way to gather points.

The categories Nature of the Course and Suggestions ranked third and fourth lowest in frequency (Table 3). Overall, the first four categories in Table 3 are concerned with more global views of the learning experience, whereas the last three, Quality of Instructor, Instructor Traits, and Instruction directly address the instructor and instruction and account for two-thirds of the total comments. The prevalence of these categories may indicate students' legitimate interest in instruction. Subsequent comparison with the SEI showed that these three categories matched six of the eight items on the institutional evaluation (Table 3). One may argue that Quality of Instructor and Instructor Traits are two sides of the same coin and might be collapsed into a single "Instructor" category. Nevertheless, we desisted from this approach in view of the significance of a category specific to the global assessment of instructor quality. Furthermore, we found the Instructor Traits category to be loaded with language stressing the relevance of the affective dimension of learning environments. It is important to note, however, that it refers to the nature of the interactions between instructor and students and not to an emotional response to grades. Both positive and negative comments addressed aspects such as respectfulness, caring, helpfulness, and approachability and only in one entry did we find direct reference to grades: "[H]as no mercy at all when it comes to your grade...Also I was 5 points away from an A and she wouldn't help me out at all". These findings challenge the assumption that RMP contributors will use the site to retaliate after getting a bad grade or, at least, such behavior is not explicit. Contrary to what others may expect,[9] our sample was free of comments unrelated to the course or focused on personal aspects such as appearance. Our data do not show hostility or personal attacks of any kind; they were seemingly focused on quality of instructor and instruction and instructor–student interactions in compliance with site guidelines.[1] Kindred and Mohammed[14] arrived at similar findings in their content analysis of 1054 RMP entries.

RMP and SEI Data Comparison

Table 4 shows the trend for RMP Overall and Clarity ratings with increasing Quality of Instructor as measured using SEI E8 (Overall Assessment of Instructor). The Spearman's rank order correlation coefficients (r_s) for RMP Overall Rating and SEI E8, as well as for RMP Clarity rating and SEI E8, were 0.56 (p = 0.23). Although not significant statistically, this trend suggests that data from these two instruments are consistent in their ranking. All SEI items, E1–E8, were found to be highly correlated rendering any other RMP–SEI item comparison redundant. Inspection of Table 4 showed that I3 and I5 are slightly off in comparison with the other instructors. Program evaluators may consider further exploration of such outliers (for instance, effect of instructor experience or introduction of instructional innovations). Other authors have gathered evidence for this consistency using different samples.[7] A striking feature of Table 4 is the consistent grouping across measures of instructors into an "average-and-below end" (I3, I4, and I2) and an "above-average end" (I5, I6, and I1). Within the extremes, the values are for practical purposes the same yet they are quite distinct between the extremes. This trend is mirrored by the SEI average response rate (based on course matriculation) in which lower percentages were associated with lower SEI and RMP ratings (Table 4). Because attendance is encouraged by giving students "clicker participation" points, we believe it is unlikely that absenteeism alone would account for differences in response rates.

Table 4. Comparison of RateMyProfessors.com and Student Evaluation of Instruction Data by Instructor

Instructor Code	SEI E8 Rating[a,b]	SEI Response Rate, %[a]	RMP Overall Rating[c]	RMP Clarity Rating[c]
I3	2.6	43	3.0	2.8
I4	3.0	44	2.2	2.5
I2	3.1	54	2.0	2.2
I5	3.7	66	3.9	3.6
I6	4.4	69	3.8	3.6
I1	4.5	62	3.7	3.4

[a]SEI: Student Evaluation of Instruction. [b]E8 is the SEI item, Overall Assessment of Instructor; scale is 1–5, with 5 being the highest. [c]RMP: RateMyProfessors.com; scale is 1–5, with 5 being the highest.

Although we conducted these comparisons in an attempt to gather supplemental evidence to establish the validity of the RMP data set, we are aware of their limitations. Understandably, the power of this analysis is supplemental and we do

dx.doi.org/10.1021/ed300277n | J. Chem. Educ. 2013, 90, 289–295

not put it forth as an appropriate primary or stand-alone source of evidence. We were more focused on probing the trends observed within the group of six instructors. As we do not have evidence for the validity of the SEI itself, the implications of these patterns need to be interpreted with caution. Additionally, although we chose to use data collected during the same period of time for both RMP and SEI, we have no specific information about the samples surveyed by the instructors at the end of their courses. Contrary to what may be considered good practice[5,15] there was no formal, well-defined, and controlled protocol for the administration of the SEI. The wide range of response rate (22−87%) and the varied average values (Table 4) suggest the possibility of a bias effect. It is of utmost relevance to underline that our purpose was to gather multiple-source evidence for the validity of this set of data to judge their usefulness in assessing this general chemistry experience. Other authors have reported moderate to strong correlations between RMP and institutional evaluations.[7]

To complement the test of consistency between the numerical component of RMP and the SEI, we further probed their consistency but this time using the transformed qualitative RMP data. Data for four of the RPM model categories were not transformable: Nature of Subject, Student Responsibility, Nature of Course, and Suggestions. These categories did not lend themselves to positive or negative attributions. Similarly the subcategory Instruction: Description was not transformable. The remaining subcategory Instruction: Quality and the two categories Quality of Instructor and Instructor Traits were tallied according to their valence. In seven cases, assigning valence to the statements was not practical due to ambiguous comments such as "She can explain things quite well, but sometimes she can go off on tangents and it can confuse you." Figure 1 shows the total positive and total negative sum of

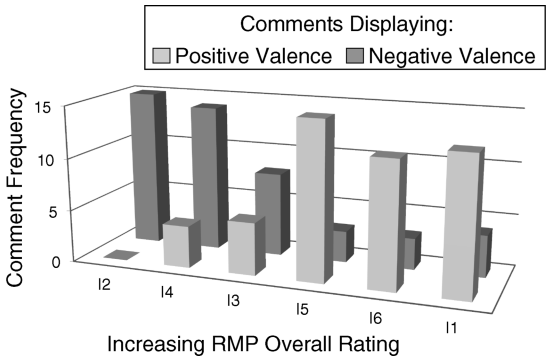


Figure 1. Comment valence by increasing RateMyProfessors.com (RMP) overall rating.

these three categories by increasing instructor RMP Overall Rating (I2 lowest, I1 highest). This figure illustrates the internal consistency of the quantitative and qualitative RMP data: positive comments are clearly more associated with higher RMP Overall Rating.

Figure 2 illustrates the total count of positive and negative valence for the three categories by increasing SEI E8 (Overall Assessment of Instructor). This analysis shows a simple trend in which instructors with higher SEI E8 ratings have a higher positive valence count in their respective RMP open-ended comments, while for those instructors with lower SEI E8 ratings, the negative valence comments outnumbered the positive ones.

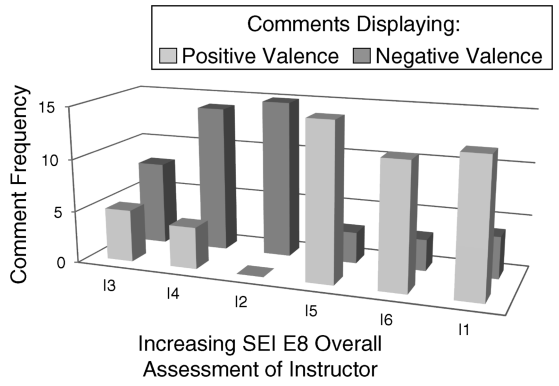


Figure 2. Comment valence by Student Evaluation of Instruction (SEI) E8 ratings of the instructors.

Interestingly, inspection of Figure 2 shows the same clusters observed in Table 4: I3, I4, and I2 at the low end and I5, I6, and I1 at the high end. It is worth emphasizing that Table 4 illustrates the consistency between RMP numerical data and SEI, whereas Figure 2 shows the consistency between qualitative data obtained from RMP and SEI. Another striking observation is the close count for the combined positive and negative comments for all instructors, 49 and 47 respectively. Although the distribution is uneven by instructor, all of them except one received positive and negative comments. These observations challenge the notion that RMP is a place for disgruntled students to vent and pursue retaliation and that contributors are more likely to post ratings on the Web site "to whine or complain".[9]

■ IMPLICATIONS FOR INSTRUCTION

From the perspective of instructional design, gathering of feedback and assessment data are indispensable to develop and improve pedagogical practices conducive to the attainment of the intended learning goals. The main purpose of our study was to assess the suitability of using RMP as a source of supporting information to analyze general chemistry programs, thereby contributing to their evaluation and improvement. Evidence shows that RMP is consistent with the SEI used at this institution during the same period. Additionally, we did not find significant differences between RMP contributors and non-contributors for general chemistry students. Similar programs can adapt the validation of the RMP data and the derived analysis procedure. This procedure allows categorizing student statements in a systematic and meaningful manner and extracting valuable supporting information usable in program evaluation. We maintain that rigorous, comprehensive sampling and analysis of a subset of courses and instructors bear several advantages. It counteracts the eventual effect of bias introduced by self-selection of contributors and, at the same time, it may appear less threatening to faculty. It is no secret that, unfortunately, in dealing with any kind of evaluation of instruction "some teachers apparently believe that nothing students say could possibly be of value to them".[16] This procedure is not a stand-alone method and we propose its use in tandem with other sources of evidence especially those that furnish direct information from the students such as in-depth interviews.

Recently, and probably derived from the influence of rating Web sites, some schools have started to post online results of their evaluations of instruction. However, it is uncommon to

dx.doi.org/10.1021/ed300277n | J. Chem. Educ. 2013, 90, 289−295

make other information available to assist in interpretation of data. In our survey, 64% of the students reviewed chemistry professors' RMP ratings. We agree with others[7,9] in that students will continue to take RMP very seriously unless they find better online alternatives for exchanging information about courses and instructors. Institutionalizing rating sites designed in accordance with scholarly assessment standards and in response to the needs of students may be an important step forward. Evidently, this constitutes an opportunity to elevate the information available above the oversimplified criteria of student satisfaction and instructor's behavior to include things such as attainment of learning goals. In an era in which creating student-centered environments, empowering students, shifting the responsibility of learning onto them, and similar mantras are a staple in instructional discourse, we may need to consider more seriously the instructors' responsibility of listening to students' voices. It is up to the individual institutions to assess how disenfranchised their students are when making relevant decisions about the learning process and environment and, if need be, to establish corrective actions. If we seek and find means to give students a voice we may learn more about their actual concerns and not have to resort to guessing based on instructors' own biases. Furthermore, we maintain that in their condition of expert students, pupils develop a unique student understanding of pedagogy that can inform expert instructors about the effectiveness of their practice.

■ ASSOCIATED CONTENT

Supporting Information

RMP Patterns of Use Survey; RMP ratings categories and their scales as they appear on the site; Student Evaluation of Instruction forms. This material is available via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: ssandi@usf.edu.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Home page for Rate My Professors. http://www.ratemyprofessors.com/ (accessed Dec 2012).
(2) Bleske-Rechek, A.; Michels, K. *Practical Assess. Res. Eval.* **2010**, *15*, 1−12.
(3) Manning, M. Where the Students Do the Grading: A Content Analysis of RateMyProfessors.com. M.S. Thesis, University of South Florida, Tampa, FL, 2005.
(4) Noer, M. America's Top Colleges. http://www.forbes.com/sites/michaelnoer/2011/08/03/americas-top-colleges/ (accessed Dec 2012).
(5) Otto, J. *Assess. Eval. High. Educ.* **2007**, *33*, 1−16.
(6) Gonyea, N. E.; Gangi, J. M. *Assess. Eval. High. Educ.* **2010**, *37*, 45−55.
(7) Brown, M. J.; Baillie, M.; Fraser, S. *Coll. Teach* **2009**, *57*, 89−92.
(8) *The Princeton Review*'s Best 300 Professors. http://www.princetonreview.com/best-professors.aspx (accessed Dec 2012).
(9) Davidson, E.; Price, J. *Assess. Eval. High. Educ.* **2009**, *34*, 65.
(10) Cashin, W. E. Student Ratings of Teaching: The Research Revisited. IDEA Paper No. 32; Center for Faculty Evaluation and Development, Kansas State University: Manhattan, KS, 1995.
(11) Patton, M. Q. *Qualitative Research and Evaluation Methods*; Sage Publications: Thousand Oaks, CA, 2002.
(12) Creswell, J. W. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*; Sage Publications, Inc.: Thousand Oaks, CA, 2009; p 260.
(13) Floriano, M. A.; Reiners, C. S.; Markic, S.; Avitabile, G. The Uniqueness of Teaching and Learning Chemistry. In *Innovative Methods of Teaching and Learning Chemistry in Higher Education*; Eilks, I., Byers, B., Eds.; RSC Publishing: Cambridge, U.K., 2009; pp 24−42.
(14) Kindred, J.; Mohammed, S. N. *J. Comput.-Mediated Commun.* **2005**, *10* (3), 00 DOI: 10.1111/j.1083-6101.2005.tb00257.x.
(15) Centra, J. A. *Res. High. Educ.* **2003**, *44*, 495−518.
(16) Heyden, T.; Henthorne, T. *Teaching English in the Two Year College* **2002**, *30*, 156−162.