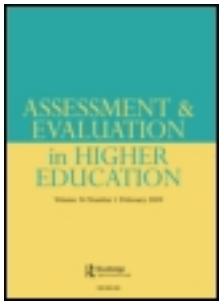

Assessment & Evaluation in Higher Education

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/caeh20

Evaluation of instruction: students' patterns of use and contribution to RateMyProfessors.com

Adrian Villalta-Cerdas[a], Patrick McKeny[a], Todd Gatlin[a] & Santiago Sandi-Urena[a]

[a] Department of Chemistry, University of South Florida, Tampa, FL, USA

Published online: 17 Mar 2014.

To cite this article: Adrian Villalta-Cerdas, Patrick McKeny, Todd Gatlin & Santiago Sandi-Urena (2014): Evaluation of instruction: students' patterns of use and contribution to RateMyProfessors.com, Assessment & Evaluation in Higher Education, DOI: 10.1080/02602938.2014.896862

To link to this article: http://dx.doi.org/10.1080/02602938.2014.896862

Assessment & Evaluation in Higher Education, 2014
http://dx.doi.org/10.1080/02602938.2014.896862

Routledge
Taylor & Francis Group

Evaluation of instruction: students' patterns of use and contribution to RateMyProfessors.com

Adrian Villalta-Cerdas, Patrick McKeny, Todd Gatlin and Santiago Sandi-Urena*

Department of Chemistry, University of South Florida, Tampa, FL, USA

RateMyProfessors.com (RMP) is the most popular commercial website to evaluate instructors, and houses a wealth of student-generated information in the form of ratings and reviews. This study investigated whether general chemistry students who use RMP were different from other students, and their reasons to use and contribute to the site. A pool of 398 students were surveyed. The survey gathered demographic information, patterns and frequency of use, and information to characterise participants in terms of their learning/grade orientation. Data analysis included descriptive and inferential statistics and the application of mixture models (latent class analysis and latent profile analysis). There were no significant differences between contributors and non-contributors in terms of gender, major, year status, grade point average, course load, previous chemistry grade and learning/grade orientation. Findings for this sample do not support the common assumption that students use RMP to shop for easy instructors. Instructor helpfulness, clarity and overall rating were reported as important aspects to visit the site. Findings agree with literature reports and support the use of RMP as supplemental data to inform assessment of instruction at the general chemistry programme level. Programmes in other disciplines can adapt the methods presented to assess the suitability of RMP data-sets for similar purposes.

Keywords: student evaluation; RateMyProfessors.com; chemical education; assessment of instruction

Introduction

Student evaluations of teaching (SET) proliferated across US campuses during the late 1960s and early 1970s (Chisholm 1977; Clayson 2009), becoming a staple in higher education. In addition to alternate uses that institutions may have for these evaluations, the original intention to capture and use student views about courses and instructors for the immediate betterment of instruction prevails, at least in principle. While research in the field is in general abundant (Otto et al. 2007), findings are often contradictory and student evaluation of instruction continues to be a very complex issue that ignites controversy (Clayson 2009; Felton, Mitchell, and Stinson 2004; Hinz 2011). Even the choice of parameters used in the research literature to assess the reliability of SETs has been recently argued to be inappropriate (Mosley 2014). Amidst the controversy, the standard methods of evaluation of instruction in higher education remained more or less stable until the emergence of online professor-rating sites at the turn of the century, an occurrence that without a doubt augmented the debate surrounding student evaluations, their validity and their potential uses.

*Corresponding author. Email: ssandi@usf.edu

In 2011, Hinz listed 28 online instructor rating sites in four different continents – some extending beyond national borders – that shared two fundamental elements: open access to submit or read reviews and contributor anonymity. Although permanence of these sites varies, their proliferation is an indication of the globalised trend of this online rating phenomenon. In the past decade, RateMyProfessors.com (RMP) has emerged as the online professor-rating site par excellence in the US. With more than 14 million entries for more than 1.3 million professors from 7000 schools in 2013, RMP is the self-proclaimed highest traffic site of its type. Although according to the site's guidelines 'Professor ratings should only be posted by users who have taken a class from the professor or who are currently taking a class with the professor', RMP is public and free, and no control is exerted over the identity of contributors.

Visitors may choose to voluntarily rate instructors using a five-point scale. Three rating items refer to instruction and instructor: easiness, helpfulness and clarity. The average of the latter two is called the overall quality. Two additional items are required to complete an entry: contributor's 'interest level prior to attending class' and quality of the 'textbook used'. Although not indispensable to complete the submission, the site collects information on the letter grade received, attendance (mandatory or not) and whether the course is online. An optional rating item, 'hotness', invites contributors to rate physical attractiveness. Contributors can choose to access a separate web page with a brief description of the rating categories. Additionally, they may leave open-ended comments that must be consistent with a series of guidelines (e.g. absence of profanity or references to professor's personal life). An RMP site moderation team is in charge of removing comments that do not abide by the guidelines. Like any online rating or review website, RMP content is user-generated and its express purpose is to serve as a resource for other users in their decision-making, in this case students weighing their course options. Therefore, RMP content is not meant as feedback for professors or institutions. Several other sources have described RMP in detail, (Bergin et al. 2013; Bleske-Rechek and Michels 2010; Davidson and Price 2009).

We view RMP as a new platform for the old practice of gathering comments from peers about instructors and courses to inform academic decisions (Kindred and Mohammed 2005), but with the added value of an expanded circle of interlocutors (Flanagin, Hocevar, and Samahito 2013; Li and Wang 2013). Nonetheless, besides affordances, the challenges of social networking and information use on the World Wide Web apply to RMP, and concerns about reliability and validity are warranted (Bergin et al. 2013; Bleske-Rechek and Michels 2010; Brown, Baillie, and Fraser 2009). Amongst the frequent criticisms expressed against RMP is that the contributors are self-selected, and students with a distinctively extreme opinion about the professor – either positive or negative – may be over-represented (Albrecht and Hoopes 2009). That contributors are self-selected is indisputable; that they are mainly driven to contribute by a desire to rant or rave is a commonly held assumption. Therefore, we believe the matter of debate should be whether this grouping is inherently different from the rest of the students in a given sample. The growing body of research on this topic has not produced statistically significant evidence to support such claim of difference (Albrecht and Hoopes 2009; Bergin et al. 2013; Bleske-Rechek and Michels 2010; Brown, Baillie, and Fraser 2009; Coladarci and Kornfield 2007).

Online instructor rating sites such as RMP are not a substitute for institutional evaluations or SETs. The nature of these two data sources is fundamentally different, because students frame the experience of providing feedback in significantly different ways (Bergin et al. 2013; Flanagin, Hocevar, and Samahito 2013; Kindred and Mohammed 2005; Otto et al. 2007). The inherently complementary nature of RMP in reference to other information sources makes it valuable (Bergin et al. 2013; Otto et al. 2007). As pointed out by Gonyea and Gangi (2010), students writing for the consumption of other students may address topics different from those they would address when communicating with instructors or institutions. Unfortunately, reflection about and research on student motives to visit RMP – why they value its information and how they use it – has been relegated, almost exclusively, in favour of the discussion about validity of the site (Felton, Mitchell, and Stinson 2004; Kindred and Mohammed 2005; Li and Wang 2013; Nasser and Fresko 2002). This emphasis on legitimacy of the site may be fuelled by personal views and anecdotal assumptions.

There is a rather generalised impression that faculty object to student evaluations regardless of the platform used for their collection (Nasser and Fresko 2002). In the case of online ratings, this reaction may be accentuated because RMP makes historical information public and extensively accessible, while word-of-mouth evaluations and opinions (Li and Wang 2013) are conveniently unrecorded and easily ignored. It will come as no surprise that negative assumptions about RMP persist even in the light of opposing evidence (Albrecht and Hoopes 2009; Bleske-Rechek and Michels 2010). However, regardless of the pre-conceived impressions of instructors and the accuracy (or lack thereof) of RMP, a large proportion of students trust the site to represent instructors' abilities (Brown, Baillie, and Fraser 2009). Literature has suggested that most students consider RMP a credible source to inform their decision-making process (Li and Wang 2013). For instance, in their survey of 216 students in various disciplines, Davidson and Price (2009) found that a staggering 95% deemed the site credible. As is the case with reviewing sites in general, RMP may influence expectations and framing of an upcoming experience, and thereby the nature of the experience itself (Edwards et al. 2007; Gonyea and Gangi 2010; Lewandowski, Higgins, and Nardone 2012). The potential influence of RMP on students' experience of learning is an incentive to explore and seek better understanding of why and how students use RMP.

It is our intention to put forth an approach that differs from other research work in the field in significant ways: (a) we view RMP as a source of supplemental data and not evaluations, (b) our analysis is at the programme level and not at the instructor level, and (c) we do not intend to generalise our findings to other data-sets; instead, we propose a methodological approach to evaluate individual data-sets obtained from RMP (or similar online instructor rating sites). Evaluation has a definitive and terminal connotation that we do not attribute to RMP. Far from validating the site (or not) and partaking in this controversy, our intention is to establish the usefulness of RMP information as data (Cashin 1995; Otto et al. 2007) that, in combination with other sources, may inform meaningful evaluation of instruction at the programme level (Bergin et al. 2013). In our view, the responsibility rests upon the evaluators to incorporate multiple sources of data to advance a proper judgement of instruction effectiveness.

Although most researches on RMP have used a large number of instructors from a variety of disciplines, the unit of analysis has continued to be the individual

instructor, and the mean rating values have been the preferred parameters to conduct analysis. Instead, we focus on validating data-sets at the programme level to then use the open-ended comments and not the ratings to inform evaluation. We envision the aggregation of data across multiple instructors and semesters as a tool to increase variation and reduce the potential effect of bias (Bergin et al. 2013). Furthermore, analysis of RMP data at the programme level removes the instructor anxiety associated with being rated personally, therefore making instructors more receptive to consider the outcomes of a comprehensive evaluation.

We do not intend to draw conclusions *about* the RMP site but to propose a methodological approach to assess the appropriateness of RMP data-sets. Ultimately, although we frame this study within a specific discipline, there is nothing exclusive that would prevent the use of this methodological approach to inform the decision regarding the use or not of RMP information by other departments or institutions.

This study is part of our undergraduate research programme constructed from student perspectives, and informed by their experiences as key participants in their own education. The fundamental purpose of this programme is to inform and impact undergraduate instruction. The goals of this study were to investigate whether students in a general chemistry course who use RMP constitute a subgroup, and to gather understanding about their reasons to visit and contribute to the site. To accomplish these goals, we addressed three guiding questions:

(1) Are students who contribute to RMP different from the general chemistry cohort? If so, in what ways?
(2) What information available on RMP do visitors find more valuable?
(3) What motivates students to contribute ratings and reviews to RMP?

Study design and procedures

RMP Patterns of Use Survey

We utilised a single instrument to capture information that addresses the three guiding questions. This 22-item instrument investigates RMP patterns of use and was modelled after work by Bleske-Rechek and Michels (2010). All participants completed the first survey section in which we collected demographic information on six criteria: (1) gender, (2) major, (3) year status, (4) grade point average (GPA), (5) course load and (6) previous chemistry grade. This section segregated participants who had never heard about or visited RMP from those who self-identified as users (Figure 1). In the second section of the survey, participants who self-identified as RMP users responded to prompts related to their use of RMP to make decisions about professors in general and chemistry professors in particular. RMP users rated the importance of six pieces of information available to them on RMP: (1) clarity, (2) easiness, (3) helpfulness, (4) number of postings, (5) reviews of other students and (6) overall rating. We refer to these as the RMP rating criteria and we used a ten-step scale (1, *not at all important*, to 10, *most important*).

RMP users also responded to two trade-off questions that examined their relative preference for learning vs. easy grades: their learning/grade orientation (Bleske-Rechek and Michels 2010). One of these items posed a hypothetical choice between an easy yet dull elective course and one that is interesting but difficult. In the second trade-off question participants selected between a course in which they would be

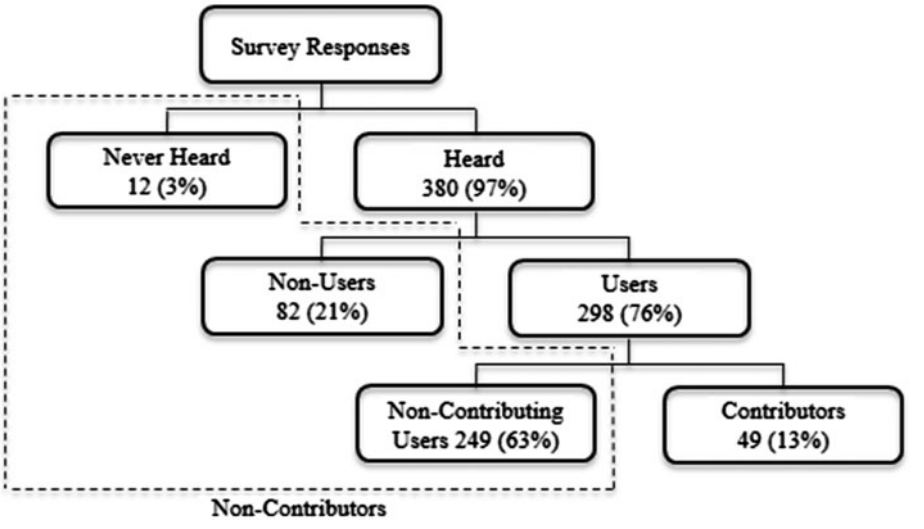

Figure 1. RMP Patterns of Use Survey subgroups.

highly likely to earn an A but learn next to nothing, and one in which they would likely learn a lot but where earning an A would be highly unlikely. The trade-off questions used a six-step scale with extremes indicating complete preference of one course over the other. Only RMP contributors – those who had rated instructors and left comments – responded to the third section of the survey, and addressed 24 possible reasons for their decision to contribute a review. Eighteen of these reasons came from the literature (Bleske-Rechek and Michels 2010); the remaining six were suggestions from three undergraduate students. Each reason was presented as a statement that participants rated in a ten-step scale (1, *not at all important*, to 10, *most important*). These statements included perceptions about the nature of the course, instruction, instructor and course assessment. Finally, an optional, open-ended question encouraged contributors to list their motivations to review professors.

Context and data collection

This study took place at a large, research intensive, south-eastern US university. Convenience sampling, a non-probability technique based on accessibility of participants produced information from two independent cohorts of General Chemistry 2 students. In the fall of 2011, we surveyed 276 students (49% of the initial course enrolment) using a combination of lecture and laboratory sampling. This included the lecture section taught by one of the co-authors, and seven laboratory sections (20–24 students each) taught by chemistry doctoral students who were not involved with this research. Six of these surveys were incomplete or missing information and were discarded.

For the second cohort, we surveyed 122 students from a single General Chemistry 2 lecture section taught by one of the co-authors (fall 2012, initial enrolment 197). In both data collections, we used hard copies of the anonymous RMP Patterns of Use Survey, and students marked their choices directly on the survey. Students

who chose to answer the optional, open-ended item wrote on the same survey. There were no written assignments in this large enrolment course (180–200 student-sections), anticipating that participants might have been preoccupied that the instructor would recognise their handwriting. Chemistry-graduate teaching assistants administered the survey in the lecture hall before the start of lecture. Research members transferred multiple-choice selections manually to electronic format in preparation for data handling and analysis. We compared the two cohorts across gender, class status, GPA, most recent chemistry grade, course load and major, finding no statistically significant differences at the 95% confidence level. We aggregated both data-sets into a single pool. Data collection followed Institutional Review Board guidelines; no compensation was offered for participation.

Comparison of RMP contributors and non-contributors

We investigated potential differences between contributors and non-contributors by two means. The first comparison used general characteristics of the subgroups: the demographic information. The second comparison focused on the learning/grade orientation and used the trade-off questions.

By means of the RMP Patterns of Use Survey, we segregated participants into those who left RMP ratings and comments, the contributors, and the rest of the respondents, the non-contributors (Figure 1). Subsequently, we used categorical statistics (χ^2 tests) to compare the response distribution of these two groups along gender, major, year status, GPA, course load and previous chemistry grade.

We utilised a multivariate technique, latent class analysis (LCA), to analyse and interpret responses to the trade-off questions. LCA, as other related mixture models (Collins and Lanza 2010; Lazarsfeld and Henry 1968; Marsh et al. 2009; Pastor et al. 2007), assumes there is an unobservable, underlying categorical variable (latent variable) that can be measured through observable variables (manifest variables). In this case, the categorical responses to the trade-off questions were the input observed variables. The learning/grade orientation was the latent variable and the output of the LCA allowed the characterisation of its classes (Collins and Lanza 2010). These classes described subgroups of participants in more detail and were more informative than the raw mean scores for the responses. Once we had established the learning/grade orientation classes, we analysed the distribution of RMP contributors and RMP non-contributors across these classes (χ^2 association test).

Importance attributed to RMP information and comparison of RMP contributors with non-contributing users

Students' importance ratings of the items available to them on RMP may act as a proxy for their reasons to access the site. We gathered this information through user ratings of the six RMP rating criteria: clarity, easiness, helpfulness, number of postings, reviews of other students and overall rating. Mean importance scores provided a general idea about what the users were looking to gather from visiting RMP. Likewise, comparison of RMP contributors, those that visit the site and leave ratings and comments, with non-contributing users, those who visit the site but leave no ratings or comments, might help to uncover whether the contributors have different motives.

We believe that the comparison of the mean values across the six RMP rating criteria only scratches the surface of the interpretation that may be drawn from the data. Consequently, we supplemented the analysis with latent profile analysis (LPA: Pastor et al. 2007), a mixture model related to LCA but different in the nature of the observed variables. In LPA, the observed variables are continuous; whereas in LCA, they are categorical (Pastor et al. 2007). In this particular case, the latent variable was the motivation to visit RMP, and student ratings of the importance of RMP rating criteria were the observed variables. The classes that emerged from the LPA represent profiles of the different users of RMP. These profiles are based on the participants' appreciation of RMP information, and they condense students' reasons for visiting the site. We performed a χ^2 test across these classes to test whether there were differences in the ways that RMP contributors and non-contributing users utilised RMP.

Reasons to contribute ratings and reviews to RMP

Self-identified RMP contributors rated a series of 24 statements that probed their agreement with reasons to rate and review instructors. In this particular case, the large number of variables and small number of contributors prevented use of mixture model analyses such as LCA and LPA (Collins and Lanza 2010; Pastor et al. 2007). We assigned valence to those items that allowed positive or negative affective attributions (Bergin et al. 2013). We considered praise of instructor or instruction, and comments that reflected satisfaction with the experience as positive valence. Negative valence included disapproval and perceived instructor's faults or mistakes. Initially, each co-author coded item valence individually and final assignments were the product of consensus. All but four items were clear-cut and presented no discrepancies amongst coders. We identified these four outliers as being more subject to interpretation than the rest.

For instance, while a challenge may seem an unnecessary aggravation for some students, others may celebrate it as a learning episode. Therefore, we deemed 'I felt the course challenged me' valence neutral, since in this context we could not conclude the possible attributions. Likewise, the other three discrepant items were assigned valence neutrality. In their individual coding, all four coders concurred the item 'I just felt like posting' was valence neutral, bringing the count of such items to five. In addition to these five valence neutral (0) items, eight statements reflected positive emotions or attitudes and carry a positive valence (+) and 11 statements negative valence (−). We used a conventional analysis in which we ranked the mean values and inspected the data for general trends in valence and content of the items. Additionally, we performed basic content analysis of the responses to the open-ended question 'What was your motivation behind writing a review on RateMyProfessors.com?' to compare those responses with the potential reasons we provided in the statements.

Results

Demographic and learning/grade orientation comparisons

Figure 1 shows the use of the RMP Patterns of Use Survey to identify subgroups within the sample. Most students, 97%, were aware of the existence of RMP, 76%

acknowledged they had specifically reviewed ratings for chemistry professors and 13% of the respondents had rated their chemistry professors on RMP. The non-contributing users group, 63%, describes the segment of the sample that visited the website to gather information but that did not rate a professor. The non-contributors group assembles non-contributing users with all other participants who did not rate a professor, that is, those who did not know the site or knew about the site but never visited. In addition, survey data also showed that 67% of the participants (88% of RMP users) agreed they found RMP useful for making choices about professors.

Table 1 shows the distribution for the RMP contributors and non-contributors across the demographic criteria. Participants had the option not to answer individual survey items; this freedom led to occasional small differences in participant counts. We excluded first-year students for two reasons: the small count and that students taking General Chemistry 2 in their first college semester do not have a prior college chemistry experience to report on. The results in Table 1 suggest students who contributed to RMP were not statistically significantly different from non-contributors in terms of the six demographic criteria probed.

We utilised a second dimension to compare RMP contributors and non-contributors: the learning/grade orientation. The interpretation of the LCA results produced two distinctive classes for the learning/grade orientation latent variable: the learning-

Table 1. Comparison of contributors and non-contributors, RMP Patterns of Use Survey.

Criteria	Values (*N*)	Contributors (%)	Non-contributors (%)	χ^2 (*p*) values
Gender	Male (161)	34.6	42.3	1.10 (0.29)
	Female (229)	65.4	57.7	
Major	Biology or related (124)	40.4	30.5	2.24 (0.52)
	Chemistry or related (42)	7.7	11.2	
	Biomedical Sciences (123)	28.8	32.0	
	Other or undecided (101)	23.1	26.3	
Status (year in school)	Second year (210)	48.9	57.2	1.23 (0.54)
	Third year (126)	40.4	32.7	
	Fourth year (38)	10.6	10.1	
GPA (four-point scale)	4.00–3.75 (44)	11.7	11.3	9.70 (0.10)
	3.74–3.50 (91)	35.3	21.7	
	3.49–3.25 (93)	27.5	23.4	
	3.24–3.00 (84)	9.8	23.4	
	2.99 or below (76)	15.7	20.2	
Course load (credits)	1–7 (15)	0	4.4	2.94 (0.57)
	8–10 (20)	5.8	5.0	
	11–13 (132)	30.8	34.2	
	14–15 (167)	48.1	41.9	
	16–18 (57)	15.4	14.5	
Most recent chemistry grade	A+ to A− (96)	23.1	24.8	1.61 (0.45)
	B+ to B− (212)	61.5	53.1	
	C+ and below (83)	15.4	22.1	

Note: Participants had the option not to answer individual items; therefore the sum may differ from the total 392.

oriented class and the learning-grade ambivalent class. These classes emerged from the statistical analysis of the response patterns and the descriptions or labels are reached through substantive analysis of membership in these classes.

Table 2 shows the probability of each class to choose the options presented for both trade-off questions. The learning-oriented class was a small group (13%) that clearly favoured learning over attaining a high grade. Given a choice, these students would prefer a course that is interesting but difficult in which they may learn substantially in spite of the risk of not earning an A. The remaining 87% did not report a clear-cut preference for learning over grades or vice versa, therefore we called it the learning-grade ambivalent class. However, it is evident from Table 2 that their preference excluded the options that indicated 'absolute' learning orientation. Although these students were *likely* to consider a course that was 'interesting but difficult', they also knew they would not *definitely* take it. That is, despite ambivalence, to some extent their preference gravitated closer to the grade orientation end.

We were interested in figuring out whether learning/grade orientation was associated with tendency to contribute to RMP; therefore, once the orientation groups emerged from the LCA analyses, we proceeded to run statistical comparisons. Table 3 shows there was no statistically significant difference in the distribution of

Table 2. Probability of choosing each option for the trade-off questions by learning/grade orientation class.

Trade-off question 1: with Section X's instructor you would learn a lot but you would be unlikely to earn an A. With Section Y's instructor you would learn next to nothing but you would be highly likely to earn an A. Which section would you choose?

	Section choice					
	Section X Learn much, unlikely A			Section Y Learn little, likely A		
Class	Definitely X	Likely X	Maybe X	Maybe Y	Likely Y	Definitely Y
Learning-oriented (LO)	0.48	0.38	0.003	0	0.03	0.09
Learning/goal ambivalent (LGA)	0	0.20	0.22	0.21	0.25	0.11

Trade-off question 2: Course A is guaranteed to be very dull but also very easy. Course B is guaranteed to be very interesting but also very difficult. Which course would you be more likely to choose?

	Course choice					
	Course A Dull but easy			Course B Interesting but difficult		
Class	Definitely A	Likely A	Maybe A	Maybe B	Likely B	Definitely B
Learning-oriented (LO)	0.16	0	0	0.03	0.28	0.53
Learning/goal ambivalent (LGA)	0.18	0.19	0.20	0.23	0.20	0

Table 3. Comparison of RMP contributors, non-contributing users and non-users by learning/grade orientation class.

	RMP group (*n*)		
Class	Contributors (49)	Non-contributing users (235)	Non-users (67)
Learning-oriented, LO, (%)	92	86	88
Learning/grade ambivalent, LGA, (%)	8.0	14	12

Notes: $\chi^2 = 1.51$, $p = 0.47$.

contributors, non-contributing users and non-users (Figure 1) between the two learning/grade orientation classes ($\chi^2 = 1.51$, $p = 0.47$). Likewise, there was no significant difference when we compared the contributors with the non-contributors ($\chi^2 = 1.22$, $p = 0.27$). That is, there is no statistically significant evidence the contributors were any different from the other groups in terms of their learning/grade orientation.

Importance attributed to RMP information and comparison of RMP contributors with non-contributing users

Table 4 shows the mean score for the six information criteria available to RMP users ($n = 298$, Figure 1). The 10-step scale ranged from 1, not at all important, to 10, most important. On average participants found all criteria to be important (all means >5); however, helpfulness, overall rating and clarity scored at the top whereas reviews, easiness and number of postings fell in the lower end.

In addition to considering the mean scores, we studied the patterns of response across the six RMP rating criteria (Table 4) using LPA. Our intention was to investigate the occurrence of subgroups of users to judge whether contributors displayed different patterns in their reasons to use RMP. Using standard analysis criteria and interpretation of the LPA results, we inferred the presence of four distinct classes. Table 5 shows these four classes and their corresponding mean ratings for the RMP criteria. The largest RMP users group, the endorsers (71%), subscribed strongly to the use of RMP and rated all criteria high. Ratings for the moderates (13%) were barely above middle values, suggesting that this group uses RMP information somewhat cautiously. A very small percentage of the users, the antagonists (2%), reported use of the site despite not finding any of the items important at all. The fourth group, the discerners (14%), made significantly different attributions to different criteria, and they valued helpfulness and clarity substantially more than easiness.

Table 4. Mean scores for RMP rating criteria.

RMP rating criteria (*n*)	Mean score (SD)
Helpfulness rating (291)	8.5 (2.0)
Overall rating (284)	8.2 (2.2)
Clarity rating (292)	8.1 (2.1)
Reviews of other students (289)	7.6 (2.4)
Easiness rating (291)	7.2 (2.5)
Number of postings (290)	6.2 (2.9)

Table 5. Mean ratings for RMP criteria by RMP patterns of use class.

RMP class (292)	RMP criteria mean rating					
	Helpfulness	Clarity	Overall rating	Reviews	Easiness	Number of postings
Endorsers (209)	9.1	8.6	9.2	8.4	8.1	6.8
Moderates (37)	5.1	6.2	6.3	5.7	5.8	5.0
Antagonists (6)	1.4	3.0	1.8	1.4	1.8	3.2
Discerners (40)	9.4	7.8	5.7	5.6	4.7	4.5

To further compare the RMP contributors with the non-contributing users, we conducted an association test across the RMP patterns of use classes in Table 5. Table 6 shows the results for this comparison. The small count of the antagonists group precluded its use for the χ^2 test. Lack of significant difference in the distribution of the RMP groups (Table 6) suggests there is no association between RMP patterns of use class and tendency to contribute ratings and reviews. That is, there is no statistically significant evidence to support the idea that participants who contributed to RMP perceived the usefulness of the site in any different way than participants who visited and did not contribute.

Reasons to contribute ratings and reviews to RMP

Table 7 shows the assigned valence, the mean score and standard deviation for each of the 24 statements addressing reasons to post on RMP. The six top-ranked reasons (Table 7, Group 1, mean scores 7.6–8.0) were positive valence statements that described overall satisfaction, especially with the instructor, but also with instruction and the course. In addition, giving others tips on how to do well in the class ranked in this top group. There were no items related to grades, course difficulty or negative assessment of instructor or course quality in this top tier.

Negative aspects about teaching (difficult explanations) and undesirable instructor traits (unhelpfulness) first appeared in Group 2 (mean scores 6.6–6.9). We considered that finding a course challenging is ambivalent, therefore valence neutral, as was warning students about the instructor. In this latter case, the assessment of the instructor was negative but the disposition to alert others was positive. In terms of valence, this second group was balanced, whereas Group 3 (mean scores 5.8–6.4) was predominantly negative. Group 3 contained the highest proportion of valence negative reasons and emphasised negative instructor traits (unfair, hard to understand) and undesirable aspects about the course (hard, waste of time, heavy workload). The remaining six items clustered in Group 4 (mean scores 2.5–5.0) and had no discernible commonality.

Table 6. Comparison of RMP contributors and RMP non-contributing users.

RMP patterns of use class	RMP group (*n*)	
	Contributors (48)	Non-contributing users (238)
Moderates (%)	20.8	11.3
Endorsers (%)	68.8	73.9
Discerners (%)	10.4	14.7

Notes: $\chi^2 = 3.45$, $p = 0.18$.

Table 7. Mean ratings for RMP users' reasons to contribute to RMP ($N = 48$).

Group	Reason statement	Valence	Mean (SD)	Rank	Rank in Bleske et al.
1	'I thought the instructor explained the material very well'.	+	8.0 (2.8)	1	N/A
	'I thought the instructor was an excellent teacher'.	+	8.0 (2.6)	2	2
	'I felt the instructor truly cared about their students'.	+	7.8 (3.0)	3	N/A
	'I thought the instructor was very helpful'.	+	7.8 (2.8)	4	4
	'I thought the course was really interesting'.	+	7.6 (2.9)	5	5
	'I wanted to give others tips on how to do well in the class'	+	7.5 (2.6)	6	3
2	'I thought the instructor was not at all helpful'.	−	7.0 (3.1)	7	1
	'I felt it was my duty to warn others about the instructor'.	0	6.9 (3.0)	8	6
	'The instructor had bad ratings, but I felt that they were a good teacher'.	+	6.9 (2.8)	9	13
	'I felt that the course challenged me'.	0	6.7 (2.7)	10	N/A
	'I found the instructor's explanations difficult to understand'	−	6.6 (2.8)	11	9
3	'I thought the workload was too heavy'.	−	6.3 (2.8)	12	7
	'I felt like the class was a waste of time'.	−	6.0 (3.5)	13	12
	'I thought the course was too hard'.	−	6.0 (2.9)	14	14
	'I felt the instructor was very lenient with their grading'.	+	5.9 (3.3)	15	N/A
	'The instructor had good ratings, but I felt they were a bad teacher'.	−	5.9 (3.0)	16	16
	'I thought the instructor was unfair to me or others in class'.	−	5.7 (3.2)	17	8
	'I found that the instructor's accent made it difficult to understand what they were saying'	−	5.7 (3.3)	18	N/A
4	'I thought the course demanded very little of my time outside of class'.	0	4.9 (3.3)	19	18
	'I felt the instructor was biased towards some students'.	−	4.8 (3.0)	20	N/A
	'I thought the course was ridiculously easy'.	0	4.7 (3.4)	21	17
	'I thought the instructor was boring'.	−	4.4 (3.2)	22	11
	'I was angry with the instructor'.	−	3.5 (3.1)	23	10
	'I just felt like posting'	0	2.6 (2.6)	24	15

Discussion

Online rating systems are now ubiquitous and available for almost anything, including higher education institutions and professors. According to the Pew Research Center website (2014), 78% of adult internet users in the US look for information online about a service or product. The percentage of students in this study who visited RMP (76%, Figure 1) is consistent, not only with this general public use of online rating sites, but also similar to values reported in the RMP literature (Brown,

Baillie, and Fraser 2009; Davidson and Price 2009). Thirteen per cent of our respondents rated their chemistry professors on RMP. This is considerably lower than the 26% of adult US internet users that contributed an online review for a product, service or person in 2013 (Pew Research Center 2014). Other RMP studies across broad disciplines have reported rates of contribution that range from 23 to 36% (Bleske-Rechek and Michels 2010; Brown, Baillie, and Fraser 2009; Davidson and Price 2009). Nevertheless, some of these studies revealed an association between discipline and posting with a lower contribution rate in the natural sciences (Bleske-Rechek and Michels 2010).

Lack of other reports specific to science courses or programmes prevents us from judging whether the contribution rate we observed is representative of this type of population. A significant 67% of our sample agreed they found RMP useful for making choices about professors. We postulate that a combination of strong identification with the source of information – other students – and lack of alternative sources contribute to users conferring credibility to RMP and similar online rating sites (Flanagin, Hocevar, and Samahito 2013). Our cohort was not unique in this regard: Davidson and Price (2009) reported that 95% of their sample found the site credible, three-quarters used the website to 'decide whether to take a particular instructor or not', and a third used it to 'select or avoid a course completely'. The extent to which this trust may persuade students to make decisions is hard to assess and may differ by institution, discipline and other indicators; however, it is apparent that in general students weigh this information favourably as a factor that influences their academic planning.

Our group comparisons (Table 1) add to the body of evidence that suggests RMP contributors are not significantly different from other students in terms of basic demographic indicators. Moreover, it would be reasonable to expect that, if differences existed, contributor bias would lead to significant differences between RMP ratings and other more widely accepted measures (e.g. institutional surveys). Several studies have investigated this kind of relationship, finding evidence to the contrary: an overall positive correlation between RMP ratings and corresponding items on institutional student evaluations of instruction (Albrecht and Hoopes 2009; Bergin et al. 2013; Brown, Baillie, and Fraser 2009; Coladarci and Kornfield 2007). For instance, Coladarci and Kornfield (2007) found evidence for concurrent validity using data from 426 instructors at a single institution. Brown, Baillie, and Fraser (2009) contrasted ratings for 312 instructors and found similar evidence.

Arguably, RMP contributors could differ from other students in aspects other than basic demographics. Another facet of the self-selection criticism is that students who are more focused on grades may be more inclined to contribute to RMP. To test this dimension of the assumption, Bleske-Rechek and Michels (2010) introduced an approach that compared learning/grade orientation of RMP contributors and non-contributors. We utilised this approach to strengthen our comparison, yet we used a different analytical tool to interpret the data. Whereas Bleske-Rechek and Michels (2010) compared the group mean values for the responses to the two trade-off questions (an item-centred approach), we utilised LCA (a person-centred approach). LCA is a powerful tool to elicit information about the characteristics of respondents that otherwise may be averaged out when the analysis exclusively uses mean scores. Using LCA, we first teased out two classes of respondents: a group that had a preference towards learning over grades and a group that was ambivalent in its orientation (Table 2).

Our main interest was in elucidating whether the evidence would uphold the persistent assumption that particularly biased, grade-focused students are over-represented in the RMP contributors group. Even though we did not identify a group that was exclusively inclined towards grades over learning, the learning/grade ambivalent group was not completely neutral in its preference. As seen in Table 2, this group was drawn closer to the grade orientation. It is worthy of attention here that these results are data dependent, and the outcome may be different for different sample. The essence of the matter is whether the RMP contributors for a given sample differ significantly from the rest of the sample. The association test showed that there was no statistically significant difference in the membership of the RMP groups as a function of learning/grade orientation (Table 3), an outcome that is aligned with previous findings (Bleske-Rechek and Michels 2010). Notwithstanding, in spite of this and other evidence (Bleske-Rechek and Fritsch 2011), the effect of self-selection continues to be a point of contention, with other researchers raising serious questions about the validity of the site (Davidson and Price 2009; Felton, Mitchell, and Stinson 2004; Legg and Wilson 2012).

Another widely held assumption, this time involving RMP users in general and not only contributors, is that visitors are particularly drawn to the website by their interest in identifying easy instructors. RMP users in this study did not consider easiness more important than other information available in the website: easiness was the second lowest ranked criterion (Table 4). No doubt these results must be interpreted cautiously, since all the mean scores in Table 4 clustered at the top half of the scale, with considerable overlap. Understandably, this may not be satisfactory for educators who would prefer easiness to be negligible when students make course decisions. However, given the associated dispersion, one may argue the mean values are, for practical purposes, not different. As it pertains to this data-set, this, in and of itself, challenges the presumption that students visit RMP especially motivated to identify easy instructors (Davidson and Price 2009).

A deeper analysis of the perceived importance of the RMP criteria using LPA produced four distinct classes of users (Table 5). The class with the largest membership of users, the endorsers (76%), attributed high importance to all pieces of information they accessed in RMP. It makes sense then that 67% of the sample reported to have used the site when making decisions related to their chemistry courses. The difference in these values may be associated with some students' choices being locked in by availability of sections within a course, because of schedule conflicts or similar situations. We hypothesise that the antagonists class, with a count of only 2%, corresponds to former users whose experience made them wary about the site's usefulness. Students whose expectations based on RMP conflicted significantly with their lived experiences might have developed scepticism towards the usefulness of the site. The moderates (13%) and the discerners (14%) are more discriminating, though apparently in different ways. The former group makes cautious use of RMP information, and we entertain the idea of students in this group viewing RMP as one of several sources of information. Unlike the other three groups, for which there was little dispersion among the mean scores for the six criteria, the discerners (14%) showed clear preference towards helpfulness and clarity over easiness. Evidently, the use of this statistical technique uncovers a wealth of information overlooked by simpler analyses. We envision the use of LPA to draw a purposeful sample that includes representatives of these four classes. In-depth interviews could shed further

light on the differences among these classes, especially on their perceptions about easiness.

The association test showed the distribution for RMP contributors and RMP non-contributing users across the RMP patterns of use classes are not statistically different (Table 6). However, it is interesting that the proportion of moderates – those we presume were more cautious in their use of RMP information – is greater for the contributors than for the non-contributing users. On the other hand, the opposite is true for the discerners. Felton, Mitchell, and Stinson (2004) studied the relationship between perceived quality, easiness and sexiness in web-based evaluations of 3190 professors at 25 universities, and concluded that course easiness influenced students who voluntarily evaluated their professors' teaching quality in a public forum. That did not seem to be the case for our sample: evidence does not support a significant difference between RMP contributors and the rest of the students.

It seems that, for our RMP contributors, highlighting positive aspects about instructors and instruction and communicating their impressions (good or bad) to other students were the main motivations to rate and review instructors (Table 7). Contrary to common assumptions (Bleske-Rechek and Michels 2010; Davidson and Price 2009), ranting about course difficulty, workload or grades was not a major motivation for this cohort. Eighteen of the 24 possible reasons in our survey came from Bleske-Rechek and Michels' work (2010). Interestingly, participants in their study and ours chose the same six top options out of those 18 reasons (albeit in a different sequence, Table 7), and in both cases items related to difficulty/easiness ranked low. These findings are in alignment with the content analysis of RMP comments reported by Kindred and Mohammed (2005). In their investigation of students' motives for use of RMP, these authors concluded that students in their sample were primarily concerned with instructors' competence and the classroom experience.

There were limitations in our probing of the possible reasons to use RMP in this study. First and foremost, the survey constrained participants to choose from options set by the researchers, which may not represent the universe of reasons that students would otherwise consider. Although the six reasons we added represented the input of undergraduates, it is not clear whether Bleske-Rechek and Michels (2010) validated the other 18 statements with actual students. With this in mind, in the fall of 2012 we asked students the open-ended question: 'What was your motivation behind writing a review on RateMyProfessors.com?' We hoped responses to this question would enlighten us about other probable reasons. The resemblance between the responses and the possible reasons on the survey was substantial. Student responses emphasised quality of instructor and instruction, traits of the instructor (helpful, caring), and a sense of responsibility to inform and help other students. Easiness, grades and ranting were not among the reasons that contributors reported.

Conclusion

The main purpose of the research project in which this study is embedded is to assess the suitability of using RMP data-sets as source of supplemental information to evaluate instruction at the programme level. While we framed this research project within a single college subject, chemistry, we maintain the methods are transferable and adaptable across disciplines. Furthermore, research gathered across multiple disciplines at this fine-grained level will complement those findings that stem from using broader samples. In previous work, we utilised mixed methods to establish the

convergence of RMP information with institutional student evaluations of instruction (Bergin et al. 2013). In the present study, we have presented evidence contradicting common assumptions about students who use and contribute to RMP: (a) that RMP contributors are substantially different from the rest, (b) that RMP visitors are especially drawn to the site to gather information about course/instructor easiness and (c) that ranting and raving are particularly important motives for students to contribute to RMP.

Our work revealed that, for this cohort of participants, RMP contributors were not different in terms of six indicators that included GPA and prior chemistry grades. Neither did they manifest a disproportionate preference for grades at the expense of learning, at least, not to a greater extent than their peers. In general, RMP users did not assign higher importance rating to easiness than they did to the other information available in the site. Furthermore, when queried about their reasons to post ratings and reviews, RMP users highlighted instructor and instruction quality and commitment to inform others over concerns related to grades, easiness/difficulty and entertainment or likeability of the instructor. Our evidence adds to a growing body of research that points at RMP as a source of information that should be considered seriously (Bleske-Rechek and Michels 2010; Otto et al. 2007).

Research suggests that RMP users develop expectations that not only alter their course choices but also their learning experiences and perceptions (Gonyea and Gangi 2010; Lewandowski, Higgins, and Nardone 2012). Therefore, accurate or not, RMP is important. Irrespective of their personal views and thoughts, educators benefit from understanding how the internet generation uses RMP, their reasons, motivations and expectations. This understanding can inform the development of better alternatives for students to access the type of information they seek in RMP. Additionally, instructors can encourage a sample that is more representative by promoting the use of the site among their own students (Coladarci and Kornfield 2007; Legg and Wilson 2012).

Institutions have limited resources and opportunities to gather information about what actually happens in class from the students' perspectives. Tapping RMP information may be a fruitful alternative. If nothing else, the controversy around RMP may be provocative enough to enliven a broader discussion about evaluation of instructors and instruction. College departments should make the outcomes of their institutional evaluations public and easily accessible, and promote their perusal by students (Brown, Baillie, and Fraser 2009; Coladarci and Kornfield 2007; Nasser and Fresko 2002). After all, as Kindred and Mohammed (2005, 20) put it, the 'main reason for accessing Ratemyprofessors.com is because students want information to aid in course selection; they also desire a variety of opinions and easy access to information'.

Evidence suggests there is value in giving voice to students, and systematic analysis of data collected online can inform instructional decisions. Despite instructors' common reticence to give credit to students' ability to judge good learning experiences, evidence has shown over several decades that generally they are good judges (McKeachie 1990). In our view, rather than making or accepting as valid unwarranted generalisations, educators interested in considering the information in RMP should test data-sets separately. As part of our undergraduate research programme, we are currently engaging in the investigation of general chemistry instructor perceptions of online and traditional student evaluations of instruction, their perceived use and instructor actions in light of student feedback.

Notes on contributors

Santiago Sandi-Urena received his PhD in chemistry from Clemson University, Clemson, South Carolina, USA, under the mentorship of Dr. Melanie M Cooper. He is currently an assistant professor with the Department of Chemistry at the University of South Florid, where his research focuses on chemical education. His research interests include the self-explaining effect in the ecology of large enrolment general chemistry courses, research on learning in the academic chemistry laboratory and evaluation of instruction. He has organised several national and international research symposia on learning in the academic laboratory and on multicultural and international perspectives of chemistry education.

Adrian Villalta-Cerdas is a 2008 graduate from the University of Costa Rica. He completed his master's degree in chemistry from the University of South Florida, where he is currently a doctoral candidate in chemistry focusing on chemical education. His doctoral research focuses on the investigation of the self-explaining effect in naturalistic large enrolment general chemistry courses.

Todd A. Gatlin is a doctoral candidate in chemical education at the University of South Florida. He completed his bachelor of arts in secondary science at the University of Mississippi and received a master's degree in chemistry from Clemson University. His dissertation research has focused on the metacognitive and epistemological development of graduate teaching assistants and general chemistry students in traditional and reformed academic laboratories.

Patrick McKeny is an undergraduate student in the Department of Chemistry at the University of South Florida. He has presented his undergraduate research in chemical education and inorganic chemistry at regional and national professional conferences. Upon graduation, he will pursue an MD/PhD programme.

References

Albrecht, W. S., and J. Hoopes. 2009. "An Empirical Assessment of Commercial Web-based Professor Evaluation Services." *Journal of Accounting Education* 27: 125–132.

Bergin, A., K. Sharp, T. A. Gatlin, A. Villalta-Cerdas, A. Gower, and S. Sandi-Urena. 2013. "Use of RateMyProfessors.com Data as a Supplemental Tool for the Assessment of General Chemistry Instruction." *Journal of Chemical Education* 90: 289–295.

Bleske-Rechek, A., and A. Fritsch. 2011. "Student Consensus on RateMyProfessors.com." *Practical Assessment, Research and Evaluation* 16 (18): 1–12. http://pareonline.net/getvn.asp?v=16&n=18.

Bleske-Rechek, A., and K. Michels. 2010. "RateMyProfessors.com: Testing Assumptions about Student Use and Misuse." *Practical Assessment, Research and Evaluation* 15 (5): 1–12. http://pareonline.net/getvn.asp?v=15&n=5.

Brown, M. J., M. Baillie, and S. Fraser. 2009. "Rating RateMyProfessors.com: A Comparison of Online and Official Student Evaluations of Teaching." *College Teaching* 57 (2): 89–92.

Cashin, W. E. 1995. *Student Ratings of Teaching: The Research Revisited*. IDE Paper No. 32. Center for Faculty Evaluation and Development. Manhattan, KS: Kansas State University.

Chisholm, M. G. 1977. "Student Evaluation: The Red Herring of the Decade." *Journal of Chemical Education* 54 (1): 22–23.

Clayson, E. J. 2009. "Student Evaluations of Teaching: Are They Related to What Students Learn? A Meta-analysis and Review of the Literature." *Marketing Education* 31 (1): 16–30.

Coladarci, T., and I. Kornfield. 2007. "RateMyProfessors.com versus Formal In-class Student Evaluations of Teaching." *Practical Assessment, Research and Evaluation* 12 (6): 1–15. http://pareonline.net/getvn.asp?v=12&n=6.

Collins, L. M., and S. T. Lanza. 2010. *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. Hoboken, NJ: Wiley & Sons.

Davidson, E., and J. Price. 2009. "How Do We Rate? An Evaluation of Online Student Evaluations." *Assessment & Evaluation in Higher Education* 34 (1): 51–65.

Edwards, C., A. Edwards, Q. Qing, and S. T. Wahl. 2007. "The Influence of Computer-mediated Word-of-mouth Communication on Student Perceptions of Instructors and Attitudes toward Learning Course Content." *Communication Education* 56 (3): 255–277.

Felton, J., J. Mitchell, and M. Stinson. 2004. "Web-based Student Evaluations of Professors: The Relations between Perceived Quality, Easiness and Sexiness." *Assessment & Evaluation in Higher Education* 29: 91–108.

Flanagin, A. J., K. P. Hocevar, and S. N. Samahito. 2013. "Connecting with the User-generated Web: How Group Identification Impacts Online Information Sharing and Evaluation." *Information, Communication and Society* : 1–12. doi:10.1080/1369118X.2013.808361.

Gonyea, N. E., and J. M. Gangi. 2010. "Reining in Student Comments: A Model for Categorising and Studying Online Student Comments." *Assessment & Evaluation in Higher Education* 37: 45–55.

Hinz, A. 2011. "Attitudes of German Teachers and Students towards Public Online Ratings of Teaching Quality." *Electronic Journal of Research on Educational Psychology* 9 (2): 745–764.

Kindred, J., and S. N. Mohammed. 2005. "He Will Crush You like an Academic Ninja!: Exploring Teacher Ratings on RateMyProfessors.com." *Journal of Computer-Mediated Communication* 10 (3), Article ID 9. doi:10.1111/j.1083-6101.2005.tb00257.x.

Lazarsfeld, P. F., and N. W. Henry. 1968. *Latent Structure Analysis*. New York: Houghton Mifflin.

Legg, A. M., and J. H. Wilson. 2012. "RateMyProfessors.com Offers Biased Evaluations." *Assessment & Evaluation in Higher Education* 37 (1): 89–97.

Lewandowski Jr., G. W., E. Higgins, and N. N. Nardone. 2012. "Just a Harmless Website?: An Experimental Examination of RateMyProfessors.com's Effect on Student Evaluations." *Assessment & Evaluation in Higher Education* 37: 987–1002.

Li, C., and X. Wang. 2013. "The Power of eWOM: A Re-examination of Online Student Evaluations of Their Professors." *Computers in Human Behavior* 29: 1350–1357.

Marsh, H. W., O. Lüdtke, U. Trautwein, and A. J. Morin. 2009. "Classical Latent Profile Analysis of Academic Self-concept Dimensions: Synergy of Person- and Variable-centered Approaches to Theoretical Models of Self-concept." *Structural Equation Modeling: A Multidisciplinary Journal* 16 (2): 191–225.

McKeachie, W. J. 1990. "Research on College Teaching: The Historical Background." *Journal of Educational Psychology* 82 (2): 189–200.

Mosley, D. 2014. "Assessing the Reliability of Student Evaluations of Teaching: Choosing the Right Coefficient." *Assessment & Evaluation in Higher Education* 39 (2): 127–139.

Nasser, F., and B. Fresko. 2002. "Faculty Views of Student Evaluation of College Teaching." *Assessment & Evaluation in Higher Education* 27 (2): 187–198.

Otto, J., D. A. Sanford Jr., A. Douglas, and D. N. Ross. 2007. "Does RateMyProfessor.com Really Rate My Professor?" *Assessment & Evaluation in Higher Education* 33 (4): 355–368.

Pastor, D. A., K. E. Barron, B. Miller, and S. L. Davis. 2007. "A Latent Profile Analysis of College Students' Achievement Goal Orientation." *Contemporary Educational Psychology* 32 (1): 8–47.

Pew Research Center. 2014. "Pew Research Center's Internet and American Life Project." Accessed February 11. http://www.pewinternet.org