ASSESSMENT & EVALUATION IN HIGHER EDUCATION
2019, VOL. 44, NO. 3, 449–462
https://doi.org/10.1080/02602938.2018.1520193

Routledge
Taylor & Francis Group

Understanding the perceived quality of professors' teaching effectiveness in various disciplines: the moderating effects of teaching at top colleges

Ya-Ling Chiu[a], Ku-Hsieh Chen[b], Yuan-Teng Hsu[c] and Jying-Nan Wang[a]

[a]College of Economics and Management, Chongqing University of Posts and Telecommunications, Chongqing, China; [b]Department of Applied Economics, Fo Guang University, Yilan, Taiwan; [c]Research Center of Finance, Shanghai Business School, Shanghai, China

ABSTRACT

Measuring the perceived quality of professors' teaching effectiveness is a critical issue in higher education. This study involves a large-scale data exploration with a sample of 16,802 professors, in which each professor had received at least 20 ratings from the RateMyProfessors website. We find that perceived difficulty (from the students' perspective) has a significantly negative effect on perceived quality. However, when professors teach more difficult courses at top colleges, the decline in perceived quality is relatively small when compared to other colleges. In other words, whether professors come from top colleges has a moderating effect on the relationship between perceived quality and perceived difficulty. Furthermore, through a consideration of the characteristic differences among disciplines in terms of the relationship between perceived quality and perceived difficulty, we obtain three specific groups of disciplines. These findings facilitate a better understanding of quality for professors from different disciplines. We suggest that the measurement of teaching effectiveness should avoid the use of a single criterion because differences in courses, disciplines or schools can influence the measurement results, and these factors are beyond the control of professors.

KEYWORDS

Perceived quality; online evaluations; teaching effectiveness; discipline; brand awareness

Introduction

Student evaluations of teaching effectiveness (SETs) are a key component in higher education. In general, official SETs, designed by schools or administrations, are conducted following well-defined procedures in the classroom or by means of certain Web-based systems which students can use to complete the evaluations. For professors, these SET reports are important because they might contribute to improving the course quality (Knol et al. 2013; Safavi et al. 2013), and they are also a major consideration for promotion, tenure and awards at most colleges (Otto, Sanford, and Ross, 2008; Silva et al. 2008). A more detailed introduction to SETs can be found in the study of Marsh (2007). When students attempt to choose among courses offered by several professors, it is not easy for them to access the related SETs in order to make a good decision. With the emergence of RateMyProfessors.com (RMP), a popular and easily accessible public forum for rating professors and universities, evaluations began to take on another relevance.

CONTACT Yuan-Teng Hsu yuanteng.hsu@gmail.com Research Center of Finance, Shanghai Business School, Shanghai, China.

Since its initiation in 1999, RMP has recorded more than 17 million ratings for over 1.6 million professors and 7000 schools across the United States, Canada and the United Kingdom. Moreover, about 4 million college students use this website each month (RateMyProfessors.com 2017). Unlike when they use official SETs, through this website students can not only directly provide evaluations on professors according to their experience but also easily obtain past evaluations for many professors. Therefore, professors' RMP evaluations can be regarded as a form of electronic word-of-mouth (Hartman and Hunt 2013; Li and Wang 2013). In other words, RMP works in a way similar to common online shopping websites where customers can browse others' reviews to aid them in making purchasing decisions and can post their own product reviews/ratings. Before exploring the determinants of RMP evaluation, it is necessary to first confirm the site's accuracy in reflecting students' evaluation of their professors. Several studies have examined the correlation between RMP evaluations and official SETs to verify the accuracy of the former (Timmerman 2008; Brown, Baillie, and Fraser, 2009; Sonntag, Bassett, and Snyder, 2009). Legg and Wilson (2012) concluded that in-class and RMP evaluations may produce inconsistent results in the rating of professors. Furthermore, since RMP evaluations are completely anonymous, there is no guarantee that every rating or comment indeed comes from a student of the professor being rated (Johnson and Crews 2013).

Despite skepticism about the validity of RMP, its popularity among students is still high. Steffes and Burgee (2009) found that when students seek information about a specific professor, the information they access from RMP is more influential in their decisions than discussions about the professor with friends in person. In addition, a study by Davison and Price (2009) shows that 92% of students are aware of RMP, 80% of students visit it more than once, 95% of students think it is credible and 75% use it to choose instructors. To sum up, the RMP evaluations not only provide useful feedback about instructor quality but also potentially affect students' perceptions of instructors and courses (Edwards et al. 2007; Brown et al. 2009; Bleske-Rechek and Fritsch 2011; Kowai-Bell et al. 2011; Edwards and Edwards 2013; Li and Wang 2013). Furthermore, students' behavior in class, such as participation and note-taking, is also influenced by RMP (Kowai-Bell et al. 2012).

Although the audience of RMP is primarily students, not professors, it influences professors in a similar way to SETs, in spite of the lower accuracy of RMP (Kowai-Bell et al. 2012; Boswell 2016). Specifically, RMP evaluations significantly affect professors' self-efficacy. Self-efficacy refers to people's beliefs about their capabilities to finish a task or accomplish a goal (Bandura 1977). With high self-efficacy, people have greater motivation to attempt new tasks (Bandura 1982, 1989). For example, if professors receive more positive feedback and perceived successes, they generally are more willing to make greater effort in their classes with the result that students will have more opportunities to enjoy improved learning situations. Moreover, the professor's self-efficacy is significantly related to the students' learning outcomes such as engagement (Tschannen-Moran and Woolfolk Hoy 2001; van Uden, Ritzen, and Pieters, 2014; Boswell 2016), motivation (Midgley, Feldlaufer, and Eccles 1989), achievement (Ross 1992) and students' own sense of efficacy (Anderson, Greene, and Loewen 1988).

In this study, we assume that the perceived quality of a professor can be measured effectively by their RMP evaluations. We want to emphasise that this notion of quality is the reflection of students' personal perceptions and may not correspond fully to the true quality of a professor's teaching. Perceived quality has a major effect on both students (Kowai-Bell et al. 2011, 2012; Lewandowski et al. 2012) and professors (Tschannen-Moran and Woolfolk Hoy 2001; van Uden et al. 2014; Boswell 2016). Therefore, greater understanding of the meaning of perceived quality can contribute to the ongoing development of higher education. Although a growing body of literature has investigated possible determinants of perceived quality such as hotness (attractiveness), easiness (level of difficulty), race and gender (Felton, Mitchell, and Stinson 2004; Riniolo et al. 2006; Felton et al. 2008; Stuber et al. 2009; Reid 2010; Leung, Jiang, and Busser 2013; Rosen 2018), there are still some aspects of this notion worthy of further exploration.

This paper makes three novel contributions. First, our sample is more representative than those of most past studies. We included in our sample over 16,800 professors who had at least 20 student ratings each, from 1592 schools. Based on this large-scale sample from RMP, we were able to carry out a broad exploration of perceived quality in 20 different disciplines. Second, we consider two important factors which might affect perceived quality values. One is the student's perceived level of difficulty in connection with each professor. In general, a higher perceived difficulty goes together with lower perceived quality. The other factor is whether the professor is from a top college (TOP) according to Forbes' America's Top Colleges 2016 report. We found that the average perceived quality of top college professors (4.11) is larger than that for professors from other colleges (3.69). More importantly, a professor's coming from a top college has a significant moderating effect on the relationship between perceived difficulty and perceived quality. Third, using a universal standard to evaluate professors from different disciplines is not appropriate. Since there are essentially different benchmarks used by students to rate professors from different disciplines, we use perceived difficulty and perceived quality as criteria to determine three groups of disciplines; recognition of these categories can contribute a new perspective on evaluations of professors.

Literature review

Antecedents of quality

Several academic studies have analysed various factors affecting perceived quality of professors. Felton et al. (2004) examine the relationships between perceived quality, easiness and hotness found with 3190 professors at 25 universities. They find that the correlation between quality and easiness is 0.61, and the correlation between quality and hotness is 0.30. Moreover, students tend to give hotter professors higher quality and easiness scores. Riniolo et al. (2006) also point out that more attractive professors have higher overall quality scores than less attractive ones based on a study of 522 professors at four schools. In addition, Felton et al. (2008) investigate this issue using a larger database, including 6582 professors from 369 American and Canadian institutions. Their results support the findings of Felton et al. (2004). Moreover, they find significant differences of the perceived quality of professors among disciplines. The three fields with the highest average quality are languages, sociology and political science, and the three with the lowest average are engineering, computer science and chemistry.

Whether gender matters in evaluations of effective teaching is another prevailing question. Some studies have found that women receive less favorable ratings than male instructors (Tatro 1995; Heckert et al. 2014), but others do not find significant gender effects (Feldman 1993; Liddle 1997). The inconsistency found for the effect of gender could be explained as the result of gender interacting with other factors (Sprinkle 2008). For RMP evaluations, based on 500 professors in five public universities, Stuber et al. (2009) do not find significant evidence of gender differences in overall quality ratings. When traditionally female disciplines, such as the arts and humanities, are considered, female professors have more positive ratings (Sandler, Silverberg, and Hall 1996). Based on a large-scale study of RMP for about 190,000 US professors with at least 20 student ratings each, Rosen (2018) applies the R gender package to determine each professor's gender and concludes that the effect of gender on RMP evaluation is small but statistically significant.

Studies focusing on the impact of the instructor's race on SET have shown contradictory evidence. For example, Boatright-Horowitz and Soeung (2009) find that racial minority instructors receive more negative evaluations than White professors, but Ho, Thomsen, and Sidanius (2009) do not find significant evidence of racial bias. In the study of Reid (2010), data on 3079 White, 142 Black, 238 Asian, 130 Latino and 128 other race faculties in colleges of art were collected from RMP. Reid found that Blacks and Asians were evaluated more negatively than White faculty

in terms of overall quality, helpfulness and clarity. Several interactions between race and gender were found, such as that Black male faculty were rated more negatively than others. Accordingly, due to negative racial stereotypes, administrations should consider more suitable standards when making decisions on the tenure or promotion of racial minority faculty.

Top colleges and quality

In this study, we explore whether professors working in top colleges is a factor impacting quality. This question is related to the concept of brand awareness, an important factor for consumers evaluating products (Aaker 1996). Specifically, the term brand awareness refers to the strength of a brand's connection in consumers' memory and reflects consumers' ability to identify brand differences; consumers with higher brand awareness have stronger links to the brand in their memory (Keller 1993). Therefore, brand awareness has a very important influence on the consumer decision-making process. Richardson, Dick, and Jain (1994) found that brand awareness is often used by consumers as an external heuristic to assess product quality. Similarly, the reliability of the brand influences consumers when they consider product selection because consumers usually believe that the brand represents product quality (Jacoby, Olson, and Haddock, 1971; Dodds, Monroe, and Grewal, 1991; Erdem and Swait 2004). Dodds et al. (1991) and Grewal et al. (1998) also put forward the same argument, which explains in part why products with high brand awareness usually have a higher market share and a higher quality evaluation. This study considers the idea that professors' working in top colleges can be regarded as a kind of brand awareness for students and that this factor might have a positive effect on perceived quality.

In addition to the direct influence on perceived quality, we consider whether professors' working in top colleges might also have a moderating effect on perceived quality. Since brand awareness is the link consumers make to a certain brand in their memory, products with higher brand awareness are more familiar to consumers (Keller 1993, 2002). In general, consumers also have more knowledge about familiar brands (Lemon and Nowlis 2002). Therefore, brand awareness as a factor in consumers' decision-making process, due to the corresponding differences in familiarity and knowledge, will further affect their behavior (Kamins and Marks 1991). In our context, given the same perceived difficulty, students in top colleges (those with higher brand awareness) are expected to have higher perceived quality (Dodds et al. 1991; Keller 2002). Furthermore, although it is expected that quality will decline with difficulty, the rate of decline for students in top colleges will be lower than that for others.

Methods

Data collection

The RateMyProfessors.com is a publicly accessible website. The site design allows students to evaluate professors according to three key criteria: clarity, helpfulness and easiness. Overall perceived quality is computed by combining the clarity and helpfulness ratings. However, on 18 May 2016, RMP simplified the evaluating process, and it now uses only the criterion of overall perceived quality; the easiness criterion is now presented as the level of difficulty, on a 1-to-5 scale, with 1 being very easy. Our sample was automatically extracted from RMP in December 2016. We implemented the web crawler technique, which is a common approach in recent empirical studies. For example, public data have been collected for research samples from Amazon (Singh et al. 2017; Wu 2017), Apple's App Store (Yin, Mitra, and Zhang 2016) and RateMyProfessors (Rosen 2018). Nevertheless, collecting data on all the professors on this site as the sample would require too much time. Thus, we randomly selected 85,306 professors as our sample, using their unique ID numbers for random selection. These professors were from 3799 colleges and universities within the United States. For each one, we obtained the first and last

Table 1. Data descriptions.

Variable	Description	Mean	St. dev.	Min	Median	Max
PQP	Perceived Quality of Professor	3.717	0.810	1.1	3.8	5
PLD	Perceived Level of Difficulty	2.899	0.770	1	2.9	5
TOP	Top Colleges (Yes = 1, No = 0)	0.115	0.319	0	0	1
DIS_1	English (Yes = 1, No = 0)	0.125	0.330	0	0	1
DIS_2	Mathematics (Yes = 1, No = 0)	0.111	0.315	0	0	1
DIS_3	History (Yes = 1, No = 0)	0.065	0.247	0	0	1
DIS_4	Psychology (Yes = 1, No = 0)	0.060	0.237	0	0	1
DIS_5	Biology (Yes = 1, No = 0)	0.054	0.227	0	0	1
DIS_6	Chemistry (Yes = 1, No = 0)	0.039	0.192	0	0	1
DIS_7	Communication (Yes = 1, No = 0)	0.034	0.180	0	0	1
DIS_8	Political Science (Yes = 1, No = 0)	0.030	0.170	0	0	1
DIS_9	Business (Yes = 1, No = 0)	0.029	0.169	0	0	1
DIS_{10}	Economics (Yes = 1, No = 0)	0.029	0.168	0	0	1
DIS_{11}	Philosophy (Yes = 1, No = 0)	0.027	0.161	0	0	1
DIS_{12}	Computer Science (Yes = 1, No = 0)	0.026	0.160	0	0	1
DIS_{13}	Sociology (Yes = 1, No = 0)	0.024	0.152	0	0	1
DIS_{14}	Science (Yes = 1, No = 0)	0.023	0.150	0	0	1
DIS_{15}	Languages (Yes = 1, No = 0)	0.023	0.148	0	0	1
DIS_{16}	Accounting (Yes = 1, No = 0)	0.020	0.140	0	0	1
DIS_{17}	Education (Yes = 1, No = 0)	0.015	0.121	0	0	1
DIS_{18}	Physics (Yes = 1, No = 0)	0.015	0.120	0	0	1
DIS_{19}	Music (Yes = 1, No = 0)	0.015	0.120	0	0	1
DIS_{20}	Anthropology (Yes = 1, No = 0)	0.014	0.119	0	0	1

name, department, institution, city, overall quality, level of difficulty and number of ratings. In order to ensure the reliability of this study, professors with fewer than 20 ratings or for whom some pieces of information were missing were removed from our sample. Finally, we retained 16,802 professors from 1592 schools in our research dataset.

Measurement

Table 1 lists the variables used in this study. The dependent variable, PQP, or the perceived quality of the professor, was obtained from RMP. Students are asked the question, 'how would you rate this professor as an instructor' and then give scores to professors based on their individual opinion. These scores range from 1 to 5 in integer increments, with 1 being the worst rating and 5 being the best rating. The value for the perceived quality of a certain professor is measured by the average of all their received scores. We also introduce two major independent variables, perceived difficulty and teaching at a top college. For measuring the perceived level of difficulty of the professor (PLD), students are asked the question, 'how hard did you have to work for this class' and provide scores ranging from 1 to 5, with 1 being very easy, to indicate the difficulty of the course. Thus, the difficulty is calculated as the average of these scores for each professor. The variable TOP is a dummy variable coded as 1 if the professor's institution belongs to Forbes America's Top 200 Colleges 2016 (http://www.forbes.com/top-colleges), by which the ranking score is calculated using five components: post-graduate success (32.5%), student debt (25%), student satisfaction (25%), graduation rate (7.5%) and academic success (10%). The calculation method of the ranking score can be found at https://www.forbes.com/sites/carolinehoward/2016/07/06/top-colleges-ranking-2016-the-full-methodology/. However, we believe that most students don't know about this methodology in detail. Therefore, this ranking might be roughly regarded as a proxy variable for college reputation due to the high visibility of the Forbes magazine. In our sample, there are 1938 (11.5%) professors from these top colleges.

Twenty other variables are dummy variables related to disciplines, which were selected depending on the number of professors in each discipline (DIS). The variable DIS_i, $i = 1, 2, \cdots, 20$, is coded as 1 if the professor belongs to one specific discipline as listed in the second column of Table 1. For example, a professor with $DIS_2=1$ is in the 'mathematics' department.

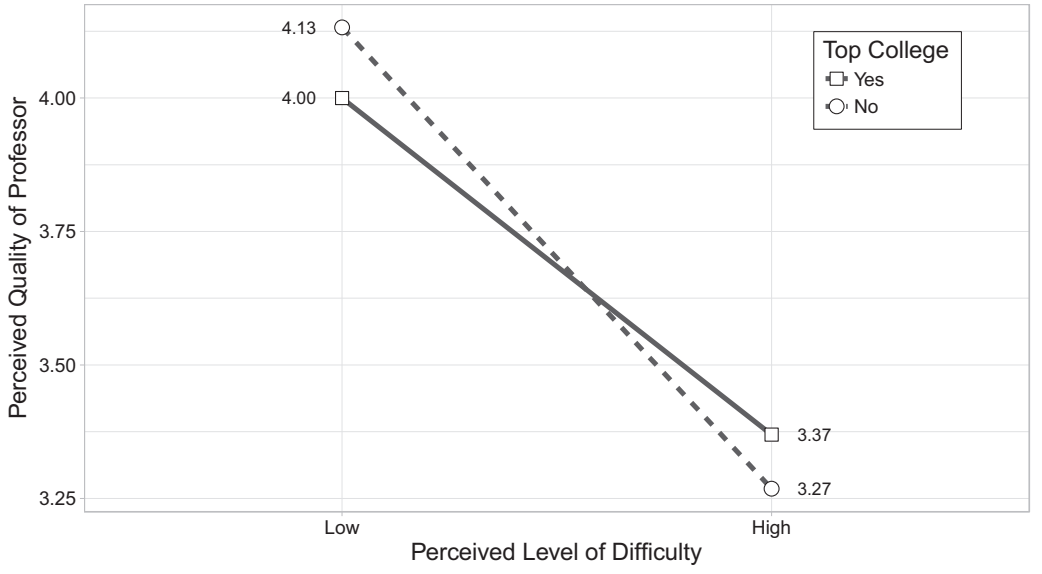

Figure 1. Moderating role of top college.

Robustness of the sample selection

In terms of sample size, our generated dataset is more representative than those used in previous studies about RMP, except for the work of Rosen (2018), in which 190,006 professors with a minimum of 20 ratings are considered. This large-scale study of RMP data can be used to check the robustness of our sample selection. A comparison of our data to that in Figure 1 in Rosen shows that our mean and median quality, at 3.72 and 3.80, are consistent with his overall quality statistics. More evidence for the robustness of our sample can be found from Rosen's Figure 4. For instance, for the physics and psychology disciplines, our averages of quality are 3.379 and 3.931, respectively. Both of these numbers are consistent with Rosen's findings. Therefore, although the size of the sample in this study is much smaller than that of Rosen (2018), our sample remains very representative thanks to random sampling.

Statistical analysis

To examine the research question of whether perceived difficulty and teaching at a top college will affect perceived quality, a multiple linear regression model is constructed as presented in Model (1),

$$\mathrm{PQP}_i = \beta_0 + \beta_1 \mathrm{PLD}_i + \beta_2 \mathrm{TOP}_i + \varepsilon_i \tag{1}$$

The above model is useful for examining the main effects of perceived difficulty and teaching at a top college on perceived quality. In regression analysis, a moderating effect can be characterised as an interaction that influences the strength of the relationship between the dependent (PQP) and independent (PLD) variables. To investigate the moderating effect of TOP, we further considered a multiple regression model with an interaction term, as shown in Model (2),

$$\mathrm{PQP}_i = \beta_0 + \beta_1 \mathrm{PLD}_i + \beta_2 \mathrm{TOP}_i + \beta_3 \mathrm{PLD}_i \times \mathrm{TOP}_i + \varepsilon_i \tag{2}$$

Table 2. Top college effects ($N = 16{,}802$).

	Dependent variable: Perceived Quality of Professor (PQP)		
Variable	Model (0)	Model (1)	Model (2)
PLD	−0.640*** (0.006)	−0.641*** (0.006)	−0.649*** (0.007)
TOP		0.030* (0.016)	0.020 (0.016)
PLD × TOP			0.084*** (0.022)
Constant	3.717*** (0.005)	3.713*** (0.005)	3.713*** (0.005)

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Results

Effects of difficulty and top colleges

The regression results of Model (1) are reported in Table 2. First, the coefficient of PLD is −0.641 ($p < 0.01$), which indicates a higher difficulty of a course generally comes with a lower perceived quality of the professor in the students' perception. The same result has been found in the previous literature (Felton et al. 2004). In addition, as shown in previous studies, a customer with higher brand awareness tends to perceive the quality of a product as higher (Dodds et al. 1991; Grewal et al. 1998; Teas and Agarwal 2000). This implies that professors in top colleges are likely to get higher ratings. Our results also support this prediction: the coefficient of TOP is 0.03 ($p < 0.05$). Although compared with that of the perceived difficulty, the effect of top colleges on perceived quality is slight, it is worthwhile to explore whether the moderating effect of top colleges has a significant influence on the relationship between perceived difficulty and perceived quality.

We examine the moderating effect of top colleges through Model (2) and show the corresponding results in Table 2. The interaction term is obtained by multiplying the independent variable and the moderator variable, so it will generate a multicollinearity problem. That is, there will be a high degree of correlation between the independent variable, moderator variable and interaction terms. Therefore, we used mean-centering to reduce the chances of multicollinearity (Aiken, West, and Reno 1991). Since the coefficient of interaction term is 0.084 ($p < 0.01$), it is clear that whether a professor is from a top college does affect the impact of perceived difficulty on perceived quality. With TOP = 0, perceived quality will drop by 0.649 when perceived difficulty increases by one unit; while given TOP = 1, perceived quality will drop by only 0.565 when perceived difficulty increases by one unit. The corresponding interaction plot is shown in Figure 1, which demonstrates that the rate of decline of perceived quality is slower when professors come from top colleges. Furthermore, when perceived difficulty is relatively low, a perceived quality with TOP = 0 is larger than a perceived quality with TOP = 1 ($4.13 > 4.00$); however, when perceived difficulty is relatively high, a perceived quality with TOP = 0 is lower than a perceived quality with TOP = 1 ($3.27 < 3.37$). These results all support the conclusion that the moderating effect of TOP does exist in our context.

Differences across disciplines

For purposes of investigating the impact of discipline on perceived quality, 16,802 professors were divided into different groups according to their disciplines. We include the 20 disciplines with the largest numbers of professors in the following discussion. The professors in these top 20 disciplines account for 77.6% of the total sample as shown in Table 3. For example, the discipline of English has the largest number of professors, 2092. By contrast, only 242 professors are from the discipline of anthropology. We discuss the differences in the characteristics of these 20 disciplines in terms of perceived quality and perceived difficulty. On this basis, we determine three specific groups and then investigate the moderating effect of TOP in each group.

Table 3 shows that the 6 disciplines with the highest perceived quality are psychology (3.931), communication (3.910), languages (3.864), music (3.851), English (3.810) and sociology (3.793). Similarly, the 3 disciplines with the highest perceived difficulty are chemistry (3.366), biology

Table 3. Means of PQP and PLD in 20 disciplines.

		PQP			PLD		
Discipline	*N*	Yes	No	Difference	Yes	No	Difference
English	2092	3.810	3.703	0.107***	2.827	2.909	−0.082***
Mathematics	1871	3.560	3.736	−0.176***	2.983	2.888	0.095***
History	1099	3.777	3.712	0.065***	2.976	2.893	0.083***
Psychology	1006	3.931	3.703	0.228***	2.718	2.910	−0.192***
Biology	915	3.697	3.718	−0.021	3.277	2.877	0.400***
Chemistry	647	3.469	3.726	−0.257***	3.366	2.880	0.486***
Communication	563	3.910	3.710	0.200***	2.640	2.908	−0.267***
Political Science	501	3.771	3.715	0.056*	2.970	2.897	0.073**
Business	494	3.645	3.719	−0.074*	2.883	2.899	−0.016
Economics	486	3.484	3.723	−0.239***	3.096	2.893	0.203***
Philosophy	447	3.654	3.718	−0.064*	2.934	2.898	0.036
Computer Science	441	3.531	3.722	−0.190***	2.875	2.899	−0.024
Sociology	396	3.793	3.715	0.078**	2.657	2.905	−0.247***
Science	386	3.563	3.720	−0.157***	3.183	2.892	0.291***
Languages	379	3.864	3.713	0.151***	2.756	2.902	−0.146***
Accounting	337	3.506	3.721	−0.215***	3.227	2.892	0.335***
Education	248	3.719	3.716	0.003	2.559	2.904	−0.345***
Physics	247	3.379	3.722	−0.343***	3.217	2.894	0.323***
Music	246	3.851	3.715	0.136***	2.439	2.906	−0.467***
Anthropology	242	3.729	3.716	0.012	2.771	2.901	−0.130***

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Note: For each discipline, our sample is divided into two sub-samples in which one consists of *N* professors belonging to the corresponding discipline and the other consists of remaining professors. The column 'Yes' ('No') reports the means of PQP or PLD calculated by the pre (post) subsample.

(3.277) and accounting (3.227), while the 3 disciplines with the lowest perceived difficulty are music (2.439), communication (2.640) and sociology (2.657). We apply a two-sample *t*-test to examine the difference of means between two subsamples, one of which consists of the professors in a certain discipline and the other of the professors who do not come from this discipline. We are also interested in the standard deviations of perceived quality and perceived difficulty in 20 disciplines, so we follow the same grouping procedures and present the corresponding results in Table 4. The variance ratio test is used to verify whether the difference of standard deviations between two subsamples is significant. For example, the perceived quality received by the professors in mathematics has relatively higher diversity than those from other disciplines (F-value $= 1.270$ and $p < 0.01$), but the perceived quality received by the professors in political science shows relative lower diversity than those from other disciplines (F-value $= 0.732$ and $p < 0.01$).

We summarise the differences among disciplines in Table 5 using the relevant data from Tables 3 and 4. The symbol + (−) means that a value, e.g., the mean of perceived quality or the standard deviation of perceived difficulty, in a certain discipline is significantly larger (smaller) than it is in others. The symbol '?' shows there is no significant result in that case. As shown in Table 5, we make three conclusions. First, with the exception of history and political science, when the mean of perceived quality is relatively high (low), the mean of perceived difficulty is relatively low (high). This outcome also supports the conclusion that the effect of perceived difficulty on perceived quality is negative. Second, in general, when the mean of perceived quality is relatively high or the mean of perceived difficulty is relatively low, the standard deviation of perceived quality is relatively low. However, since as many as nine disciplines do not show statistically significant differences in the standard deviation of perceived difficulty, we cannot make further inferences about this measure.

Finally, based on these features, we define three groups of disciplines as follows. The characteristics of a relatively low perceived quality and a relatively high perceived difficulty are used to define Group 1, which includes physics, chemistry, economics, accounting, mathematics and science. This type of discipline seems to require greater mathematical ability. Group 2 is determined by a relatively high perceived quality and a relatively low perceived difficulty. It includes the disciplines of psychology, languages, communication, music, English and sociology. Group 3 is more special; in it,

Table 4. Standard deviations of PQP and PLD in 20 disciplines.

		PQP			PLD		
Discipline	*N*	Yes	No	*F*-value	Yes	No	*F*-value
English	2092	0.795	0.811	0.961	0.766	0.770	0.989
Mathematics	1871	0.897	0.796	1.270***	0.727	0.775	0.879***
History	1099	0.738	0.815	0.820***	0.741	0.772	0.921*
Psychology	1006	0.720	0.813	0.784***	0.759	0.770	0.973
Biology	915	0.762	0.813	0.879***	0.682	0.769	0.784***
Chemistry	647	0.845	0.807	1.096*	0.625	0.770	0.659***
Communication	563	0.739	0.812	0.829***	0.771	0.769	1.007
Political Science	501	0.696	0.813	0.732***	0.678	0.773	0.769***
Business	494	0.849	0.809	1.102	0.752	0.771	0.951
Economics	486	0.767	0.810	0.897	0.671	0.772	0.756***
Philosophy	447	0.748	0.812	0.850**	0.676	0.773	0.766***
Computer Science	441	0.869	0.808	1.157**	0.783	0.770	1.035
Sociology	396	0.754	0.811	0.864**	0.724	0.770	0.884*
Science	386	0.799	0.810	0.973	0.724	0.770	0.885
Languages	379	0.742	0.811	0.836**	0.676	0.772	0.766***
Accounting	337	0.808	0.809	0.996	0.730	0.770	0.900
Education	248	0.923	0.808	1.303***	0.818	0.768	1.132
Physics	247	0.903	0.807	1.250**	0.696	0.770	0.816**
Music	246	0.749	0.811	0.854*	0.758	0.768	0.972
Anthropology	242	0.700	0.812	0.745***	0.644	0.772	0.697***

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Note: For each discipline, our sample is divided into two sub-samples in which one consists of *N* professors belonging to the corresponding discipline and the other consists of remaining professors. The column 'Yes' ('No') reports the standard deviations of PQP or PLD calculated by the pre (post) subsample.

Table 5. Summary of difference in disciplines.

Discipline	Mean (PQP)	Mean (PLD)	SD (PQP)	SD (PLD)
Philosophy	−	?	−	−
Business	−	?	?	?
Computer Science	−	?	+	?
Economics	−	+	?	−
Accounting	−	+	?	?
Science	−	+	?	?
Chemistry	−	+	+	−
Mathematics	−	+	+	−
Physics	−	+	+	−
Anthropology	?	−	−	−
Education	?	−	+	?
Biology	?	+	−	−
Languages	+	−	−	−
Sociology	+	−	−	−
Communication	+	−	−	?
Music	+	−	−	?
Psychology	+	−	−	?
English	+	−	?	?
History	+	+	−	−
Political Science	+	+	−	−

Note: The symbols + (−) indicates that the statistic (Mean or SD) of the discipline is significantly larger (smaller) than others. The symbol ? means the relation is not significant.

both perceived quality and perceived difficulty are relatively high. This group includes history and political science. We organise these characteristics in Figure 2. The *x*-axis and the *y*-axis represent the average perceived difficulty and the average perceived quality, respectively. The circle size indicates the standard deviation of perceived difficulty (larger circles reflecting greater standard deviation), and the color depth represents the standard deviation of perceived quality (the darker colors reflecting smaller standard deviation). The differences in nature across disciplines, in terms of perceived quality and perceived difficulty, are presented visually in Figure 2. Specifically, disciplines located in the lower right corner belong to Group 1 (PLD $>$ 2.98 and PQP $<$ 3.56), and disciplines located in the upper left corner belong to Group 2 (PLD $<$ 2.83 and PQP $>$ 3.79).

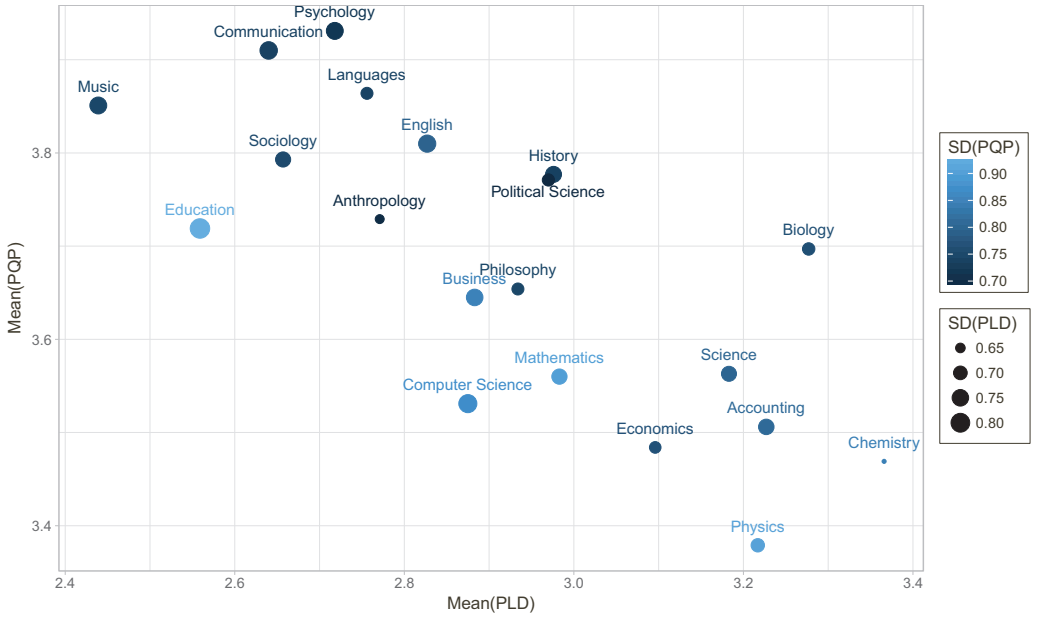

Figure 2. PQP and PLD in different disciplines.

We then explore the moderating effects of teaching at a top college in these three groups and show the results in Table 6. For all groups, the coefficients of the interaction terms are significantly positive, showing that the moderating effect of top colleges on the relationship between perceived quality and perceived difficulty does exist. Specifically, given TOP = 0, the groups with the declining rates of quality ranged from large to small are Group 1 (−0.804), Group 2 (−0.605) and Group 3 (−0.580). On the other side, given TOP = 1, the groups with the declining rates of perceived quality ranged from large to small are Group 1 (−0.678), Group 2 (−0.513) and Group 3 (−0.363). We also show three corresponding interaction plots in Figure 3. In summary, our results support the conclusion that top colleges can lessen the rate of decline of quality when perceived difficulty increases. However, the moderating effects of top colleges have different degrees of impact among three groups, especially in Group 3.

Conclusion and discussion

This study uses a large-scale sample from the RMP website to explore two factors affecting the perceived quality of professors. Based on the empirical results we arrive at two major findings. First, perceived difficulty has a significantly negative effect on perceived quality. When professors develop more difficult courses, students tend to give them poorer ratings as compared to easier courses. In addition, a top college has a significantly positive moderating effect on the relationship between perceived difficulty and perceived quality. A declining rate of perceived quality will be lessened when professors come from top colleges. In addition to the effect of brand awareness, another possible interpretation of this finding is that top colleges usually have better students who have higher learning abilities or are more willing to take time to study, which makes these students more likely to adapt to difficult courses. Thus, the perceived quality of professors from top colleges is less influenced by perceived difficulty. Second, we identify differences in the nature of disciplines with regard to perceived quality and perceived difficulty. As shown by the summarised results in Table 5, differences in the degree of difficulty of courses lead to differences in the assessment of professors' quality (Felton et al. 2008; Rosen 2018). Based on these characteristics of disciplines, we can identify three specific groups of disciplines and further verify the moderating effect of top colleges in each group.

Table 6. Top college effects: three groups of disciplines.

	Dependent variable: Perceived Quality of Professor (PQP)		
Variable	Group (1)	Group (2)	Group (3)
PLD	−0.804*** (0.015)	−0.605*** (0.012)	−0.580*** (0.022)
TOP	0.008 (0.036)	0.086*** (0.030)	−0.063 (0.045)
PLD × TOP	0.126*** (0.048)	0.092** (0.044)	0.217*** (0.069)
Constant	3.687*** (0.011)	3.749*** (0.010)	3.821*** (0.016)
N	3974	4682	1600
R^2	0.436	0.354	0.313

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

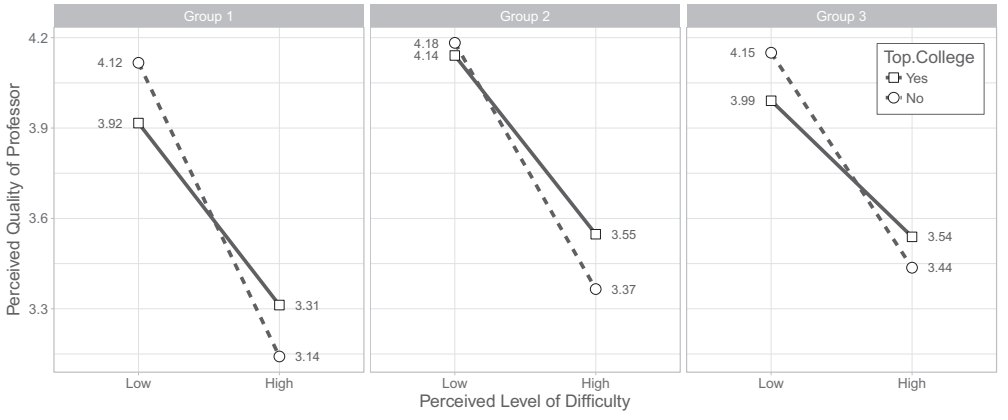

Figure 3. Moderator effects of top college in three groups.

When students attempt to find a good professor through the RMP website, they should consider perceived quality, perceived difficulty and whether the professor teaches at a top college in evaluating a professor's quality. Since the effects of perceived difficulty and teaching at a top college on perceived quality are complex, the website should consider providing further information to help students make decisions more effectively. For example, the website might present students with an average perceived quality for specified degrees of perceived difficulty or within the same college. On the other hand, professors responsible for tougher courses face a dilemma between pursuing good ratings and including the required teaching content.

Of course, professors should continue to improve their teaching techniques and content to lower the perceived difficulty to students so that their perceived quality can increase. However, our empirical results show that the discrepancies in the nature of disciplines cannot easily be eliminated. For example, the average perceived quality of professors in physics is 3.379, while that of professors in psychology is 3.931. Since perceived quality is significantly related to students' learning outcomes (Tschannen-Moran and Woolfolk Hoy 2001; Kowai-Bell et al. 2011; Kowai-Bell et al. 2012; van Uden et al. 2014; Boswell 2016), we advise teachers to understand these ratings correctly and avoid losing their teaching enthusiasm.

Furthermore, our results indicate that students in top colleges are more accepting of difficult courses, so the curriculum design should take into account the characteristics of students in the particular school. Finally, our results also lead to suggestions for schools or government administrations. While we do not propose a perfect evaluation method, administrators should realise that it is not generally easy for professors whose specialty is more closely related to mathematics to receive high ratings, while professors whose specialty is related to language or policy usually have higher ratings. In other words, since the nature of disciplines is different, it is not appropriate to apply the same criteria to assess the quality of professors from different disciplines.

Acknowledgements

The authors would like to express their sincere appreciation to the editor and the anonymous referees for their valuable comments and suggestions, which helped improve the quality of the paper tremendously. All authors contributed equally to this manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work is supported by the Chongqing University of Posts and Telecommunications Teaching Project of China (Project Number XFZS1803).

Notes on contributors

Ya-Ling Chiu is currently a Professor in the faculty of School of Economics and Management, Chongqing University of Posts and Telecommunications, China. She received her PhD in College of Management (Major in Marketing) from Yuan Ze University. Her research interests involve online Word-of-Mouth behavior, online community behavior and relationship marketing.

Ku-Hsieh Chen is a professor of Economics and chief executive officer (dean) of the Commission of General Education at Fo-Guang University, Taiwan. He obtained his Ph.D. from the Tamkang University (2003-2008). His research interests involve the Industrial Economics, Managerial Economics, Financial Economics, Productivity and Efficiency Analysis, Applied Econometrics, Financial Market and Institution, and Energy Economics. To date, he has published numerous research papers in high quality refereed academic journals, including Journal of Banking and Finance, Research in Higher Education, Energy Economics, China Economic Review, Journal of Productivity Analysis, Small Business Economic, Energy Efficiency, Applied Economics, and Agricultural Economics (Zemedelska Ekonomika).

Yuan-Teng Hsu is an Assistant Professor at the Research Center of Finance, Shanghai Business School, China. He received his PhD in College of Management (Major in Finance) from Yuan Ze University. His research examines corporate finance issues, investment and financial accounting.

Jying-Nan Wang is currently a Professor in the faculty of School of Economics and Management, Chongqing University of Posts and Telecommunications, China. He received his PhD in College of Management (Major in Finance) from Yuan Ze University. His research interests involve business intelligence analysis, consumer reviews analysis, financial and economics information system, programming trading and market microstructure.

ORCID

Ya-Ling Chiu http://orcid.org/0000-0001-5466-0993
Ku-Hsieh Chen http://orcid.org/0000-0002-8330-9892
Yuan-Teng Hsu http://orcid.org/0000-0001-8805-3964
Jying-Nan Wang http://orcid.org/0000-0002-2315-2520

References

Aaker, D. A. 1996. *Building Strong Brands*. New York: Simon and Schuster.

Aiken, L. S., S. G. West, and R. R. Reno. 1991. *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: Sage.

Anderson, R., M. Greene, and P. Loewen. 1988. "Relationships among Teachers' and Students' Thinking Skills, Sense of Efficacy, and Student Achievement." *Alberta Journal of Educational Research* 34 (2): 148–165.

Bandura, A. 1977. "Self-efficacy: Toward a Unifying Theory of Behavioral Change." *Psychological Review* 84 (2): 191–215.

Bandura, A. 1982. "Self-efficacy Mechanism in Human Agency." *American Psychologist* 37 (2): 122–147.

Bandura, A. 1989. "Human Agency in Social Cognitive Theory." *The American Psychologist* 44 (9): 1175–1184.
Bleske-Rechek, A., and A. Fritsch. 2011. "Student Consensus on RateMyProfessors.com." *Practical Assessment Research & Evaluation* 16 (18): 1–12.
Boatright-Horowitz, S. L., and S. Soeung. 2009. "Teaching White Privilege to White Students Can Mean Saying Good-Bye to Positive Student Evaluations." *American Psychologist* 64 (6): 574–575.
Boswell, S. S. 2016. "Ratemyprofessors is Hogwash (but I Care): Effects of Ratemyprofessors and University-Administered Teaching Evaluations on Professors." *Computers in Human Behavior* 56: 155–162.
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.
Davison, E., and J. Price. 2009. "How Do we Rate? An Evaluation of Online Student Evaluations." *Assessment & Evaluation in Higher Education* 34 (1): 51–65.
Dodds, W. B., K. B. Monroe, and D. Grewal. 1991. "Effects of Price, Brand, and Store Information on Buyers Product Evaluations." *Journal of Marketing Research* 28 (3): 307–319.
Edwards, A., and C. Edwards. 2013. "Computer-Mediated Word-of-Mouth Communication: The Influence of Mixed Reviews on Student Perceptions of Instructors and Courses." *Communication Education* 62 (4): 412–424.
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.
Erdem, T., and J. Swait. 2004. "Brand Credibility, Brand Consideration, and Choice." *Journal of Consumer Research* 31 (1): 191–198.
Feldman, K. A. 1993. "College Students' Views of Male and Female College Teachers: Part II—Evidence from Students' Evaluations of Their Classroom Teachers." *Research in Higher Education* 34 (2): 151–211.
Felton, J., P. T. Koper, J. Mitchell, and M. Stinson. 2008. "Attractiveness, Easiness and Other Issues: Student Evaluations of Professors on RateMyProfessors.com." *Assessment & Evaluation in Higher Education* 33 (1): 45–61.
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 (1): 91–108.
Grewal, D., R. Krishnan, J. Baker, and N. Borin. 1998. "The Effect of Store Name, Brand Name and Price Discounts on Consumers' Evaluations and Purchase Intentions." *Journal of Retailing* 74 (3): 331–352.
Hartman, K. B., and J. B. Hunt. 2013. "What RateMyProfessors.com Reveals about How and Why Students Evaluate Their Professors: A Glimpse into the Student Mind-set." *Marketing Education Review* 23 (2):151–162.
Heckert, T. M., A. Latier, A. Ringwald, and B. Silvey. 2014. "Relation of Course, Instructor, and Student Characteristics to Dimensions of Student Ratings of Teaching Effectiveness." *American Journal of Business Education (AJBE)* 3 (1): 195–203.
Ho, A. K., L. Thomsen, and J. Sidanius. 2009. "Perceived Academic Competence and Overall Job Evaluations: Students' Evaluations of African American and European American Professors." *Journal of Applied Social Psychology* 39 (2): 389–406.
Jacoby, J., J. C. Olson, and R. A. Haddock. 1971. "Price, Brand Name, and Product Composition Characteristics as Determinants of Perceived Quality." *Journal of Applied Psychology* 55 (6): 570–579.
Johnson, R. R., and A. D. Crews. 2013. "My Professor Is Hot! Correlates of RateMyProfessors.com Ratings for Criminal Justice and Criminology Faculty Members." *American Journal of Criminal Justice* 38 (4): 639–656.
Kamins, M. A., and L. J. Marks. 1991. "The Perception of Kosher as a Third Party Certification Claim in Advertising for Familiar and Unfamiliar Brands." *Journal of the Academy of Marketing Science* 19 (3): 177–185.
Keller, K. L. 1993. "Conceptualizing, Measuring, and Managing Customer-Based Brand Equity." *Journal of Marketing* 57 (1): 1–22.
Keller, K. L. 2002. "Branding and Brand Equity." In *Handbook of Marketing*, edited by Bart Weitz, Robin Wensley, 151–178. London, UK: Sage Publications.
Knol, M. H., R. Veld, H. C. M. Vorst, J. H. van Driel, and G. J. Mellenbergh. 2013. "Experimental Effects of Student Evaluations Coupled with Collaborative Consultation on College Professors' Instructional Skills." *Research in Higher Education* 54 (8): 825–850.
Kowai-Bell, N., R. E. Guadagno, T. Little, N. Preiss, and R. Hensley. 2011. "Rate My Expectations: How Online Evaluations of Professors Impact Students' Perceived Control." *Computers in Human Behavior* 27 (5): 1862–1867.
Kowai-Bell, N., R. E. Guadagno, T. E. Little, and J. L. Ballew. 2012. "Professors are People Too: The Impact of Informal Evaluations of Professors on Students and Professors." *Social Psychology of Education* 15 (3): 337–351.
Legg, A. M., and J. H. Wilson. 2012. "RateMyProfessors.com Offers Biased Evaluations." *Assessment & Evaluation in Higher Education* 37 (1): 89–97.
Lemon, K. N., and S. M. Nowlis. 2002. "Developing Synergies between Promotions and Brands in Different Price-Quality Tiers." *Journal of Marketing Research* 39 (2): 171–185.
Leung, X. Y., L. Jiang, and J. Busser. 2013. "Online Student Evaluations of Hospitality Professors: A Cross-Cultural Comparison." *Journal of Hospitality, Leisure, Sport & Tourism Education* 12 (1): 36–46.

Lewandowski, G., E. Higgins, and N. N. Nardon. 2012. "Just a Harmless Website? An Experimental Examination of RateMyProfessors.com's Effect on Student Evaluations." *Assessment & Evaluation in Higher Education* 37 (8): 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 (4): 1350–1357.

Liddle, B. J. 1997. "Coming out in Class: Disclosure of Sexual Orientation and Teaching Evaluations." *Teaching of Psychology* 24 (1): 32–35.

Marsh, H. W. 2007. Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness. In *The Scholarship of Teaching and Learning in Higher Education: An Evidence-based Perspective*, 319–383. Dordrecht: Springer.

Midgley, C., H. Feldlaufer, and J. Eccles. 1989. "Change in Teacher Efficacy and Student Self- and Task-Related Beliefs in Mathematics during the Transition to Junior High School." *Journal of Educational Psychology* 81 (2): 247–258.

Otto, J., D. A. Sanford, Jr., and D. N. Ross. 2008. "Does Ratemyprofessor.com Really Rate my Professor?." *Assessment & Evaluation in Higher Education* 33 (4): 355–368.

RateMyProfessors.com 2017. "About RateMyProfessors.com." www.ratemyprofessors.com/about.jsp

Reid, L. D. 2010. "The Role of Perceived Race and Gender in the Evaluation of College Teaching on RateMyProfessors.com." *Journal of Diversity in Higher Education* 3 (3): 137–152.

Richardson, P. S., A. S. Dick, and A. K. Jain. 1994. "Extrinsic and Intrinsic Cue Effects on Perceptions of Store Brand Quality." *Journal of Marketing* 58 (4): 28–36.

Riniolo, T. C., K. C. Johnson, T. R. Sherman, and J. A. Misso. 2006. "Hot or Not: Do Professors Perceived as Physically Attractive Receive Higher Student Evaluations?." *The Journal of General Psychology* 133 (1): 19–35.

Rosen, A. S. 2018. "Correlations, Trends and Potential Biases among Publicly Accessible Web-Based Student Evaluations of Teaching: A Large-Scale Study of RateMyProfessors.com Data." *Assessment & Evaluation in Higher Education* 43 (1): 31–44.

Ross, J. A. 1992. "Teacher Efficacy and the Effect of Coaching on Student Achievement." *Canadian Journal of Education* 17 (1): 51–65.

Safavi, S. A., K. A. Bakar, R. A. Tarmizi, and N. H. Alwi. 2013. "Faculty Perception of Improvements to Instructional Practices in Response to Student Ratings." *Educational Assessment, Evaluation and Accountability* 25 (2): 143–153.

Sandler, B. R., L. Silverberg, and R. Hall. 1996. *The Chilly Classroom Climate: A Guide to Improve the Education of Women*. Washington, DC: National Association for Women in Education.

Silva, K. M., F. J. Silva, M. A. Quinn, J. N. Draper, K. R. Cover, and A. A. Muno. 2008. "Rate My Professor: Online Evaluations of Psychology Instructors." *Teaching of Psychology* 35 (2): 71–80.

Singh, J. P., S. Irani, N. P. Rana, Y. K. Dwivedi, S. Saumya, and P. K. Roy. 2017. "Predicting the Helpfulness of Online Consumer Reviews." *Journal of Business Research* 70: 346–355.

Sonntag, M. E., J. F. Bassett, and T. Snyder. 2009. "An Empirical Test of the Validity of Student Evaluations of Teaching Made on RateMyProfessors.com." *Assessment & Evaluation in Higher Education* 34 (5): 499–504.

Sprinkle, J. E. 2008. "Student Perceptions of Effectiveness: An Examination of the Influence of Student Biases." *College Student Journal* 42 (2): 276–293.

Steffes, E. M., and L. E. Burgee. 2009. "Social Ties and Online Word of Mouth." *Internet Research* 19 (1): 42–59.

Stuber, J. M., A. Watson, A. Carle, and K. Staggs. 2009. "Gender Expectations and on- Line Evaluations of Teaching: Evidence from RateMyProfessors.com." *Teaching in Higher Education* 14 (4): 387–399.

Tatro, C. N. 1995. "Gender Effects on Student Evaluations of Faculty." *Journal of Research & Development in Education* 28 (3): 169–173.

Teas, R. K., and S. Agarwal. 2000. "The Effects of Extrinsic Product Cues on Consumers' Perceptions of Quality, Sacrifice, and Value." *Journal of the Academy of Marketing Science* 28 (2): 278–290.

Timmerman, T. 2008. "On the Validity of RateMyProfessors.com." *Journal of Education for Business* 84 (1): 55–61.

Tschannen-Moran, M., and A. Woolfolk Hoy. 2001. "Teacher Efficacy: Capturing an Elusive Construct." *Teaching and Teacher Education* 17 (7): 783–805.

van Uden, J. M., H. Ritzen, and J. M. Pieters. 2014. "Engaging Students: The Role of Teacher Beliefs and Interpersonal Teacher Behavior in Fostering Student Engagement in Vocational Education." *Teaching and Teacher Education* 37: 21–32.

Wu, J. 2017. "Review Popularity and Review Helpfulness: A Model for User Review Effectiveness." *Decision Support Systems* 97: 92–103.

Yin, D., S. Mitra, and H. Zhang. 2016. "Research Note—When Do Consumers Value Positive vs. Negative Reviews? An Empirical Investigation of Confirmation Bias in Online Word of Mouth." *Information Systems Research* 27 (1): 131–144.