e-Journal of Business Education & Scholarship of Teaching
Vol. 13, No. 3, December 2019, pp: 1-13.
http://www.ejbest.org

Student Evaluation of Teaching in Business Education: Discovering Student Sentiments Using Text Mining Techniques

Swathi Baddam*
Assistant Professor
Department of Management & Marketing
Doermer School of Business
Purdue University Fort Wayne, USA
baddams@pfw.edu
*Corresponding Author

Prasad Bingi
Associate Professor
Department of Management & Marketing
Doermer School of Business
Purdue University Fort Wayne, USA
bingi@pfw.edu

Syed Shuva
Graduate Student
Doermer School of Business
Purdue University Fort Wayne, USA
shuvs01@pfw.edu

Abstract

The purpose of this research is to assess the student sentiments from student evaluations. Student written comments in student evaluations from a midwestern university are analyzed using text mining software. Sentiments and themes are discovered that can be used to improve teaching effectiveness. Analysis shows noteworthy sentiment differences across courses and students. There are major differences in the sentiments between graduate and undergraduate students. Within the undergraduate students, the sentiments also varied from freshmen to senior students. The nature of course also has an effect on student sentiments, particularly in hybrid and online courses. Practical implications of this research are also discussed.

Key words: *Text mining; sentiment analysis; student evaluations; thematic analysis, qualitative analysis.*

JEL Classification: M31
PsycINFO Classification: 3550
FoR Code: 1301; 1503
ERA Journal ID#: 35696

Introduction

Student evaluation of teaching (SET) is a common method to assess teaching effectiveness in higher education. Traditionally, student evaluations are administered towards the end of the semester using a paper-based survey. However, recently online evaluations are becoming popular. The results of student evaluations are commonly used in decisions related to hiring, promotion and tenure, merit raises, and other performance measurements of instructors (Stupans, 2016; Amr, Michael, & Tiffany, 2010). Student evaluations are also significant because they may be the only opportunity for students to provide constructive feedback that could improve the future student learning outcomes. Although there are some concerns about the effectiveness of student evaluations (Slade & McConville, 2006; Clayson, 2009; Hornstein, 2017) they continue to be widely used in higher educational institutions.

SET questionnaires contain both standardized questions which are rated on a Likert scale, and open-ended questions which capture student opinions that are not covered by the standardized questions. Typically, the responses on standardized questions are summarized while the qualitative part is open for interpretation by the faculty. As it is time consuming for faculty to undertake a systematic analysis of the student written comments on their own, they often ignore it. However, such an analysis can be a valuable source of feedback to improve course outcomes (Kabanoff, Richardson, & Brown, 2003). Further, every course has unique features that cannot be evaluated by a standardized set of questions. Analyzing textual data is easily accessible now through a number of tools or models found in text mining software. Sentiment analysis that is very popular in marketing and social media analytics can be used to understand the student opinions in a different lens when applied to qualitative data in student evaluations. This study will apply sentiment and thematic analyses to explore and understand teaching effectiveness. Further, it will investigate how the student and course characteristics influence the student sentiments.

Literature Review

A number of researchers examined how course and instructor characteristics influence differences in student evaluations (Macfadyen, Dawson, Prest, & Gašević, 2016; Mardikyan & Badur, 2011; Narayanan, Sawaya, & Johnson, 2014). Even among the business schools, evaluations collected from business courses showed that using the same survey instrument for different courses was not effective because ratings differed significantly across different types of courses (Whitworth, Price, & Randall, 2002). In another comprehensive study conducted by (Narayanan et al., 2014), the course level and course grades were found to be significant. Uttl & Smibert (2017) showed that class subject was strongly associated with SET ratings that also had a substantial impact on professors being labeled satisfactory vs. unsatisfactory and excellent vs. non-excellent.

Beran & Violato (2005) found that students who attend classes regularly and have high grade expectations provide high ratings for their instructors. Additionally, Mardikyan and Badur (2011) explored many instructor characteristics and found that both preparedness for the class and fair grading significantly influenced the student evaluations.

While much research focused on studying quantitative data, there is a rising interest in examining the qualitative part of SET's. Early research in this area by Alhija & Fresko (2009) found that student comments revealed unique aspects of courses that were not captured in closed-ended questions. Amr, Michael, & Tiffany (2010) utilized text mining techniques to extract positive and negative comments using co-occurrence logic for quantifying students' open-ended responses. Brockx, Van Roy, & Mortelmans

(2012) showed that surveys that received low scores were more likely to have negative comments while higher scores tend to have positive comments indicating the consistency between student comments and the quantitative SET scores.

The validity of using SET's in measuring teaching effectiveness has been questioned in the literature (Clayson, 2009; Hornstein, 2017; Uttl, White, & Gonzalez, 2017). Further, Stark and Freishtat (2014) suggested that the use of average SET scores does not reflect teaching effectiveness and recommend utilizing student comments that can add valuable information. Özgüngör & Duru (2015) utilized qualitative analysis to reveal four themes that explain the differences between high and low performing instructors: lecturing, relationship with the students, knowledge and expertise, and exams and evaluation. Stupans et al. (2016) employed text analysis techniques on open-ended comments to extract various themes which include mode, concept, exam, content, work, assignments, lectures, intensive school, staff, interesting, practical. Text analysis from student comments is also able to identify domains and sub-domains in student experience that need to be improved for better student satisfaction (Grebennikov & Shah, 2013; Scott, Grebennikov, & Shah, 2008). Qualitative analysis, particularly in business schools is also explored to extract themes concerning course content, student learning environment, assessment, and teaching styles (Steyn, Davies, & Sambo, 2019).

The relationship between student satisfaction and student characteristics such as class standing (freshman, sophomore, junior, senior, and graduate) is largely unexplored in the literature. Similarly, influence of course characteristics such as discipline or subject matter and quantitative nature of a course on student satisfaction and sentiment has not received attention in the qualitative research on SET. Therefore, this research proposes the following null-hypotheses to address the paucity in the SET literature.

H_{01}: Student characteristics will not influence the student sentiment.
H_{02}: Course characteristics will not influence the student sentiment.

Further, this research will use thematic analysis to discover teaching dimensions and concepts that influence the student sentiment.

Data

Student evaluation data used in this study is anonymously collected as part of the course evaluations that are conducted regularly in the business school of a midwestern university in the United States. The data is obtained from a secondary source in the university that collected and stored student evaluations. Any references to instructor such as instructor names are redacted from the student written comments at the source. The Institutional Review Board at the university reviewed and approved the research proposal as it meets the ethical standards and in a manner that promotes the protection of the rights and welfare of human subjects. A total of 939 usable comments are retrieved across 110 business courses taught over three semesters in 2016-17 time period. The data in its entirety consists of student responses to one open-ended section ("please write your comments below") at the end of SET questionnaire. The following is the frequency distribution of comments based on the student class standing- freshmen (11%), sophomore (15%), junior (34%), senior (27%), and graduate (12%). The data is also grouped depending on the type of course - marketing (20%), operations (13%), information technology (23%), and management (44%). Further, the courses were categorized as quantitative (13%), non-quantitative (general-61%), and technical (25%) based on the nature of the course content. The last course characteristic is on course requirements that belonged one of the following course types - general education (11%), core (63%), major required (14%), and major elective (12%).

Method

The text analysis method used in this study is an unsupervised learning technique, which means that the model does not build on a known pattern or response variable. The method will search for unrecognized patterns making it less prone to bias. The NVivo software is selected for this research as it minimizes researcher bias by using a predefined dictionary of words. Student comments were corrected for minor spelling and grammatical errors without impacting the original intent. The data is prepared and coded for student and course characteristics while the software automatically recognizes the open-ended comments and evaluates them accordingly. The sentiment analysis can be executed automatically on the textual data to discover underlying sentiments. Thematic analysis, on the other hand, extracts themes based on the semantics of frequently observed words or phrases and may require manual adjustments.

Sentiment Analysis

Sentiment analysis is performed to extract people's opinion from textual data. As opinions are subjective and lie in the contextual nature of words and phrases, it is difficult to extract the precise and accurate opinions from data. Sentiment analysis helps to identify the polarity and the strength of opinions. The polarized nature of sentiment is extracted through unsupervised machine learning algorithm available within the NVivo software. The software first searches for expression of opinions in the source material with each word in isolation. The words with expression have a predefined score that are scored on a range of sentiment scale: "Very Negative", "Moderately Negative", "Neutral", "Moderately Positive", and "Very Positive". Then each word is coded under the corresponding sentiment score, except for the "Neutral" sentiment. Scores are affected if they are preceded by a modifying word (example: very, somewhat). The software allows to view words and references under each sentiment code. It also allows the researchers to modify or omit words from a sentiment node (NVivo, 2019a).

Sentiment analysis is performed on each category to explore the distribution of sentiment. The influence of student characteristic on the opinion is explored by comparing the proportion of each sentiment node across the type of class standing (freshman, sophomore, junior, senior, and graduate). Similarly, the influence of course characteristic is explored by comparing the proportion of sentiment nodes across each type of course characteristics.

Thematic Analysis

The thematic analysis is used to identify patterns in the textual data. Themes are extracted with unsupervised machine learning algorithm available in NVivo. The themes are identified by analyzing the content and sentence structure. The process first detects significant meaningful phrases in the source to identify frequently occurring themes. Themes are then grouped together by comparing words with the same stem, for example, network, networks, and networking are all grouped under one theme (NVivo, 2019b). This list is then examined manually and some themes are combined based on the common meanings and context found in the automated themes. For example, "course" and "class" are merged in the top themes but they are essentially referring to the course on which the student comments are collected. This process is repeated to arrive at a final the list of themes that capture the unique aspects of student opinions.

Results

Sentiment Analysis Results

The results from the sentiment analysis show varied opinions expressed by the students in their comments. Table 1 shows summary sentiment results for all course and student characteristics included in the study. The percentage sentiments shown in the table indicate that overall, the student sentiments vary across the categories and sub-categories. In order to make the interpretation simple, moderately positive sentiment and very positive sentiment are both ¬¬grouped under 'positive' while moderately negative sentiment and very negative sentiment are grouped under 'negative'. After sentiment categories are merged, freshmen show higher positive sentiment (89%=48%+41%) in comparison to all the other undergraduate students. On the other hand, higher negative sentiment (42%=27%+15%) is observed in the graduate students among all students who commented indicating that class standing may have an effect on the student sentiment. These differences found in student sentiments across student class standing are tested for statistical significance using the chi-square test. The sentiments generated across class standing are significant with X2 (9, N=599) =39.27, p=0.00. Given the statistical significance the null-hypothesis is rejected - class standing influences student sentiment (H_{01}).

Similarly, viewing from the course characteristics perspective, students express higher negative sentiment (60%) in quantitative classes compared to non-quantitative classes. Further, the field of course also show differences in student sentiment with operations courses showing the highest negative sentiment with 50%. Finally, the required classes have negative sentiment of 40% while the elective classes show a 35%. The sentiments in field of the course are also statistically different with X2 (9, N=744) =39.27, p=0.00. The chi-square tests are also significant across type of course requirement and the sentiments with X2 (9, N=774) =37.63, p=0.00. Finally, the relationship between sentiment categories and the quantitative nature of the course is also found to be significant X2 (6, N=774) =76.56, p=0.00. As the statistical tests indicate that course characteristics influenced student sentiment the null-hypothesis (H_{02}) is also rejected.

Table 1:
Sentiment Results

Dimension	Category	Subcategory	Very Negative	Moderately Negative	Moderately Positive	Very Positive
Student Characteristics	*Class Standing*	Freshman	6%	6%	48%	41%
		Sophomore	8%	11%	52%	29%
		Junior	11%	25%	39%	25%
		Senior	14%	12%	44%	30%
		Graduate	15%	27%	42%	16%
Course Characteristics	*Field of the course*	Management	9%	11%	49%	31%
		Marketing	14%	25%	38%	23%
		Operations	18%	32%	32%	19%
		Information Technology	10%	13%	46%	31%
	Type of Course requirement	General Education	6%	6%	48%	41%
		Core	9%	17%	47%	27%
		Major Required	13%	27%	38%	21%
		Major Elective	21%	15%	36%	29%
	Quantitative nature of course	General	7%	15%	47%	32%
		Quantitative	26%	34%	26%	13%
		Technical	12%	13%	47%	28%

Thematic Analysis Results

Given the statistical significance of the sentiment results both null-hypotheses H_{01} and H_{02} are rejected. Consequently, it appears that the use of thematic analysis has potential for exploring teaching concepts. The data was further analyzed to extract themes that are commonly observed in the student comments. Table 2 provides the list that is based on the frequency of themes found from the automated process of NVivo. Once a full list of themes was generated, the themes were categorized based on the underlying common features in them. For example, both 'hybrid' and 'online' themes fit into the course delivery methods. Further, the themes are categorized based on commonalities into higher level of grouping depending on whether the concepts are linked to the course or instructor. After repeating this process for all themes, a comprehensive list of categories, concepts, and themes are developed as shown in Table 2.

Frequently, students in all class standings include general evaluation of class (31%) and instructor (31.5%) in their comments which are labeled as 'overall' in Table 2. Comments related to course delivery methods which are hybrid and online are merged into the 'delivery' concept. Students' comments on course material are grouped into the resources category while workload concept includes both course related work and the time aspects. The group and project themes are placed in the collaboration concept. The concept of relevancy is unique to the instructor category where comments are linked to relevant examples provided by an instructor. Since the frequencies on some concepts such as delivery, resources, and relevancy, are much lower than other concepts, we will be judicious in generalizing our conclusions regarding them.

Table 2:
Thematic Results

Category	Concepts	Theme	Frequency	Percentage
Course	Overall	Great/good/excellent course, etc.	183	31.0%
	Delivery	Hybrid, online	46	7.8%
	Resources	Material	29	4.9%
	Workload	Work, time	60	10.2%
	Collaboration	Group, project	57	9.6%
Instructor	Overall	Amazing/great professor	186	31.5%
	Relevancy	Examples	30	5.1%

Further, exploring sentiments across these themes may provide a great opportunity to identify target areas to improve. In order to facilitate easy comparison of sentiments across different student and course characteristics, we would like to propose a weighted sentiment index that would factor in the number of references in each sentiment category. The sentiment index is formulated by assigning weights to the frequency of very positive (VP), moderately positive (MP), moderately negative (MN), and very negative (VN) references. The very positive and very negative sentiments are assigned positive one (1) and negative one (-1) weights respectively, while the moderately positive and moderately negative sentiments are each worth half, 0.5 for moderately positive and -0.5 for moderately negative categories. The weights are assigned to each reference according to the sentiment category it belongs to, and then the sum of these weighted references is compared to the total number of references to calculate the index. The final index calculated will be referred to as aggregate sentiment index (ASI). By design, ASI values will range from -1 to +1 with positive values indicating positive sentiment and negative values indicating negative sentiment. The magnitude of the index shows the strength of the sentiment.

$$ASI = \frac{VP + 0.5 * MP - 0.5 * MN - VN}{VP + MP + MN + VN}$$

**VP: # of Very Positive sentiment references; MP: # of Moderately Positive sentiment references; VN: # of Very Negative sentiment references; MN: # of Moderately Negative sentiment references*

Table 3 summarizes the sentiments using ASI across all concepts identified. If there are not enough references to calculate ASI they are labelled as "NA". Most of the ASI values in Table 3 are positive except for junior and graduate students who have some negative ASI values. However, the negative index values are associated with low number of references and hence do not warrant any further generalization.

Table 3:
Aggregate Sentiment Index in Themes: Class Standing

Category	Concepts	Theme	FRSH	SOPH	JUNR	SENR	GRAD
Course	Overall	Great course	0.50 *(11)*	0.50 *(8)*	0.26 *(27)*	0.36 *(37)*	0.04 *(14)*
	Delivery	Hybrid, Online	NA	0.50 *(3)*	NA	0.50 *(1)*	0.17 *(6)*
	Resources	Material	NA	0.17 *(6)*	NA	0.17 *(6)*	0.50 *(1)*
	Workload	Work, Time	NA	0.13 *(12)*	-0.25 *(6)*	0.50 *(5)*	0.33 (3)
	Collaboration	Group, Project	0.43 *(7)*	NA	-0.05 *(10)*	NA	-0.10 *(5)*
Instructor	Overall	Amazing/great professor	0.55 *(22)*	0.52 *(31)*	0.69 *(29)*	0.59 *(37)*	0.06 *(9)*
	Relevancy	Examples	0.50 *(2)*	NA	0.50 *(3)*	0.60 *(5)*	0.40 *(5)*

** FRSH: *Freshman*, SOPH: *Sophomore*, JUNR: *Junior*, SENR: *Senior*, GRAD: *Graduate; The values in parentheses denote number of sentiment references*

The ASI values in Table 3 show that junior students express highest positive index (0.69) when evaluating instructors in general. Consistent with the results from the sentiment analysis reported in Table 1, freshmen display higher positive sentiment among undergraduate students, while graduate students overall express lower positive sentiment.

The ASI is also calculated for all three types of course characteristics - field of study, type of course requirement, and nature of course in Tables 4, 5, and 6 respectively. Course characteristics are also explored for their influence on sentiments in the concepts. Largely, marketing courses are associated with high positive sentiment. A negative sentiment is seen when a course is quantitative in nature, as seen from negative ASI in Table 5. Technical courses also show some negative sentiment specifically in course workload and collaborative projects. General (GEN) type of courses as seen in Table 5, largely have the highest positive sentiment in all areas. Finally, courses that are in major both required and elective have lower positive sentiment than general education and core courses as shown in Table 6. Mostly, a high positive sentiment observed in general education type courses, in all concepts. Overall, the instructors are positively evaluated by students as shown by positive ASI values in all course and student characteristics.

Table 4:
Aggregate Sentiment Index in Themes: Field (Course Characteristics)

Category	Concepts	Theme	MGMT	MKTG	OP	IT
Course	Overall	Great course	0.37 *(56)*	0.43 *(22)*	0.12 *(17)*	0.29 *(12)*
	Delivery	Hybrid, online	-1.00 *(1)*	0.67 *(3)*	0.17 *(6)*	0.30 *(5)*
	Resources	Material	NA	0.60 *(5)*	-0.13 *(12)*	0.47 *(18)*
	Workload	Work, time	NA	1.00 *(2)*	0.00 *(2)*	-0.03 (16)
	Collaboration	Group, project	0.19 *(13)*	0.25 *(2)*	0.00 *(6)*	0.13 *(4)*
Instructor	Overall	Amazing/great professor	0.58 *(59)*	0.38 *(25)*	0.42 *(6)*	0.50 *(34)*
	Relevancy	Examples	0.35 *(10)*	0.15 *(13)*	NA	-1.00 *(1)*

**MGMT: *Management*, MKTG: *Marketing*, OP: *Operations*, IT: *Information Technology*
The values in parentheses denote number of sentiment references

Table 5:
Aggregate Sentiment Index: Quantitative Nature of Course (Course Characteristics)

Category	Concepts	Theme	GEN	QUAN	TECH
Course	Overall	Great course	0.45 *(75)*	-0.13 *(19)*	0.22 *(29)*
	Delivery	Hybrid, online	0.33 *(3)*	0.21 *(7)*	0.30 *(5)*
	Resources	Material	NA	-0.36 *(11)*	0.00 *(5)*
	Workload	Work, time	0.22 *(9)*	-0.08 *(6)*	-0.08 *(13)*
	Collaboration	Group, project	0.17 *(9)*	0.00 *(4)*	-0.17 *(6)*
Instructor	Overall	Amazing/great professor	0.60 *(78)*	-0.13 *(4)*	0.52 *(43)*
	Relevancy	Examples	0.68 *(11)*	NA	0.63 *(4)*

** GEN: *General*, QUAN: *Quantitative*, TECH: *Technical*
The values in parentheses denote number of sentiment references

Table 6:
Aggregate Sentiment Index: Type of Course Requirement (Course Characteristics)

Category	Concepts	Theme (% comments)	GE	CO	MJR	MJEL
Course	Overall	Great course	0.47 *(15)*	0.24 *(55)*	0.32 *(19)*	0.19 *(18)*
	Delivery	Hybrid, online	NA	0.15 *(10)*	0.50 *(1)*	NA
	Resources	Material	NA	0.44 *(9)*	NA	0.00 *(5)*
	Workload	Work, time	NA	0.24 *(41)*	-0.40 *(5)*	0.33 *(3)*
	Collaboration	Group, project	0.50 *(4)*	0.13 *(8)*	-0.25 *(6)*	-0.75 *(2)*
Instructor	Overall	Amazing/great professor	0.52 *(25)*	0.52 *(63)*	0.65 *(10)*	0.52 *(21)*
	Relevancy	Examples	0.50 *(1)*	0.50 *(11)*	NA	0.64 *(7)*

** GE: *General Education*, CO: *Core*, MJR: *Major Required*, MJEL: *Major Elective*
The values in parentheses denote number of sentiment references

Conclusion and Discussion

The sentiment analysis (Table 1) shows that both freshmen and sophomore students have higher positive sentiment in comparison to students belonging to higher class standings. This is possibly due to high school students entering university level education as freshmen may have higher satisfaction in the first couple of years, because of high degree of freedom in class selection, class schedules, etc. Additionally, the university in which this particular dataset belongs to has various programs (e.g. first-year student satisfaction survey) and mechanisms in place to retain students and improve student satisfaction. The decline in student sentiment from the second year may be attributed to higher workload and difficult subject matter in higher level classes as they get introduced to major core and elective courses. The class standing can explain some statistically significant differences in SET (Narayanan et al., 2014) particularly finding that freshmen courses are different from others.

Graduate and junior students express higher negative sentiment (Table 1) indicating a need for improvement to meet their expectations. Further exploration on graduate student sentiments using ASI (in Table 3) shows that the graduates expressed negative sentiments resulting from group work including projects and other group activities. However, the frequency of sentiment references is low. The institution has to investigate further on this aspect by specifically including questions related to group work. Nevertheless, students expressed positive sentiments for both the courses and the instructors.

Whitworth et al. (2002) found significant differences in student evaluations across different course types. This study also found that the course characteristics such as the field of study, the reason for taking the course, and the quantitative nature of course content influence the student sentiments. Quantitative type of courses received higher negative sentiment, which can be explained by the presence of math-based material and increased workload in those courses, as seen in Table 5. Prior research also shows these differences in SET ratings between math-based and non-math courses (Beran & Violato, 2005; Hodges & Stanton, 2007). Surprisingly in this study, information technology-oriented courses did not have negative sentiments even though they typically have difficult course content. With regard to the positive sentiments related to IT courses, perhaps the students felt that IT is pervasive in the workplace, and they need to be IT savvy. The students could have accepted the extra effort and challenges associated with the IT-intensive courses as necessary to gain mastery of the IT skills that could lead to successful employment in the future. Moreover, the availability of tutoring labs where students can get help on their information technology projects and assignments could have mitigated the stress of students and resulted in improved positive sentiment.

Student sentiments also varied from required courses to elective courses (Table 6). The elective courses received higher positive sentiment possibly because of students choosing them due to their prior interest in the subject matter. Kozub (2010) found evidence in their study that students perceived electives to be more valuable than non-electives. The students in required courses expressed negative opinions on working in groups as seen in the negative ASI values (Table 6). This is the common theme that has been observed in the data pointing out that the students do not hold positive opinion on working in groups. This sentiment can be also be attributed to the introduction of many online courses in the department, which make working in groups more challenging for students.

This research has practical implications for both instructors and university administrators. To improve the overall student experience, the institution needs to leverage factors that contribute to positive sentiment and address the issues that affect students negatively. On the other hand, the instructors were perceived positively by the students. The higher positive sentiment was due to instructors bringing relevant or real-life examples from the business world that enriched the student experience. Largely, students at this institution have negative sentiment towards collaboration in the coursework. These results are being shared with the faculty in open forums to discuss further the reasons that contribute to this negative sentiment and methods to improve the collaborative work. The automated mining tools are useful for universities to gain additional insights that are easier to discover. The researchers plan to discuss needed intervention strategies to mitigate or prevent the negative behaviors and contribute to the overall positive student experience.

Limitations

This research like any other is not without any limitations. One of the main limitations is the lack of detail in student information in the dataset since the data is provided by a secondary source. Individual level of student information such as expected grade would have given more robustness to the study. The data is also relatively small with only courses taught over three semesters in the business school. Additionally, the comments included in the analysis do not have corresponding quantitative evaluations. The conclusions are limited to the students who wrote comments and may not represent the entire class. There are also limitations in using automatic process in NVivo as it does not identify sarcasm, double negatives, etc. in the analysis. Further, manual intervention used in the analysis means that there is a possibility of researcher bias particularly in when grouping themes into concepts. Future research may utilize both quantitative and qualitative data to explore if the sentiments in student comments will also reflect in the quantitative ratings.

References

Alhija, F. N. A., & Fresko, B. (2009). Student evaluation of instruction: What can be learned from students' written comments? *Studies in Educational Evaluation*, *35*(1), 37–44.

Amr, A.-E., Michael, A., & Tiffany, A. (2010). Using text data mining techniques for understanding free-style question answers in course evaluation forms. *Research in Higher Education Journal*, *9*, 12-23.

Beran, T., & Violato, C. (2005). Ratings of university teacher instruction: How much do student and course characteristics really matter? *Assessment and Evaluation in Higher Education*, *30*(6), 593–601.

Brockx, B., Van Roy, K., & Mortelmans, D. (2012). The student as a commentator: Students' comments in student evaluations of teaching. *Procedia - Social and Behavioral Sciences*, *69*, 1122–1133.

Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn? A meta-analysis of the review and literature. *Journal of Marketing Education*, *31*(1), 16–30.

Grebennikov, L., & Shah, M. (2013). Student voice : Using qualitative feedback from students to enhance their university experience. *Teaching in Higher Education 18*(6), 606–618.

Hodges, L. C., & Stanton, K. (2007). Translating comments on student evaluations into the language of learning. *Innovative Higher Education*, *31*(5), 279–286.

Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, *4*(1), 1–8.

Kabanoff, B., Richardson, A., & Brown, S. (2003). Business graduates ' perceptions of the quality of their course : A view from their workplace. *Journal of Institutional Research*, *12*(2), 1-12.

Kozub, R. M. (2010). Relationship of course, instructor, and student characteristics to dimensions of student ratings of teaching effectiveness in business schools. *American Journal of Business Education*, *3*(1), 33-40.

Macfadyen, L. P., Dawson, S., Prest, S., & Gašević, D. (2016). Whose feedback? A multilevel analysis of student completion of end-of-term teaching evaluations. *Assessment and Evaluation in Higher Education*, *41*(6), 821–839.

Mardikyan, S., & Badur, B. (2011). Analyzing teaching performance of instructors using data mining techniques. *Informatics in Education*, *10*(2), 245–257.

Narayanan, A., Sawaya, W. J., & Johnson, M. D. (2014). Analysis of differences in nonteaching factors influencing student evaluation of teaching between engineering and business classrooms. *Decision Sciences Journal of Innovative Education*, *12*(3), 233–265.

NVIVO QSR (2019a). How auto coding sentiment works. Retreived from: http://help-nv11.qsrinternational.com/desktop/concepts/How_auto_coding_sentiment_works.htm (Accessed on March, 2019).

NVIVO QSR (2019b). How auto coding themes works. Retreived from: http://help-nv11.qsrinternational.com/desktop/concepts/how_auto_coding_themes_works.htm (Accessed on March, 2019).

Özgüngör, S., & Duru, E. (2015). Course and instructor characteristics distinguishing highest and lowest student ratings of instructors. *Eurasian Journal of Educational Research*, *15*(61), 118–136.

Scott, G., Grebennikov, L., & Shah, M. (2008). Using qualitative data to prove and improve quality in Australian higher education. *Evidence Based Decision Making: Scholarship and Practice,* 97-111.

Slade, P., & McConville, C. (2006). The validity of student evaluations of teaching. *International Journal for Educational Integrity*, 2(2), 140-155.

Stark, P., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research,* 1-7.

Steyn, C., Davies, C., & Sambo, A. (2019). Eliciting student feedback for course development: The application of a qualitative course evaluation tool among business research students. *Assessment and Evaluation in Higher Education*, *44*(1), 11–24.

Stupans, I., McGuren, T., & Babey, A. M. (2016). Student evaluation of teaching: A study exploring student rating instrument free-form text comments. *Innovative Higher Education*, *41*(1), 33–42.

Uttl, B., & Smibert, D. (2017). Student evaluations of teaching: Teaching quantitative courses can be hazardous to one's career. *PeerJ*, *5*, e3299.

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, *54*, 22–42.

Whitworth, J. E., Price, B. A., & Randall, C. H. (2002). Factors that affect college of Business student opinion of teaching and learning. *Journal of Education for Business*, *77*(5), 282–289.