What Rate My Professors Ratings Indicate About Effective Teaching in Economics

Junaid Jahangir

_____

What Rate My Professors ratings indicate about effective teaching in Economics

Abstract: Four questions pertaining to effective teaching are addressed through content analysis of student comments at the Rate My Professors website for Economics instructors at two universities in Edmonton. First, do student comments capture effective teaching? Second, why are Economics instructors rated lower than those in other social sciences? Third, do contract instructors lower teaching standards through easy classes and grade inflation? Finally, are quality instructors penalized for rigorous standards? The analysis suggests that student comments capture many attributes of effective teaching but not critical thinking. There exists some evidence on contract instructors offering easy As just as there exist individual cases of instructors being penalized for maintaining rigorous standards. These findings suggest that administration should focus on student comments instead of average numerical scores. Effective instructors should not be penalized when student comments emphasize easy grades as part of the corporate model of education instead of learning and critical thinking.

What Rate My Professors ratings indicate about effective teaching in Economics

1. **Introduction**

Hollywood's depiction of inspirational instructors includes Mr. Hundert in "The Emperor's Club" and Mr. Keating in "Dead Poet's Society." The former is recognized for emphasis on discipline and transfer of values, whereas, the latter is noted for pushing boundaries and thinking outside the box. However, their methods of calling out a student to enforce discipline or pushing one to think on the spot are not necessarily viewed as effective teaching. Generally, pedagogical methods that consist of asking pupils to think on the spot through pop quizzes, prepare their own notes, attend class on time and write exams based on applications instead of regurgitating information are not favourably rated by students in teaching evaluations. What, then, do students rate as effective teaching?

While the literature on teaching evaluations is vast, the contribution in this paper is to gauge the extent to which teaching effectiveness, as depicted in the literature that includes factors like subject mastery, clarity and active learning methods, can be determined from both qualitative and quantitative analysis of student comments at the Rate My Professors (RMP) website. The focus in this study is on addressing four questions in the context of teaching Economics. First, do student comments capture effective teaching? Second, why are Economics instructors rated lower than those in other social sciences? Third, do contract instructors lower teaching standards through easy classes and grade inflation? Finally, are quality instructors penalized for maintaining rigorous standards?

In terms of data, Universal Student Ratings of Instruction (USRI) are officially used for teaching evaluations but are not publicly available. However, ratings and student comments on the RMP site are readily available and since these ratings are strongly correlated with formal university evaluations (Albrecht and Hoopes, 2009; Timmerman, 2008) and are consistent (Silva et al., 2008) with them, they will be used for the purposes of this study. Beyond the Likert scale responses, the richness of this data, as with the USRI data, lies in the student comments that provide detailed feedback on instructor teaching style and student concerns.

This study emphasizes qualitative content analysis as opposed to Likert scale responses because student comments are often not systematically analyzed even as they offer more nuanced and valuable information on improving teaching effectiveness (Santhanam and Jones, 2018). Student comments reveal what they really feel and think and comments from a large student body reveal issues and preferences common to students through word patterns (Jordan, 2011), which are not necessarily captured by Likert scale responses. Additionally, Likert scale responses can be misleading, as a few disgruntled students can have a large and disproportionate effect on teaching evaluation averages (Kane and Staiger, 2002).

The focus of this study is on Economics because instruction in this field has been consistently ranked amongst the lowest due to various reasons. These include math-based instruction and the fact that Economics is often required of many students from diverse disciplines (Ongeri, 2009). Other reasons for low teaching evaluations in Economics

include the perception of the subject as boring with little real-life application (Ghosh, 2013), fast paced lectures with lots of information (Reimann, 2004), low grades in Economics classes (Cashin, 1990), and instructors emphasizing lecture-based teaching instead of active learning methods (Becker and Watts, 2001). There are also concerns about the simplicity of economic models and lack of discussions on current issues (Andreopoulos and Panayides, 2009) and the observation that Economics has not always valued teaching compared to research (Becker, 1997). In short, both factors within and outside the control of Economics instructors affect their evaluations.

A multitude of studies on teaching evaluations based on the RMP data focus on bias based on ethnicity (Boatright-Horowitz and Soeung, 2009; McPherson and Jewell, 2007; Smith, 2007), accent (Ogier, 2005), gender (Davison and Price, 2009; Schmidt, 2015), and attractiveness (Feeley, 2002; Lawson and Stephenson, 2005; Bonds-Raacke, 2007; Sen, Voia and Woolley, 2010). However, an indepth study of bias due to instructor demographics and attributes is beyond the scope of this paper. The focus in this study is to investigate whether teaching evaluations of Economics instructors are explained by rigorous teaching methods, poor pedagogical methods or dumbing down of instruction. Rigorous in this context does not refer to a teacher centred approach where the student is expected to do all the work but an emphasis on critical thinking, concept-based teaching and exams that test learning instead of the ability to regurgitate information.

This paper is divided into five sections. The next section delves into a select literature review on justifying the use of ratings and comments from the RMP website, the

determinants of teaching effectiveness and issues in teaching evaluations pertaining to student satisfaction, corporatization of education and easy expectations. The third section describes methodology on the data collected and the analytic strategy employed. The fourth section provides content analysis on student comments. The fifth section offers concluding remarks on the findings and results. The results of this study allow for policy prescription for University administration on using teaching evaluations to gauge effective teaching and/or to use alternative methods to recognize effective teaching for annual faculty evaluations, salary raises, promotions and tenure decisions (Algozzine et al., 2004).

2. Literature Review

2.1 Using the Rate My Professors data

One reason for focusing on teaching evaluations from the RMP website is that students as customers rely on peer recommendations on instructor and course selection (Harlow, 2003; Felton, Mitchell and Stinson, 2004). Evaluations at this site may influence student expectations of classes (Kowai-Bell et al., 2011) and perceptions of the instructor (Lewandowski, Higgins and Nardone, 2012). Another reason is the strong correlation between overall quality scores at the RMP website and the USRI teaching evaluation item "overall, how would you rate the instructor" (Coladarci and Kornfield, 2007). Otto and Sanford (2008) even argue that RMP ratings could be a useful supplement to teaching evaluations.

While Legg and Wilson (2012) suggest that students on the RMP website have a negative bias and therefore are not representative of classes, Bleske-Rechek and Michels (2010) indicate that RMP ratings are moderate in tone as opposed to ranting and raving and tend to be more positive than negative. A 2012 survey on formal teaching evaluations, the Universal Student Ratings of Instruction (USRI) system, reveals concerns on abusive comments by students (Sarkonak, 2016). As such, while students who have a strong like or dislike for an instructor are more likely to leave ratings at the RMP website (Sen, Voia and Woolley, 2010) thereby biasing the results, this sample selection bias is also true for formal teaching evaluations. Therefore, using data from the RMP website does not pose an additional problem compared to formal teaching evaluations.

2.2 Determinants of effective teaching

There is no consensus on the definition of effective teaching (Spooren et al., 2013). Based on teaching evaluations, Boex (2000) found that effective teaching includes clarity, presentation skills, the ability to motivate students, and that adopting a demanding stance affects these evaluations. However, Braga et al (2014) conclude that effective teachers require students to exert effort and Becker (2000) suggests avoiding structured tests that fail to challenge students. Other effective teaching methods include the use of instructional support technology (Grimes and Ray, 1993). However, based on 50 years of experience teaching Economics, Hamermesh (2019) indicates that use of technology like clickers has been a pedagogical failure as they disrupt the class flow. Likewise, Ghosh (2013) notes that smart phones, laptops and social media allow students to tune out their instructors.

Effective teaching includes subject mastery (Biggs and Tang, 2011), enthusiasm and approachability (Akerlind, 2007). It also includes motivating by real life examples and current issues (Reimann, 2004), conveying relevance and telling a good story (Hoyt and Imazeki, 2014) and humorous anecdotes and visual illustrations (Dunn and Griggs, 2000). In contrast, poor teaching is based on being disrespectful, having unrealistic expectations, and not being knowledgeable (Busler et al., 2017).

There is consensus in the literature on the effectiveness of active learning, which involves listening, writing, reading and reflecting (Meyers and Jonbes, 1993). It also includes tolerance of alternative viewpoints, emphasis on critical skills (Pozo-Munoz et al., 2000), emphasis on discovering and constructing knowledge (Barr and Tagg, 1995) and collaborative problem solving (Ongeri, 2009). Such interactive teaching is also deemed effective in intermediate and higher-level Economics classes (Hervani and Helms, 2004, 267). Similarly, Bonwell (1992) indicates that students must talk, write and make connections with daily lives, as they do not learn much by memorization. Moreover, students expect to be engaged instead of being passively lectured (Becker, 2000) and giving them notes or PowerPoint slides may result in poor attendance and very little engagement (Sheridan, Hoyt and Imazeki, 2014; Hamermesh, 2019).

However, Andreopoulous and Panayides (2009) found that even the best students, those with a GPA greater than 3.5, liked lecture-based instruction. Similarly, Jordan (2011) stated that student learning-philosophy could be based on memorization and fact

acquisition instead of understanding concepts and interpretation and therefore dependent on passive reception for learning instead of active engagement.

In summary, the literature indicates that effective teaching includes subject mastery, enthusiasm, approachability, clarity, challenge, using technology, nurturing critical thinking, real life examples, humour, story-telling, interaction and active engagement. Alternatively, effective teaching requires that teachers avoid being unapproachable, disrespectful, or have unrealistic expectations. Similarly, students do not value effective teaching when they prefer passive reception of lectures, want organized notes and preparation solely geared towards exams (McKeachie, 1997), do not appreciate challenge, adopt memorization instead of rigorous thinking and avoid active engagement in class.

2.3 Issues in teaching evaluations – student satisfaction, corporatization and easy expectations

There is concern on whether students can evaluate instruction quality, as they are neither trained observers nor privy to instructor pedagogy (Braskamp et al., 1981). According to Becker (2000) there is low correlation between student evaluations and other measures of teaching effectiveness including tests scores and alumni surveys. However, Kulik (2001) indicates that studies show that student ratings are consistent with both such measures. Similarly, Lattuca and Domagal-Goldman (2007) mention that a considerable body of research on teaching evaluations finds that students are good judges of clarity,

preparation and organization but not content. Likewise, Feldman (2007) and Pan et al. (2009) repudiate claims that students lack maturity to evaluate instruction quality. Indeed, according to Marsh (2007), teaching evaluations are helpful in improving instruction quality.

However, it has been noted that teaching evaluations may simply reflect students' particular experience or disposition (MacFadyen, Dawson, Prest and Gasevic, 2016) and primarily capture student satisfaction (Abrami et al., 2007; Beecham, 2009). This is consistent with the observation that teaching evaluations are biased due to student prior interest, expected grades and perceived workload (Marsh, 1987). The literature indicates that while faculty is focused on learning, students may simply care for grades (Hornstein 2017) that are significant determinants of teaching evaluations (Millea and Grimes, 2002; Weinberg et al., 2009). However, Feldman (2007) notes that students who learn more and receive higher grades may give higher instructor ratings.

Generally, to the extent that administration wants to retain students, associates teaching effectiveness with higher teaching evaluations and exclusively relies on them, effective teaching may be reduced to satisfying students. However, Beckman and Stirling (2005) caution that student satisfaction does not necessarily imply learning. There is some evidence that increasingly students with modest academic profiles have been admitted (Reimann, 2004) and such students, who use surface approach to education including rote learning and memorization, use the logic that education is a commodity to be bought (Biggs and Tang, 2011, p. 4-6). However, Sproule (2000) indicates that despite administration

attempts to corporatize the University, the instructor student relationship is not analogous to the customer-client covenant. This corporatization of higher education in many countries by a shift towards a business-oriented model of operation has been well noted (Marginson and Considine, 2000; Mazzarol, Soutar and Seng, 2003).

In a corporate model, faculty members, especially contract instructors or those without tenure, who are concerned about potential contract renewal or promotion, would have the incentive and/or pressure to give easier exams, contribute to grade inflation and generally "dumb down" instruction material. Hornstein (2017) notes that teaching evaluations pressurize faculty members to not rock the boat and to not push undergraduate students to maximize their intellectual potential. Therefore, given pressure to win student approval through an easy A, academic standards may have fallen (Felton, Mitchell and Stinson, 2004).

There are concerns that high teaching evaluations may not be capturing teaching effectiveness but rather dumbing down course material (Becker, 2000), personality and entertainment skills (Jordan, 2011), easy grading and light course load (Greenwald and Gillmore, 1997) and popularity (Aleamoni, 1987; Coleman and McKeachie, 1981). However, Feldman (2007) and Pan et al. (2009) suggest that it is untrue that teaching evaluations are popularity contests. Likewise, Marsh and Ware (1982) indicate that entertaining instructors do not necessarily receive higher ratings. Major reviews have consistently shown that teaching evaluations are unaffected by grading leniency and

workload, as students learn through challenge and commitment and devalue learning if success is easy due to low workload (Marsh and Roche, 2000).

According to Theyson (2015), the easiness and quality relationship may be complex as extremely hard and extremely easy may be viewed as low quality. Therefore, instructors cannot get higher evaluations by offering higher grades and easier courses (Marsh and Roche, 2000). However, positive correlations are found between instructor quality and easiness (Rosen, 2017) and instructors perceived as easier receive higher evaluation scores (Cashin, 1995; McKeachie, 1997), indicating student preference for easy classes (Miller, 2006). Specifically, in the case of instructors with low ratings, students indicate concerns on workload, grade distribution and teaching practices, whereas, majority of the instructors rated highly are described as "nice," "easy," "cool," "caring," "laid back," and "understanding," which suggests that they are liked for their easy expectations apart from personalities (Felton, Mitchell and Stinson, 2004). However, it is possible that easiness is capturing instructional quality, as Otto, Sanford and Ross (2008) suggest that easiness could be interpreted as "easy to understand" instead of "not challenging." Specifically, it is found that helpfulness and clarity are positively related to easiness (Reid, 2010). Alternatively, there is ambiguity in whether better teachers are perceived as easier or whether easier teachers are perceived as better teachers (Rosen, 2017).

In summary, there is evidence that teaching evaluations are driven by student experience, satisfaction, prior interest, expected or received grades and perceived workload. However, there is disagreement in the literature if such evaluations are

rewarding easy expectations and entertainment skills or whether better instructors are perceived as easier, which reflects an association between easiness, higher grades and therefore higher evaluations. This necessitates an investigation into student comments as to whether they are capturing teaching effectiveness, as depicted in the literature that includes factors like subject mastery, clarity and active learning methods, or other factors including easy expectations and easy grades.

3. Data and Methods

3.1 Data

Macewan University in Edmonton provides an excellent study to qualitatively and to some extent quantitatively determine what constitutes effective teaching in Economics and why Economics faculty members are rated consistently lower than their counterparts in other social sciences. The reason is because Economics is part of the department of Anthropology, Economics and Political Science (AEPS), where the same broad governance structure, including common department meetings, deadlines and general expectations, allows for comparison of teaching evaluations across the three disciplines. There are six full-time and a few contract Economics faculty members at MacEwan. While a qualitative study focus on such smaller samples makes generalization problematic, it allows for in-depth understanding of an issue (Bengtsson, 2016). Moreover, since the MacEwan Economics program is similar to large Canadian universities like the University of Alberta, results should be broadly generalizable. A similar argument for generalizability has been made by McPherson, Jewell and Kim (2009) for the University of North Texas.

Apart from MacEwan Economics instructors, this study also includes information on University of Alberta Economics instructors. The justification to do so is based on the fact that some faculty at MacEwan have been trained at the University of Alberta. Additionally, many students have historically transferred course credits from one institution to the other. Economics courses at MacEwan University, a teaching-intensive institution, are parallel to those taught at the University of Alberta, a research-intensive institution, where both tenure track and contract-based teaching faculty teach Economics students. This facilitates comparison of teaching evaluations between tenure track faculty and contract-based lecturers and allows investigating whether the latter respond to teaching evaluations pressures through easy classes and grade inflation. The two institutions, however, differ due to larger class sizes and relatively stringent entrance requirements at the University of Alberta.

The RMP website provides ratings on 83 instructors of six distinct categories. Enumerated in Table 1, they include instructors in Anthropology, Political Science and Economics at MacEwan University and full time, retired and contract Economics instructors at the University of Alberta in Edmonton. The website provides instructor evaluations based on six student level variables including instructor difficulty level (on a 1-5 scale), percentage of students that would take another class with the instructor, the course level, textbook use, attendance being mandatory, and grade received. Instructors are rated on a $1 - 5$ scale, which is associated with three emoticons - green for good $(3.5 - 4.0)$ and awesome $(4.5 - 5.0)$, yellow for average $(2.5 - 3.4)$ and red for poor $(1.5 - 2.4)$ and awful (1.0). Data points for these variables were collected for approximately a five-year

period from 2015 – 2019 and constituted about 1,702 observations. Data from 2000 onwards is not used to avoid any adverse impact of COVID-19 on the ratings.

Data on instructor demographics including ethnicity, accent and gender is based on the author's knowledge based on working at both institutions. Data on percentage of white instructors, native English speakers and male instructors along with the average instructor quality in each of the six instructor categories are summarized in Table 1. This Table also includes data on the six student/course level variables including the percentage of students that would take the course again with the instructor, students in 100 level courses, students who deemed attendance mandatory and students who used the text, along with the average instructor difficulty level and the average grade received for each of the six instructor categories. Table 1 also provides this data by distinguishing MacEwan instructors between continuing and contract instructors. A preliminary analysis based on this data on instructor demographics and student/course level variables is presented in Section 4.1.

3.2 Coding

In addition to Likert scale responses, students can leave comments, which can be subjected to qualitative content analysis. Specifically, content analysis allows making inferences from written data to describe and quantify specific phenomena (Downe-Wambolt, 1992). It allows reducing volume of texts in a systematic fashion by facilitating the grouping of categories and themes through the process of coding. This form of analysis comprises of both a qualitative and a quantitative methodology, where in the case of the latter, textual facts are expressed as percentage of key categories (Bengtsson, 2016). The

idea is to convert written data into quantitative form for systematic drawing out of categories and themes.

For qualitative content analysis, 1,702 comments that correspond with each of the 1,702 data observations were extracted and subjected to coding that yielded 5,511 words. This qualitative data is subjected to difference of means tests for subsequent analyses. Table 1 summarizes the number of comments and words coded for each of the six instructor categories and by distinguishing MacEwan instructors between continuing and contract instructors. Table 2 provides the percentage of keywords/themes sorted into eighteen key coding categories. These themes were guided by their preponderance in student comments and based on the literature review that identified attributes of effective teaching, factors that influence teaching evaluations other than teaching effectiveness and issues identified specifically in the context of Economics instruction.

The effective teaching attributes identified through the literature review were summarized in the section 'determinants of effective teaching'. They allow for coding categories that include 'mastery', 'interaction/interest/passion,' 'critical skills/learning,' 'real world relevance,' 'fun/story,' 'tech' for technology use, 'clarity/organization,' 'personality/helpful/available' and 'tough/grading/exams'. The converse coding categories of 'rude/personality' and 'boring/slide reader' were also used. The factors that influence teaching evaluations include the categories of 'Easy A/pattern/fair' for easy expectations and 'unclear/tangents/accent'. The issues in the context of Economics instruction offered the coding categories of 'fast pace' and 'math'. Other coding categories based on their

preponderance in the student comments and on the observation by McKeachie (1997) that students want organized notes include 'notes given/skip/no text,' its converse 'attend/take notes/work hard' and 'text expensive/bad'.

To illustrate the coding process, consider the following comment.

"i would say participation is key for his class, his class is enjoyable he is very funny and engages the class in discussion so you will not be bored. Very good prof, explains everything very clearly with examples easy to follow and remember."

This comment indicates that the student found the class fun, that it involved engagement through class participation, that it was not boring and that the instructor explains clearly. The keywords/themes in this comment that stand out are "funny," "clear," "CP matters," "enagages," and "enjoyable." This allowed coding this statement in the broad categories of 'Fun/Story,' 'Clarity/Organization,' and 'Interaction/Interest/Passion.'

As another example, consider the following statement, which indicates that the student found the instructor confusing, difficult to follow, that the instructor gave difficult exams but was willing to help.

"Lectures were really confusing. He did examples but they were really hard to follow. The tests were hard and I ended up doing really bad on them. He was always willing to answer questions though."

Based on this comment, the keywords/themes that stand out are "confusing," "difficult exams," and "helpful." This allowed coding this statement in the broad categories of 'Unclear/Tangents/Accent,' 'Tough/Grading/Exams,' and 'Personality/Helpful/ Available.'

Such coding is not without judgment, which can be illustrated through the coding category of 'Easy A/Pattern/Fair.' Generally, this category is predominately about easy expectations based on words and themes like 'easy A,' 'GPA booster,' 'helps succeed,' and the fact that exams were based on lectures notes and sample exams, that the instructor offered hints, gave away exams, shifted the curve, offered step by step chart for exam preparations or gave straight forward exams. While it could be argued that the category 'fair' could be construed as tough instead of easy, or that 'helps succeed,' and step by step chart for exam preparations could be reflective of outstanding teaching or useful learning approach, the student comments did not allow for such alternate coding. Consider, for instance, the following comments that associate study charts, 'helps succeed,' fair exams and fair grading with easy expectations or regurgitation-based exams instead of rigorous assessment that tests for concepts and applications.

"He gives out a study guide for each exam, use it!! Pretty much everything that is on that study sheet is on the exam. You should be able to get at least an A-."

"he really wants his students to succeed. Exams are based on the guidelines he hands out, ... Minimal effort is required."

"Exams were rather straightforward and is a fair marker, an A is easily attainable."

"Such an awesome prof. No surprises on the exam, if you do the work and do his examples you will know exactly what is on exam."

It is such value judgment that guided the coding for all the 1,702 student comments into the 18 categories. Once the 5,511 words were sorted in these categories, the percentage of keywords/themes in each category was computed to determine the relative importance of that category in student comments. For instance, there are 887 words out of a total of 5,511 words (16.1%) in the 'Easy A/Pattern/Fair' category, but only 201 (3.65%) words in the 'Critical skills/Learning' catgeory. This facilitated the use of difference in means method, which helped answer the four questions in this paper as follows.

4. Qualitative analysis of the Rate My Professors data

4.1 Preliminary Data Analysis

Table 1 indicates that MacEwan Economics faculty members receive the lowest quality ratings from students. It suggests that this may be due to four factors – the lowest average grade received by students (3.43/4), ratings by predominantly students at the first-year level (80.23%) and the highest percentage of ethnic and accented faculty members (72.73%). Table 1 also indicates that the difficulty rating of retired Economics faculty members at the University of Alberta is the highest (3.32/5). Additionally, class attendance

and the willingness to take a class again with the instructor is much higher for Anthropology and Political Science instructors compared to those in Economics.

Distinguishing MacEwan faculty between continuing and contract instructors, Table 1 indicates that the difficulty level of contract instructors, whether at MacEwan (2.48/5) or at the University of Alberta (2.69/5), is the lowest and the quality rating is higher compared to continuing faculty members in Economics (3.69 versus 3.26 for MacEwan and 3.95 versus 3.50 for the University of Alberta). Additionally, the average grade received by students is also higher for contract instructors compared to continuing instructors in Economics (3.55 versus 3.41 for MacEwan and 3.72 versus 3.65 for the University of Alberta). This suggests either that contract instructors have lowered academic standards or that their exposition is clearer compared to continuing Economics instructors.

In summary, the preliminary analysis shows that low ratings of Economics instructors are driven by first year level students, several of whom are required to take the class, who have received lower grades and who are penalizing ethnic instructors that speak with an accent. It also provides potential evidence for contract instructors lowering teaching standards through easy classes and grade inflation since the difficulty levels are lower and both instructor quality and grades received by students are higher for these instructors. However, a detailed analysis of bias in student evaluations based on instructor demographics is beyond the scope of this paper and the focus in the following is retained on answering the four questions raised in the introduction.

4.2 Do student comments capture teaching effectiveness?

Content analysis allows addressing whether student comments capture effective teaching. Specifically, Table 2 shows that the largest coding categories comprise of the 'Easy A/Pattern/Fair' and the 'Attend/Take notes/Work hard' categories (16% - 17% compared to 0.5% - 9% for other categories). These categories capture students' overwhelming concern with easy As, routine sample based easy exams, easy grades, attending class, paying attention, taking notes and working hard. In contrast, the categories of 'Critical Skills/Learning' and subject 'Mastery,' which capture challenge, concepts instead of memorization, debates, in depth analysis, critical thinking, instructor command of the subject and brilliance, are quite small categories (2% - 4%). Interestingly, real world relevance (1.42%) and math content (0.54%) constitute very small categories. The latter indicates that any perceived toughness of Economics instructors is not necessarily related to the quantitative aspect of the subject compared to other social sciences.

This analysis can be complemented by a difference of means test based on the percentage of keywords between the eighteen categories for highly rated instructors (rated above 3.5) and average or poorly rated instructors (rated below 3.5). Table 3 shows the results of a difference of means test and indicates that significant differences between highly and poorly rated instructors arise from categories that comprise of great personality and helpfulness, disagreeable personality, clarity and organization, instructor accent and unclear exposition, engagement and passion, fun and stories, boring instructor and real-world relevance. Interestingly, the largest categories of easy As, workload and tough exams

remain insignificant. Additionally, provision of notes is not a significant category for both highly rated and poorly rated instructors provide notes.

Indeed, both highly rated and poorly rated instructors are noted to provide notes as illustrated through the following comments respectively. Both comments also indicate a preoccupation with easy grades or attending class.

"... superb Economics Professor. Do not buy the textbook his notes are sufficient and clear. You can easily pass his tests with his notes alone and his tips during the lecture so show up ..."

"notes are very straight forward and dry at times, hands out usually 1 work sheet per chapter (not nearly enough). extremely test heavy ... would suggest showing up as reading the notes on your own can be tough."

More generally, the following comments on a highly rated and poorly rated instructor respectively indicate that while student preoccupation remains with easy expectations and grades, comments do capture many facets of effective teaching as highlighted in the literature, such as, instructor enthusiasm and clarity. The second comment on the poorly rated instructor confirms that instructors cannot obtain higher ratings simply through easy expectations if the quality of engaging students is missing.

"He is very enthusiastic about teaching the material and he had examples that were very easy to understand. He has powerpoints although he bases most of his exam questions from his lectures so if you absolutely have to miss a class, ask someone for the notes. ... Pay attention and you will do well. Amazing prof."

"I j-walk on the way to this class because getting taken out by a bus would be more enjoyable than sitting through another lecture . tests are easy though"

In summary, these findings indicate that while students are predominately concerned with easy As, workload and tough exams, their rating of instructors includes several of the attributes of effective teaching. However, student ratings do not effectively emphasize challenge and critical thinking. Overall, this suggests that achieving high instructor ratings requires clarity, real life examples, humour, storytelling, interaction and engagement even as some students do not appreciate the challenge of rigorous thinking. Alternatively, avoiding lower ratings means not being disrespectful, unhelpful, unclear, disorganized and dull in class. Moreover, the result that the categories of easy As, workload and tough exams remain insignificant, as they are similar across highly rated and poorly rated instructors confirms the findings in the literature, specifically Marsh and Roche (2000), that instructors cannot get higher evaluations by just offering higher grades and easier courses.

4.3 Why are Economics instructors rated lower than those in other social sciences?

The second question is about determining reasons for lower ratings for Economics instructors compared to those in relatively less quantitative social sciences at MacEwan. Based on difference of means tests, Table 4 indicates that Economics instructors receive significantly lower comments for engagement and passion (4.57% vs. 8.78%) and significantly higher comments on toughness (12% vs. 5.68%). At 10% significance, Economics instructors also receive significantly higher comments on unclear exposition and accent (15.24% vs. 8.96%). Additionally, while they receive significantly lower percentage of comments in the critical skills and learning category and significantly higher comments on real world relevance, these are small coding categories.

As an illustration, note the contrast between the following two student comments for an Economics and Political Science instructor respectively, where the distinction arises on the basis of engagement and toughness.

"Tests are very hard because they will include information that didn't really seem relevant to the course and wasn't. You don't really need to show up to class because all he does is read off of the slides, and rarely goes through examples."

"... makes lectures fun and engaging. He draws mind maps which are really helpful if youre a visual learner, and he emphasizes key terms that you need to know later. Hes a great prof, would recommend."

Overall, the analysis suggests that lower Economics ratings are not due to quantitative methods but rather due to Economics instructors being perceived as less interactive and tougher due to hard exams or high expectations. The concern on being less interactive confirms the analysis of Becker and Watts (2001) that students rate Economics instructors relatively lower because they do not use active learning methods and instead emphasize lecture-based teaching.

4.4 Do contract instructors lower teaching standards?

The third question is about investigating the lowering of teaching standards by contract instructors. Based on difference of means tests, Table 4 distinguishes between continuing and contract instructors at MacEwan and between retired and contract instructors at the University of Alberta. The comparisons between full time and contract instructors at the University of Alberta are not significantly different. Table 4 indicates that contract instructors receive higher percentage of comments in the Easy A category at MacEwan (21.07% vs. 14.95%) and significantly higher at the University of Alberta (15.92% vs. 8.49%). This may be explained by lowering of academic standards or better clarity. However, the categories of both clarity and unclear exposition are insignificant.

Compared to contract instructors, continuing instructors receive a significantly higher percentage of comments on subject mastery at MacEwan, although this is a small category. Likewise, retired instructors receive significantly larger percentage of comments on subject mastery (8.05% vs. 1.11%), critical skills/learning (8.28% vs. 3.8%) and significantly lower percentage of comments on easy As (8.49% vs. 15.92%) compared to

contract instructors at the University of Alberta. The fact that subject mastery and critical skills/learning are both insignificant categories when distinguishing between highly rated and poorly rated instructors in Table 3, but are significantly higher for continuing/retired instructors, suggests that to the extent contract instructors offer easy As and to the extent subject mastery, critical skills and learning are ignored in such ratings, lower academic standards can be associated with contract instructors.

As an illustration, consider the following two student comments for a retired and a contract Economics instructor respectively. The comments contrast the difference between subject mastery and emphasis on learning in the case of the former and easy expectations and grades in the case of the latter.

"He knows what he is talking about. His assignments/exams are VERY tough. Most people hate him, and class is non-stop notetaking. He can be cruel/disrespectful to students, but he seems nice outside of class. Forces you to learn. Recommended if you like challenge."

"If you take his class you won't need a textbook because his notes and practice exams provide more than enough material. Don't expect to learn much because he doesnt take ... in depth as other profs making it the easiest course ... Good gpa booster"

In summary, there exists some evidence on lower academic standards associated with contract instructors through easy As, a finding that gets masked in the analysis when

overall highly rated and poorly rated instructors are compared, in which case the category of easy As remains similar. The finding that retired Economics instructors are rated lower despite being rated for subject mastery and critical skills, along with the finding that student ratings do not effectively emphasize challenge and critical thinking, suggests that to the extent student ratings penalize instructors for maintaining rigorous standards and award higher ratings for easy As, student learning philosophy may not be consistent with effective teaching practices. However, the finding on lower ratings due to less interaction, as noted in Section 4.3, indicates that Economics instructors may have to replace lecture-based teaching with active learning methods. In essence, changes in both student philosophy on learning and instructor emphasis on the lecture method are warranted.

4.5 Are quality instructors penalized for maintaining rigorous standards?

The fourth question is about investigating whether quality instructors are penalized for maintaining rigorous standards. Rigorous in this context does not mean using a teacher centred approach where the student is expected to do all the work but an emphasis on critical thinking, concept-based teaching and giving exams that test learning instead of the ability to regurgitate information. In Table 5, the focus is on six Economics instructors at MacEwan who are rated adequately well or very highly. They are distinguished between those rated more than 4.5 (instructors, I1 and I2) and those below 4.5 (instructors, I3 – I6). It is discovered that very highly rated instructors, I1 and I2 (rated 4.9/5 and 4.7/5) receive very high percentage of comments in the easy A category (46.15% and 32.86% compared to the average of 18.06%) and 0% on being tough. Compared to instructors I1 and I2, instructor I3, while being rated relatively highly on interaction and passion, clarity and

26

organization humour and story-telling, critical skills and learning, real world relevance and subject mastery, all categories identified for effective teaching, is rated relatively lower at 4.1/5. This is potentially explained by I3 receiving much lower percentage of comments on easy As (6.49%) and much higher on toughness (8.11%). It is therefore not surprising that I3 received much lower percentage of ratings on personality and helpfulness and relatively more on accent and unclear exposition. The key student comments also substantiate that I1 and I2 are noted for easy expectations, whereas I3 is described as requiring hard work. This indicates that sometimes effective instructors are penalized for maintaining rigorous academic standards and very highly rated instructors are rewarded for giving out easy As.

As an illustration, note the contrast between the following student comments on instructor I3 and I1 respectively, where the former is noted for pushing critical thinking and challenge and the latter is noted for easy expectations and high grades.

"... tough, but he wants to get you thinking. He doesn't teach you for the intention of letting you get an easy A, he wants to challenge you. His lectures are very enjoyable and definitely necessary if you want to pass. His exams are passable if you show you really put the effort in. He's willing to go slow to make sure everyone understands."

"... practically econ for dummies. easiest class to get a 4.0"

However, generalization based on a single example is not warranted. Table 5 also provides difference of means tests for various coding categories for two groups of

instructors, those rated above 4.5 and those rated relatively lower between 3 and 4.5 but who nonetheless are still identified as bearing the attributes of effective teaching. This analysis includes all categories of AEPS instructors at MacEwan and all categories of Economics instructors at the University of Alberta. Seventeen instructors fall in the former category and fifty-one in the latter category. It is found that very highly rated instructors (4.5 and above) receive significantly higher percentage of comments on easy As and significantly lower on toughness. This can be explained by their receiving significantly higher percentage of comments on clarity and organization, interaction and passion, humour and story-telling and significantly lower percentage of comments on disagreeable personality, unclear exposition and accent and being boring.

In summary, the above analysis suggests that while individual cases of instructors being penalized for maintaining rigorous standards exist, such cases get masked in over all data analysis based on averages. This substantiates looking at detailed comments instead of numerical averages from teaching evaluations. Moreover, achieving high instructor ratings requires a composite of several factors, over and above subject mastery, including clarity and organization, interaction and passion, and humour and story-telling, a result that substantiates the findings from Section 4.2. In other words, while student ratings capture several attributes of teaching effectiveness, as identified in the literature review, it is also important to note that there is no single criterion for effective teaching.

5. Conclusions

The objective in this paper was to address four questions pertaining to teaching effectiveness through qualitative content analysis, which is warranted, as student comments highlight issues that are not necessarily captured by Likert scale responses. Content analysis revealed that student comments capture many attributes of effective teaching and that achieving high instructor ratings requires clarity, real life examples, humour, storytelling, interaction and engagement. However, these comments do not effectively emphasize challenge and critical thinking. The result that the categories of easy As, workload and tough exams are similar across highly rated and poorly rated instructors confirms the conclusions of Marsh and Roche (2000) that instructors cannot get higher evaluations by just offering higher grades and easier courses.

For the second question, content analysis provides evidence that low Economics ratings compared to other social sciences are not due to quantitative methods but due to Economics instructors being perceived as less interactive and tougher. This confirms the observation by Becker and Watts (2001) that Economics instructors do not use active learning methods and suggests that they may have to replace or complement lecture-based teaching with active learning methods.

For the third question, content analysis provides evidence on lower academic standards associated with contract instructors through easy As and especially as subject mastery, critical skills and learning are ignored in student comments. This finding gets

masked when overall highly rated and poorly rated instructors are compared, which shows that the category of easy As is similar across both groups of instructors. The results also suggest that to the extent student ratings penalize instructors for maintaining rigorous standards and pushing critical thinking, student learning philosophy may not be consistent with effective teaching practices.

For the final question, the analysis suggests that while individual cases of instructors being penalized for maintaining rigorous standards exist, such cases get masked in over all data analysis based on averages. This substantiates looking at detailed comments instead of numerical averages from teaching evaluations.

Overall, the analysis indicates that there is no single criterion for effective teaching. It indicates that instructors cannot get higher ratings by just entertainment skills or easy expectations but require clarity, enthusiasm and active learning methods. The analysis also indicates that student-learning philosophy may not affirm effective teaching to the extent that student ratings discount critical thinking and challenge. Additionally, the analysis suggests that teaching effectiveness should not be reduced to numerical averages that mask information and which allows the penalization of effective instructors to go unnoticed. The analysis suggests that for policy, administration should focus qualitatively on student comments and distinguish why they are high or low. It should be ascertained whether student comments emphasize learning (as opposed to easy grades) and critical thinking and learning (as opposed to easy expectations).

This recommendation of carefully looking at student comments confirms the literature, which indicates that teaching evaluations should be cautiously used for promotion and tenure decisions (Spooren et al., 2013; Beecham, 2009; Stark and Freishtat, 2014; Braga et al., 2014). An alternative to teaching evaluations would be the use of reflective instructor statements, three years of syllabi and evidence of improving teaching activities (Seldin, 1999), as no single source of evidence can reasonably evaluate teaching effectiveness (Stark and Freishtat, 2014).

References

Abrami, P. C., d'Apollonia, S. and Rosenfield, S. (2007) 'The dimensionality of student ratings of instruction: What we know and what we do not', in Perry, R. P. and Smart, J. C. (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective*, Springer, New York, pp. 385–456.

Akerlind, G. S. (2007) 'Constraints on academics' potential for developing as a teacher', *Studies in Higher Education,* Vol. 32 No. 1, pp. 21-37.

Albrecht, S. and Hoopes, J. (2009) 'An empirical assessment of commercial web-based professor evaluation services', *Journal of Accounting Education,* Vol. 27 No. 3, pp. 125-132.

Aleamoni, L. (1987) 'Student rating myths versus research facts', *Journal of Personnel Evaluation in Education,* Vol. 1 No. 1, pp. 111-119.

Algozzine, B., Gretes, J., Flowers, C., Howley, L., Beattie, J., Spooner, F., Mohanty, G. and Bray. M. (2004) 'Student Evaluation of College Teaching: A Practice in Search of Principles', *College Teaching,* Vol. 52 No. 4, pp. 134−141.

Anderson, H. and Siegfried, J.J. (1997) 'Gender Differences in Rating the Teaching of Economics', *Eastern Economic Journal* Vol. 23 No. 3, pp. 347-357.

Andreopoulos, G.C. and Panayides, A. (2009) 'Teaching Economics to the best undergraduates: What are the problems?' *American Journal of Business Education,* Vol. 2 No. 6, pp. 117-121.

Baek, S.G and Shin, H.J. (2008) 'Multilevel analysis of the effects of student and course characteristics on satisfaction in undergraduate liberal arts courses', *Asia Pacific Education Review,* Vol. 9 No. 4, pp. 475-486.

Barr, R.B. and Tagg, J. (1995) 'From teaching to learning: A new paradigm for undergraduate education', *Change,* Vol. 27 (November/December), pp. 13-25.

Becker, W. E. (1997) 'Teaching economics to undergraduates', *Journal of Economic Literature,* Vol. 35 No. 3, pp. 1347-1373.

Becker, W.E. (2000) 'Teaching Economics in the 21st century', *The Journal of Economics Perspectives,* Vol. 14 No. 1, pp. 109-119.

Becker, W. E. and Watts, M. (2001) 'Teaching Methods in U.S. Undergraduate Economics Courses', *The Journal of Economic Education,* Vol. 32 No. 3, pp. 269-279.

Beckman, M. and Stirling, K. (2005) 'Promoting critical thinking in Economics Education', Working Paper.

Beecham, R. (2009) 'Teaching quality and student satisfaction: Nexus or simulacrum?' *London Review of Education,* Vol. 7, pp. 135–146.

Bengtsson, M. (2016) 'How to plan and perform a qualitative study using content analysis', *NursingPlus Open,* Vol. 2, pp. 8-14.

Biggs, J. and Tang, C. (2011) *Teaching for quality learning at university*, 4th edition, Open University Press.

Bleske-Rechek, A. and Michels, K. (2010) 'RateMyProfessors.com: Testing Assumptions about Student Use and Misuse,' *Practical Assessment, Research & Evaluation,* Vol. 15 No. 5, pp. 1-12.

Boatright-Horowitz, S. and Soeung, S. (2009) 'Teaching white privilege to white students can mean saying good-bye to positive student evaluations', *American Psychologist,* Vol. 64 No. 6, pp. 574–575.

Boex, J.F.L. (2000) 'Attributes of effective Economics instructors: An analysis of student evaluation'*, Journal of Economic Education,* Vol. 31 No. 3, pp. 211-227.

Bonds-Raacke, J. and Raacke, J. D. (2007) 'The Relationship between Physical Attractiveness of Professors and Students' Ratings of Professor Quality', *Journal of Psychiatry, Psychology and Mental Health,* Vol. 1 No. 2, pp. 1-7.

Bonwell, C.C. (1992) 'Risky business: Making active learning a reality. Teaching Excellence', *POD Network in Higher Education*, 1992/93 Essay Series, pp. 1-3.

Braga, M., Paccagnella, M. and Pellizzari, M. (2014) 'Evaluating students' evaluations of professors', *Economics of Education Review,* Vol. 41, pp. 71–88.

Braskamp, L., Ory, J. and Pieper, D. (1981) 'Student written comments: Dimensions of instructional quality', *Journal of Educational Psychology,* Vol. 73 No. 1, pp. 65-70.

Busler, J., Kirk, C., Keeley, J. and Buskist, W. (2017) 'What constitutes poor teaching? A preliminary inquiry into the misbehaviours of not-so-good instructors', *Teaching of Psychology,* Vol. 44 No. 4, pp. 330-334.

Cashin, W. (1990) 'Students do rate different academic fields differently', in Theall, M. and Franklin, J. (Eds.), Student ratings of instruction: Issues for improving practice, New Directions for Teaching and Learning 43 (Fall), Jossey-Bass, San Francisco, pp. 113-21.

Coladarci, T. and Kornfield, I. (2007) 'RateMyProfessors.com versus formal in-class student evaluations of teaching,' *Practical Assessment, Research & Evaluation,* Vol. 12 No. 6, pp. 1- 15.

Coleman, J. and McKeachie, W.J. (1981) 'Effects of instructor/course evaluations on student course selection', *Journal of Educational Psychology,* Vol. 73 No. 2, pp. 224-26.

Davison, E. and Price, J. (2009) 'How do we rate? An evaluation of online student evaluations,' *Assessment & Evaluation in Higher Education,* Vol. 34 No. 1, pp. 51-65.

Downe-Wamblot, B. (1992) 'Content analysis: method, applications and issues', *Health Care for Women International,* Vol. 13, pp. 313-321.

Dunn, R. and Griggs, S.A. (2000) *Practical Approaches to Using Learning Styles in Higher Education*, Bergin & Garvey, Westport, CT.

Feeley, T. H. (2002) 'Evidence of halo effects in student evaluations of communication instruction', *Communication Education,* Vol. 51 No. 3, pp. 225–236.

Feldman, K. A. (2007) 'Identifying exemplary teachers and teaching: Evidence from student ratings,' in Perry, R. P. and Smart, J. C. (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective,* Springer, New York, pp. 93–143.

Felton, J., Mitchell, J. and Stinson, M. (2004) 'Web-based student evaluations of professors: the relations between perceived quality, easiness and sexiness', *Assessment & Evaluation in in Higher Education,* Vol. 29 No. 1, pp. 91-108.

Ghosh, I.K. (2013) 'Learning by doing models to teach undergraduate Economics', *Journal of Economics and Economic Education Research,* Vol. 14 No. 1, pp. 105-119.

Greenwald, A. G. and Gillmore, G.M. (1997) 'Grading leniency is a removable contaminant of student ratings', *American Psychologist,* Vol. 52 No. 11, pp. 1209–1217.

Grimes, P. W. and Ray, M.A. (1993) 'Economics: Microcomputers in the College Classroom: A Review of the Academic Literature', *Social Science Computer Review,* Vol. 11, pp. 452-63.

Harlow, R. (2003) '"Race doesn't matter, but …": The effect of race on professors' experiences and emotion management in the undergraduate college classroom', *Social Psychology Quarterly,* Vol. 66 No. 4, pp. 348–363.

Hamermesh, D.S. (2019) '50 years of teaching introductory Economics', *The Journal of Economic Education*, Vol. 50 No. 3, pp. 273-283.

Hamermesh, D., and Parker, A. (2005) 'Beauty in the Classroom: Instructors' Pulchritude and Putative Pedagogical Productivity', *Economics of Education Review,* Vol. 24 No. 4, pp. 369-376.

Hervani, A., and Helms, M.M. (2004) 'Increasing Creativity in Economics: The Service Learning Project', *Journal of Education for Business,* Vol. 79 No. 5, pp. 267-274.

Hornstein, H.A. (2017) 'Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance', *Cogent Education,* Vol. 4 No. 1, 1304016.

Jordan, D.W. (2011) 'Re-thinking student written comments in course evaluations: Text mining unstructured data for program and institutional assessment', Doctoral dissertation. California State University.

Kane, T. J. and Staiger, D.O. (2002) 'The promise and pitfalls of using imprecise school accountability measures', *The Journal of Economic Perspectives,* Vol. 16, pp. 91–114.

Kowai-Bell, N., Guadagno, R. E., Little, T., Preiss, N. and Hensley, R. (2011) 'Rate My Expectations: How online evaluations of professors impact students' perceived control', *Computers in Human Behavior,* Vol. 27 No. 5, pp. 1862-1867.

Kulik, J.A. (2001) 'Student ratings: validity, utility, and controversy', *New Directions for Institutional Research,* 109, pp. 9-25.

Lattuca, L.R. and Domagal-Goldman, J.M. (2007) 'Using qualitative methods to access teaching effectiveness', *New Directions for Institutional Research,* Vol. 136, pp. 81-93.

Lawson, R. A. and Stephenson, E. F. (2005) 'Easiness, Attractiveness, and Faculty Evaluations: Evidence from RateMyProfessors.com', *Atlantic Economic Journal,* Vol. 33 No. 4, pp. 485–486.

Legg, A.M. and Wilson, J.H. (2012) 'RateMyProfessors.com Offers Biased Evaluations', *Assessment & Evaluation in Higher Education,* Vol. 37 No. 1, pp. 89-97.

Lewandowski, G. W., Higgins, E. and Nardone, N.N. (2012) 'Just a harmless website?: An experimental examination of RateMyProfessors.com's effect on student evaluations', *Assessment & Evaluation in Higher Education,* Vol. 37 No. 8, pp. 987-1002.

Macfadyen, L.P., Dawson, S., Prest, S. and Gasevic, D. (2016) 'Whose feedback? A multilevel analysis of student completion of end of term teaching evaluations', *Assessment and Evaluation in Higher Learning,* Vol. 41 No. 6, pp. 821-839.

Marginson, S., and Considine, M. (2000) *The enterprise university: Power, governance, and reinvention in Australia*, Cambridge University Press, New York.

Marsh, H. W. (1987) 'Students' Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research', *International Journal of Educational Research,* Vol. 11 No. 3, pp. 253-388.

Marsh, H. W. (2007) 'Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness', in Perry, R. P. and Smart, J. C. (Eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective,* Springer, Netherlands, pp. 319-383.

Marsh, H.W. and Ware, J.E. (1982) 'Effects of expressiveness, content coverage and incentive on multidimensional student rating scales: New interpretations of the Dr. Fox effect', *Journal of Educational Psychology,* Vol. 74 No. 1, pp. 126-134.

Marsh, H.W. and Roche, L.A. (2000) 'Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders?' *Journal of Educational Psychology,* Vol. 92 No. 1, pp. 202-228.

Mazzarol, T., Soutar, G. N. and Seng, M. S. Y. (2003) 'The third wave: Future trends in international education', *International Journal of Educational Management,* Vol. 17 No. 3, pp. 90-99.

McKeachie, W. J. (1997) 'Student ratings: The validity of use', *American Psychologist,* Vol. 52 No. 11, pp. 1218–1225.

McPherson, M. A. and Jewell, R.T. (2007) 'Leveling the playing field: Should student evaluation scores be adjusted?' *Social Science Quarterly,* Vol. 88 No. 3, pp. 868–881.

McPherson, M. A., Jewell, R.T. and Kim, M. (2009) 'What determines student evaluation scores? A random effects analysis of undergraduate Economics classes', *Eastern Economic Journal,* Vol. 35 No. 1, pp. 37-51.

Meyers, C. and Jones, T.B. (1993) *Promoting active learning: Strategies for the college classroom*, Jossey Bass, San Francisco, CA.

Millea, M. and Grimes, P.W. (2002) 'Grade expectations and student evaluation of teaching', *College Student Journal,* Vol. 36 No. 4, pp. 582–590.

Miller, J.D. (2006) 'How to fight ratemyprofessors.com', *Inside Higher Ed*, January 31.

Otto, J., Sanford, D.A.J. and Ross, D.N. (2008) 'Does ratemyprofessor.com really rate my professor?' *Assessment & Evaluation in Higher Education,* Vol. 33 No. 4, pp. 355-368.

Mobius, M. M. and Rosenblat, T. S. (2006) 'Why Beauty Matters,' *American Economic Review,* Vol. 96 No. 1, pp. 222-235.

Niu, Shun-Chen. (2005) 'Comparing Two Populations', December 22, University of Texas at Dallas. https://www.utdallas.edu/~scniu/OPRE-6301/documents/Two_Populations.pdf (Accessed 26 March 2021)

Ogier, J. (2005) 'Evaluating the effect of a lecturer's language background on a student rating of teaching form', *Assessment & Evaluation in Higher Education,* Vol. 30 No. 5, pp. 477–488.

Ongeri, J.D. (2009) 'Poor student evaluation of teaching in economics: a critical survey of the literature', *Australasian Journal of Economics Education,* Vol. 6 No. 2, pp. 1-24.

Pan, D., Tan, G. Ragupathi, K., Booluck, K., Roop, R., and Ip, Y. (2009) 'Profiling teacher/teaching using descriptors derived from qualitative feedback: Formative and summative applications', *Research in Higher Education,* Vol. 50 No. 1, pp. 73-100.

Pozo-Munoz, C., Rebolloso-Pacheco, E. and Fernandez-Ramirez, B. (2000) 'The 'ideal teacher' implications for student evaluation of teacher effectiveness', *Assessment & Evaluation in Higher Education,* Vol. 25 No. 3, pp. 253-263.

Reid, L. D. (2010) 'The role of perceived face and gender in the evaluation of college teaching on ratemyprofessors.com', *Journal of Diversity in Higher Education,* Vol. 3 No. 3, pp. 137-152.

Reimann, N. (2004) 'First-year teaching-learning environments in Economics', *International Review of Economics Education,* Vol. 3 No. 1, pp. 9-38.

Rosen, A. S. (2017) 'Correlations, trends and potential biases among publicly accessible web-based student evaluations of teaching: a large-scale study of ratemyprofessors.com data', *Assessment and Evaluation in Higher Education,* Vol. 43 No. 1, pp. 31-44.

Santhanam, E., Lynch, B. and Jones, J. (2018) 'Making sense of student feedback using text analysis – adapting and expanding a common lexicon', *Quality Assurance in Education,* Vol. 26 No. 1, pp. 60-69.

Sarkonak, J. (2016) 'USRIs: students may discriminate against gender and race', *The Gateway*, October 6.

Schmidt, B. (2015) 'Gender bias exists in professor evaluations', *The New York Times*, December 16.

Seldin, P. (1999) *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions*, Anker Publishing Co., Bolton, MA, pp. 1-24.

Sen, A., Voia, M. and Woolley, F. (2010) 'Hot or Not: How appearance affects earnings and productivity in academia', *Carleton Economic Papers* 10-07.

Silva, K. M., Silva, F.J., Quinn, M.A., Draper, J.N., Cover, K.R. and Munoff, A. A. (2008) 'Rate my professor: Online evaluations of psychology instructors', *Teaching of Psychology,* Vol. 35 No. 2, pp. 71–80.

Sheridan, B.J., Hoyt, G. and Imazeki, J. (2014) 'Targeting Teaching: A primer for new teachers of Economics', *Southern Economic Journal,* Vol. 80 No. 3, pp. 839-854.

Smith, B. P. (2007) 'Student ratings of teacher effectiveness: An analysis of end-of-course faculty evaluations', *College Student Journal,* Vol. 41 No. 4, pp. 788–800.

Spooren, P., Brockx, B. and Mortelmans, D. (2013) 'On the validity of student evaluation of teaching: The state of the art', *Review of Educational Research,* Vol. 83, pp. 598–642.

Sproule, R. (2000) 'Student evaluation of teaching: A methodological critique of conventional practices', *Education Policy Analysis Archives,* Vol. 8 No. 50, pp. 1-23.

Stark, P. B. and Freishtat, R. (2014) 'An evaluation of course evaluations', *ScienceOpen Research.* https://www.scienceopen.com/document/vid/42e6aae5-246b-4900-8015-dc99b467b6e4?0 (Accessed 26 March 2021)

Theyson, K. C. (2015) 'Hot or not: The role of instructor quality and gender on the formation of positive illusions among students using ratemyprofessors.com', *Practical Assessment, Research & Evaluation,* Vol. 20 No. 4, pp. 1 -12.

Timmerman, T. (2008) 'On the Validity of RateMyProfessors.com', *The Journal of Education for Business,* Vol. 84 No. 1, pp. 55-61.

Weeden, J. and Sabini, J. (2005) 'Physical Attractiveness and Health in Western Societies: A Review', *Psychological Bulletin,* Vol. 131 No. 5, pp. 635–653.

Weinberg, B. A., Fleisher, B. M. and Mashimoto, M. (2009) 'Evaluating Teaching in Higher Education', *The Journal of Economic Education,* Vol. 40 No. 3, pp. 227-261.

Table 1: Summary of the Rate My Professors data

Instructors	# of instructors	# of Comments	Words coded	White	Native English	Male	Quality	Difficulty	Would take again (%)	100 level	Attendance	Text used	Average grade
MacEwan Economics	11	333	1,016	27.27%	27.27%	100.00%	3.46	2.71	62.87%	80.23%	49.33%	64.86%	3.43
MacEwan Political Science	9	237	859	66.67%	55.56%	77.78%	3.92	2.92	78.04%	64.29%	68.95%	44.33%	3.61
MacEwan Anthropology	14	353	1,419	85.71%	85.71%	35.71%	3.86	2.75	79.86%	65.86%	56.57%	70.57%	3.62
UfA Full Time	22	174	486	68.18%	36.36%	59.09%	3.50	3.10	63.05%	1.43%	50.48%	44.70%	3.64
UfA Retired	14	235	566	92.86%	92.86%	85.71%	3.58	3.32		7.51%		42.12%	
UfA Contract Teaching	13	370	1,165	69.23%	53.85%	84.62%	3.95	2.69	65.30%	32.87%	47.59%	42.62%	3.72
Sum	83	1,702	5,511										
MacEwan Economics Continuing	7	190	580	14.29%	0.00%	100.00%	3.26	2.87	55.72%	68.94%	48.53%	52.53%	3.41
MacEwan Political Science Continuing	7	206	748	71.43%	57.14%	85.71%	3.97	3.09	80.85%	60.12%	72.07%	40.90%	3.58
MacEwan Anthropology Continuing	10	196	781	90.00%	90.00%	30.00%	4.00	2.81	79.49%	54.43%	58.84%	64.80%	3.65
MacEwan AEPS Contract	10	331	1,185	60.00%	70.00%	70.00%	3.69	2.48	76.08%	92.40%	52.48%	79.26%	3.55

Empty cells indicate missing data

Table 2: Coding for content analysis

Coding Categories	Percentages	Key words
Personality/Helpful/Available	8.73%	accessible outside class, cares, helpful office hours, inspirational, kind, nice, professional, emails quickly, respectful, encourages, goes out of the way, hot, humble, respects opinions, works hard
Rude/Personality	3.34%	disrespectful, does not care, rude, shames, unhelpful, arrogant, calls you out, condescending, delayed grading, late responses, not helpful
Clarity/Organization	5.19%	clear expectations, explains well, organized, cares about the subject, clear exam guidelines, clear grading, clear lectures, clear notes, concise, easy to understand, good feedback, simplifies, repeats
Unclear/Tangents/Accent	8.90%	accent, auditory class, bad notes, confusing, difficult to understand, doesn't answer questions, scattered, self-study, tangents, bad examples, bad writing, disorganized, doesn't explain well, makes it hard, makes mistakes, no feedback, poor wording, vague instructions, unstructured
Critical skills/Learning	3.65%	concepts not memorization, learned lots, analyze and understand, care about learning, challenging, changed perception, critical skills, debates, ensures understanding, in-depth, pushed to think, multiple perspectives, learned nothing, don't learn much, memorization
Interaction/Interest/Passion	6.77%	CP matters, engages, interesting, passionate, discussions, sparked interest, CP optional, doesn't like disagreement
Easy A/Pattern/Fair	16.10%	easy A, easy exams, easy HW, easy marking, follow chart, gives away exams, hints, GPA booster, helps succeed, lecture based exams, similar exams, fair exams, fair marking, homework based exams, many sample exams, prepares for exams, sample based exams, review questions are exams, shifts curve, straight forward exams, text based exams
Tough/Grading/Exams	8.95%	different exams, difficult exams, hard quizzes, high expectations, no homework, pop quizzes, tough grader, work hard, tough online course, hard A, hard final, hard homework, long exams, low grades, not an easy A, not a GPA booster, study for exams, tricky exams, worked but scored low
Fun/Story	6.88%	enjoyable, fun, funny, entertaining, stories, makes dull fun, time flies
Boring/Slide reader	4.16%	boring, dry, disengaging, reads PowerPoint slides, dull, monotone, not interesting, slow paced
Real world relevance	1.42%	good examples, real world applications, relevant, real life connections
Fast Pace	0.80%	fast paced, talks fast, throws information, type fast,
Mastery	2.61%	knows stuff, smart, brilliant, smart, thorough, doesn't know stuff, gets stuck, unable to answer, reads notes
Attend/Take notes/Work hard	16.87%	attend class, discipline, lots to read, no notes, pay attention, puts on spot, read text, study hard, take notes, bonus marks in class, CP marks, don't skip class, lots of work, lots to read, old school, work hard,
Notes given/skip/no text	4.07%	can skip class, online notes provided, gives notes, don't need notes, everything posted, great notes, text not used, skip class, text not needed
Math	0.54%	lots of graphs, math heavy, calculus based
Text expensive/bad	0.54%	useless text, expensive text, text required for quizzes,
Tech	0.47%	bonus marks on e-polls, iclicker is a waste of time, video recording of lectures, can book online exam dates

Table 3: Determinants of effective teaching based on student comments

Coding Categories	HIGHLY RATED ($\overline{X_1}$)	POORLY RATED ($\overline{X_2}$)	n1	n2	s1^2	s2^2	n1+n2-2	(n1-1)s1^2	(n2-1)s2^2	Sp^2	Sqrt	$\overline{X_1} - \overline{X_2}$	t
Personality/Helpful/Available	11.99%	6.97%	45	34	0.0084	0.0037	77	0.37	0.12	0.0064	0.0182	5.01%	2.75***
Rude/Personality	1.04%	6.43%	45	34	0.0003	0.0052	77	0.01	0.17	0.0024	0.0112	-5.39%	-4.83***
Clarity/Organization	8.04%	3.19%	45	34	0.0044	0.0023	77	0.19	0.08	0.0035	0.0135	4.85%	3.61***
Unclear/Tangents/Accent	5.28%	17.94%	45	34	0.0038	0.0137	77	0.17	0.45	0.0081	0.0204	-12.67%	-6.20***
Critical skills/Learning	4.71%	3.82%	45	34	0.0031	0.0031	77	0.14	0.10	0.0031	0.0127	0.89%	0.70
Interaction/Interest/Passion	7.18%	3.81%	45	34	0.0031	0.0018	77	0.13	0.06	0.0025	0.0114	3.37%	2.96***
Easy A/Pattern/Fair	15.10%	13.26%	45	34	0.0136	0.0101	77	0.60	0.33	0.0121	0.0250	1.84%	0.74
Tough/Grading/Exams	8.88%	10.92%	45	34	0.0064	0.0091	77	0.28	0.30	0.0076	0.0198	-2.04%	-1.03
Fun/Story	6.92%	2.53%	45	34	0.0040	0.0010	77	0.17	0.03	0.0027	0.0118	4.39%	3.72***
Boring/Slide reader	1.91%	7.76%	45	34	0.0011	0.0070	77	0.05	0.23	0.0036	0.0136	-5.86%	-4.30***
Real world relevance	2.33%	0.67%	45	34	0.0012	0.0002	77	0.05	0.01	0.0008	0.0064	1.67%	2.60**
Fast Pace	0.61%	0.72%	45	34	0.0003	0.0003	77	0.01	0.01	0.0003	0.0038	-0.10%	-0.27
Mastery	3.54%	2.94%	45	34	0.0032	0.0021	77	0.14	0.07	0.0027	0.0118	0.60%	0.51
Attend/Take notes/Work hard	17.14%	14.65%	45	34	0.0116	0.0102	77	0.51	0.34	0.0110	0.0238	2.49%	1.05
Notes given/skip/no text	3.65%	3.18%	45	34	0.0020	0.0025	77	0.09	0.08	0.0022	0.0108	0.48%	0.44
Math	0.79%	0.52%	45	34	0.0006	0.0002	77	0.03	0.01	0.0004	0.0046	0.26%	0.58
Text expensive/bad	0.37%	0.70%	45	34	0.0001	0.0003	77	0.00	0.01	0.0002	0.0031	-0.33%	-1.05
Tech	0.52%	0.00%	45	34	0.0006	0.0000	77	0.03	0.00	0.0004	0.0043	0.52%	1.21

***, **, * denote significance at the 1%, 5% and 10% levels

Notes: The second and third columns indicate the average % of comments for highly and poorly rated instructors in each of the 18 categories. The fourth and fifth columns indicate the number of instructors in the specific categories. The next two columns indicate the variances of the coding categories. The four columns after that use the methods described in Niu (2005) to determine the pooled variance "S^2" for the coding categories. The pooled variances allow computing the t statistic to test for the significance of the difference between the coding categories for the highly rated and poorly rated instructors.

Table 4: Content analysis for MacEwan AEPS and UfA instructors

MACEWAN	ECON (11)	Anthro/ Pol Sci (22)	t	AEPS Continuing (23)	AEPS Contract (10)	t	UfA Retired (14)	UfA Contract (13)	t	UfA Full-time (19)	UfA Contract (13)	t
Personality/Helpful/Available	5.45%	9.80%	-1.67	8.98%	6.91%	0.75	9.66%	8.96%	0.22	13.11%	8.96%	1.37
Rude/Personality	5.21%	3.26%	0.96	3.75%	4.27%	-0.25	2.61%	1.60%	0.72	4.17%	1.60%	1.15
Clarity/Organization	2.64%	5.34%	-1.50	3.82%	5.87%	-1.09	5.60%	8.16%	-1.05	7.32%	8.16%	-0.30
Unclear/Tangents/Accent	15.24%	8.96%	1.77*	11.57%	9.87%	0.45	9.42%	9.78%	-0.08	11.77%	9.78%	0.50
Critical skills/Learning	1.22%	4.04%	-2.12**	3.56%	2.04%	1.06	8.28%	3.80%	1.95*	3.91%	3.80%	0.05
Interaction/Interest/Passion	4.57%	8.78%	-2.00**	7.37%	7.40%	-0.01	4.94%	5.90%	-0.58	3.33%	5.90%	-1.69
Easy A/Pattern/Fair	15.95%	17.24%	-0.32	14.95%	21.07%	-1.52	8.49%	15.92%	-2.04**	13.15%	15.92%	-0.66
Tough/Grading/Exams	12.00%	5.68%	2.69**	7.98%	7.34%	0.24	11.42%	9.16%	0.71	12.35%	9.16%	0.90
Fun/Story	4.46%	7.08%	-1.38	6.18%	6.27%	-0.05	5.32%	5.51%	-0.07	2.46%	5.51%	-1.56
Boring/Slide reader	7.65%	3.58%	1.51	5.58%	3.45%	0.75	6.07%	2.83%	1.22	3.43%	2.83%	0.41
Real world relevance	2.06%	0.20%	3.19**	0.47%	1.62%	-1.76*	2.46%	2.84%	-0.28	1.54%	2.84%	-1.01
Fast Pace	0.65%	0.49%	0.44	0.67%	0.25%	1.11	0.85%	0.73%	0.15	0.67%	0.73%	-0.10
Mastery	1.88%	2.51%	-0.77	2.92%	0.88%	2.65**	8.05%	1.11%	2.89***	2.96%	1.11%	1.19
Attend/Take notes/Work hard	18.35%	18.87%	-0.12	18.59%	18.96%	-0.08	12.83%	14.84%	-0.54	14.72%	14.84%	-0.04
Notes given/skip/no text	1.70%	3.45%	-1.00	3.06%	2.41%	0.36	2.48%	5.28%	-1.52	3.92%	5.28%	-0.74
Math	0.70%	0.00%	1.68	0.05%	0.67%	-1.42	0.65%	1.60%	-0.93	0.83%	1.60%	-0.80
Text expensive/bad	0.27%	0.73%	-0.86	0.52%	0.72%	-0.37	0.87%	0.17%	1.35	0.36%	0.17%	0.48

Note: The numbers in parentheses in the first row refer to the number of instructors in each of the instructor categories.

Table 5: Penalization of effective instructors

MACEWAN - ECONOMICS	I1	I2	I3	I4	I5	I6	AVERAGE (I1 – I6)	Instructors rated 4.5 and above (17)	Instructors rated 3 - 4.5 (51)	t
Personality/Helpful/Available	7.69%	14.29%	2.16%	10.13%	10.34%	2.67%	7.88%	11.63%	10.19%	0.59
Rude/Personality	0.00%	1.43%	2.16%	5.06%	3.45%	1.33%	2.24%	0.61%	2.95%	-2.04**
Clarity/Organization	2.10%	4.29%	5.41%	6.33%	3.45%	2.67%	4.04%	9.35%	5.72%	2.02**
Unclear/Tangents/Accent	0.00%	1.43%	4.32%	13.92%	31.03%	10.67%	10.23%	2.80%	11.14%	-3.40***
Critical skills/Learning	0.70%	1.43%	1.62%	0.00%	0.00%	0.00%	0.62%	4.44%	4.36%	0.05
Interaction/Interest/Passion	4.90%	11.43%	12.43%	3.80%	0.00%	8.00%	6.76%	10.30%	4.66%	4.17***
Easy A/Pattern/Fair	46.15%	32.86%	6.49%	11.39%	3.45%	8.00%	18.06%	20.76%	12.34%	2.90***
Tough/Grading/Exams	0.00%	0.00%	8.11%	24.05%	27.59%	10.67%	11.74%	5.17%	10.94%	-2.38***
Fun/Story	10.49%	1.43%	16.76%	1.27%	6.90%	5.33%	7.03%	9.94%	4.01%	4.23***
Boring/Slide reader	0.00%	0.00%	0.00%	3.80%	6.90%	0.00%	1.78%	0.63%	4.66%	-2.54***
Real world relevance	7.69%	1.43%	4.86%	1.27%	0.00%	5.33%	3.43%	2.08%	1.56%	0.61
Fast Pace	0.00%	0.00%	3.24%	0.00%	0.00%	1.33%	0.76%	0.26%	0.70%	-0.97
Mastery	0.00%	2.86%	4.86%	0.00%	6.90%	1.33%	2.66%	4.24%	3.40%	0.54
Attend/Take notes/Work hard	19.58%	25.71%	25.41%	16.46%	0.00%	36.00%	20.53%	13.81%	18.24%	-1.51
Notes given/skip/no text	0.70%	1.43%	0.00%	1.27%	0.00%	0.00%	0.57%	3.72%	3.23%	0.38
Math	0.00%	0.00%	1.08%	0.00%	0.00%	6.67%	1.29%	0.19%	0.84%	-1.14
Text expensive/bad	0.00%	0.00%	1.08%	1.27%	0.00%	0.00%	0.39%	0.07%	0.60%	-1.43
Quality	4.9	4.7	4.1	3.8	3.6	3.6	4.12			
Notes	1	1	0	1	0	0				

Comments	
I1	easy A, follow chart, attend class, funny, real life
I2	attend class, lecture-based exams, passionate, easy
I3	attend class, take notes, work hard, fun, engages, hard exams
I4	quick response, hard quizzes, easy exams, read text
I5	knows stuff, hard exams, self-study
I6	attend class, puts on spot, pop quiz, interesting