Determinants of Accounting Student Evaluations of Teaching Scores

Saad S. Albuloushi[1] & Mishari M. Alfraih[1]

[1] Accounting Department, College of Business Studies, the Public Authority for Applied Education and Training, Kuwait

Correspondence: Mishari Alfraih, Accounting Department, College of Business Studies, the Public Authority for Applied Education and Training, Kuwait. Tel: 965-997-556-66. E-mail: m@dralfraih.com

Abstract

Given the prevalent use of the student evaluations of teaching (SET) as a measure of teaching effectiveness, this study aims to investigate the determinants of SET scores among students attending the College of Business Studies at the Public Authority for Applied Education and Training (PAAET), Kuwait. A total of 678 SET were analysed using univariate and multiple regression analyses. It was found that SET scores were significantly and positively biased by expected grade, student age and course level. In contrast, class size and faculty experience were found to be significantly and negatively related to SET. Expected grade had the strongest impact on SET scores. The study findings raise concerns about the reliability and validity of the SET as well as their suitability for evaluation purposes. As SET scores have an important assessment function and serve as formative and summative measures in personnel decisions, the incentives for faculty to compromise their grading standards to receive good teaching evaluations increase. Accordingly, administrators should devote more effort to ensure a careful and complete understanding and interpretation of SET if they want to effectively incorporate them into the faculty evaluation process. To the authors' knowledge, this is the first study to explore determinants of student evaluations of teaching scores in Kuwait.

Keywords: student evaluation of teaching, expected grades, bias, accounting education, Kuwait

1. Introduction

Student evaluations of teaching (SET) are an integral part of the educational and training process commonly used in higher education institutions to measure teaching effectiveness. Prior research shows that SET have become the most prevalent measure of teaching effectiveness across universities (Bonitz, 2011). Pounder (2007) argues that the extent of reliance on the SET as the predominant measure of teaching effectiveness is not confined to the USA; it is a worldwide phenomenon. A common practice among colleges and universities is for the administration to use these SET scores as a diagnostic feedback tool for faculty (formative function) and as performance measures for personnel decisions such as hiring, tenure, promotion and salary reviews (summative function) (Emery et al., 2003; Bonitz, 2011). During the past decade or so, Johnson (2002) observes an increased emphasis on the use of SET as a measurement of teaching effectiveness across colleges and universities. In a feature for the Chronicle of Higher Education, Robin Wilson (1998, p. A12) stated that:

> "Only about 30 per cent of colleges and universities asked students to evaluate professors in 1973, but it is hard to find an institution that doesn't today. Such evaluations are now the most important, and sometimes the sole, measure of a teacher's teaching ability."

Given the prevalent use of SET across colleges and universities, there exists a substantial body of literature on student evaluations of teaching effectiveness. Eiszler (2002) argues that few topics in the popular and scholarly literature of higher education have attracted as much attention over prolonged periods as have the concern for the validity of student evaluations of college teaching effectiveness. While many researchers contend that SET scores obtained from students are valid and reliable measures of teaching effectiveness, there is still a large contingent that argue that the results from such instruments should not be relied upon for making personnel and tenure decisions (Sauer, 2012). Sproule (2000) and Bonitz (2011) argue that the use of SET for the purpose of formative and summative functions is controversial.

Both researchers and faculty alike have questioned the reliability and validity of the SET and their suitability for evaluation purposes. Johnson (2002) notes that the usage of SET to evaluate faculty teaching effectiveness raises

a number of concerns, including the basic validity of these surveys and their sensitivity to external biases. Yunker and Yunker (2003) argue that SETs are not clearly and directly related to teaching effectiveness. Instead, they are related and influenced significantly by personal characteristics and institutional factors over which the faculty members have no control. In support of Yunker and Yunker (2003), Bonitz (2011) notes that a large proportion of SET research has been devoted to the issue of bias in student evaluations of faculty. Although most university administrators believe that students can reliably evaluate teaching effectiveness, prior research in higher education literature, however, provides evidence that many faculty members believe that student evaluations are simply a popularity contest and have no relation to the measurement of effective teaching (Morgan et al. 2003).

Theoretically, student bias is an obvious possibility when adverse treatment (for example, due to grade disappointments or because of disciplinary actions against the student) is connected to a procedure in which anonymity provides a lack of rating accountability. A resulting student animus to the faculty can easily become translated into ratings that may have little to do with the faculty member's instructional effectiveness and actual performance in the course (Clardy, 2003). Centra (1993) defines bias in this context as "a circumstance that unduly influences a teachers' ratings, although it has nothing to do with the teacher's effectiveness" (p. 65). Being aware of the possibility of student bias, faculty can easily gravitate to a position of either complete indifference to student evaluations or to a modification of teaching practices (by limiting demands and assignments, grading easily, and so on) to increase the ratings faculty are given by their students (Clardy, 2003).

Some educators have voiced their concerns about how student evaluations affected the quality of education and resulted in grade inflation and lower academic performance (Stapleton & Murkison, 2001). Badri et al. (2006) empirically attempted to identify potential biasing factors in student evaluations and found that expected grade, actual grade, class size, course level, course timing, student gender and course subject significantly affect SET. Similarly, Sauer (2012) explored the predictors of student evaluations and revealed that student gender, ethnicity, age, prior interest in subject, course electivity and expected grade variables were statistically significant predictors of SET scores.

Although prior literature related to student evaluations of teaching has attempted to explore various student characteristics and contextual factors and potential biasing variables that may affect the legitimacy, validity and reliability of student evaluations, the majority of these studies have been conducted in the Western countries. The results of such studies might not be generalizable to a country like Kuwait that has a different social, cultural and educational setting. Based on this background, this study aims to investigate the determinants of SET scores among students attending the College of Business Studies at the Public Authority for Applied Education and Training (PAAET), Kuwait.

For this purpose, a SET form was constructed based on the official SET form developed by the Measurement and Evaluation Center at PAAET. The developed SET consisted of three parts. Part 1 contained demographic data and other background information about participating students. Part 2 contained questions covering a students' rating of the instructor. Part 3 contained one question that required student to report the expected final grade at the end of the course. In addition, a separate questionnaire that contained questions related to the instructor characteristics and course taught was administrated to each instructor under evaluation. Seven hundred and twenty SET were distributed in 21 various accounting classes offered to students.

Data collected was analyzed using quantitative methods. Descriptive statistics, independent sample t-tests, one-way analysis of variance (ANOVA) tests and multiple regression analyses were used to identify the determinants of SET. Multivariate regression analysis results clearly indicated that SET scores are significantly and positively biased by expected grade, student age and course level. In contrast, class size and faculty experience variables were found to be significantly and negatively related to SET. Interestingly, student gender and faculty gender variables were both negative; however, their influence on SET scores were insignificant.

The findings of this study make several important contributions. First, they raise concerns about the reliability and validity of the SET and their suitability for evaluation purposes. As SET scores have an important assessment function and serve as formative and summative measures in personnel decisions, the incentives for faculty to compromise their grading standards to receive good teaching evaluations increase. Accordingly, administrators should devote more effort to ensure a careful and complete understanding and interpretation of SET if they want to effectively incorporate them into the faculty evaluation process. Second, the findings are helpful to administrators in reviewing and interpreting SET scores. Third, the findings provide faculty with information about the potential determinants of student evaluations of teaching.

The remainder of this study is organized as follows. Section 2 reviews prior research related to student

evaluations of teaching. Section 3 outline methodology followed. The findings are presented in Section 4. Finally, section 5 presents the conclusions, contributions and implications.

2. Literature Review

The practices of using student opinions to evaluate the effectiveness of college teaching and studying the factors that affect it date back to the early 1900s (Algozzine et al., 2004). Today, student evaluation of teaching has become a routine and mandatory part of teaching in colleges and universities. In their review of student evaluation forms, Chulkov and Van Alstine (2010) show that student evaluations of teaching (SET) forms typically consist of questions that ask students to evaluate various aspects of the faculty's performance and course design. These forms are completed by the students at the end of the semester and serve as a summative measure used in personnel decisions about faculty hiring, tenure, promotion, and salary reviews. SET scores also have an important assessment function and are used as a formative measure by faculty seeking to improve their teaching skills and course design (Chulkov & Van Alstine, 2010). In a study surveying the use of student evaluations as a component of faculty evaluation systems in 600 liberal arts colleges between 1973 and 1993, Seldin (1993) finds that the use of SET increased from 28 percent to 86 percent over the study period. In 1999, nearly 90% of the colleges surveyed reported the use of SET (Seldin, 1999). Similarly, Wilson (1998) claims that it is hard to find colleges or universities that do not use SET in measuring teaching effectiveness. He argues that these evaluations are now the most important, and sometimes the sole, measure of a faculty member's teaching ability. Chulkov and Van Alstine (2010) claim that the international accrediting bodies have contributed to the increased emphasis on SET. For example, in its Accreditation Standard No. 12, the Association for Advancement of Collegiate Schools of Business (AACSB)-a leading accrediting body for business schools that has accredited more than 700 business schools in 48 countries-requires all business schools to "have a systematic program for evaluating instructional performance of faculty members. Information from instructional evaluation should be available to both faculty members and administrators. The school should use instructional evaluations as the basis for development efforts for individual faculty members and for the faculty as a whole" (AACSB, 2015).

However, the increased reliance on student evaluations as a measure of teaching effectiveness in personnel decisions about faculty hiring, tenure, promotion, and salary reviews has raised questions whether student feedback is a valid measure of effective teaching. Many faculty members argue that there is a number of potential biases in the use of student evaluations (Morgan et al., 2003). Similarly, Heine and Maddox (2009) note that these evaluations have been routinely criticized for being open to many sources of bias and error. Consequently, various faculty members have argued that students' evaluations are unfair because students rate some faculty members poorly as instructors because of the nature and amount of work assigned and the grades students earned (Stapleton & Murkison, 2001). Helterbran (2008) argues that the measurement of teaching effectiveness is a complex process that "involves the interweaving of content knowledge, pedagogy skills, and a knowledge and appreciation of the multifaceted nature of students to, in the end, be able to point to evidence that learning has occurred" (p. 126). Heine and Maddox (2009) and Morgan *et al*. (2003) argue that teaching effectiveness is a complex process to identify and nearly impossible to validly measure, and they question whether students have the capacity to actually evaluate teaching and teaching effectiveness.

Given the prevalence of SET use across colleges and universities, measurement and assessment of teaching effectiveness through student evaluations have been research topics for almost a century (Campbell & Bozeman, 2007). Similarly, Eiszler (2002) argues that few topics in the popular and scholarly literature of higher education have attracted as much attention over prolonged periods as the concern regarding the validity of student evaluations of college teaching effectiveness.

While many researchers contend that SET scores obtained from students are valid and reliable measures of teaching effectiveness, a large contingent still argues that the results from such instruments should not be relied upon for making personnel and tenure decisions (Sauer, 2012). For instance, using a sample from a Malaysian university, Liaw and Goh (2003) show that class size inappropriately influences students' judgments in teaching evaluations, suggesting that classes with small enrolment receive good teaching ratings, whereas large classes are associated with poor evaluation ratings. They further show that teaching evaluations are not significantly influenced by faculty or course characteristics. Similarly, Whitworth *et al.* (2002) analyze 12,153 student evaluations to investigate the effects of faculty gender, course type, and course level on student evaluations. Their results reveal that female faculty members were rated better than male ones and that ratings differed significantly by course type and students' perceived amount of learning. In addition, they show that graduate students tend to give higher SET scores than undergraduates. Based on data obtained from a university in Hong Kong, Kwan (1999) examines the effects of course characteristics on student evaluations, finding significant differences in student evaluations across academic disciplines, class sizes, course levels, type of course, and

modes of study.

In examining whether the use of student evaluations of teaching effectiveness have been a contributing factor to a trend of grade inflation at a mid-sized public university in the United States, Eiszler (2002) concludes that student evaluations may be used in ways that raise questions regarding consequential validity, specifically by encouraging grade inflation. Similarly, Millea and Grimes (2002) examine the link between course rigor and expected grades to evaluation scores and found that expected grades significantly and positively affect evaluation scores. In their study of grades, course evaluations, and academic incentives, Love and Kotchen (2010) develop a model that identifies a range of new and seemingly counterintuitive insights about how an institution's promotion criteria may affect student and faculty behavior. The results of their model show that placing more emphasis on student evaluations intensifies the problems of grade inflation and can even decrease a professor's teaching effort. Their findings suggest that an institution's efforts to improve teaching quality may adversely result in grade inflation.

In an attempt to identify potential biasing variables in students' evaluations of teaching in a newly AACSB accredited business school in the United Arab Emirates (UAE), Badri et al. (2006) provide evidence that supports previous research regarding the existence of potential biasing factors. Their results reveal that expected grade, actual grade, course level, class size, course timing, student gender and course subject significantly affect student evaluation of teaching. Due to the possible existence of biasing factors in SET, Badri et al. (2006) argue that comparing individual faculty members SET scores regardless of other factors might not be fair and call for the need to supplement the SET scores with other measures of teaching effectiveness.

Bonitz (2001) uses an experimental study to evaluate the influence of course type, instructor and student gender, and student individual differences on SET scores. The results reveal that student individual differences explained a significant proportion of the variance in SET scores. The most salient traits that were significantly related to SET scores were agreeableness, conscientiousness, conventional and investigative confidence, and gender role attitudes. In addition, the results of Bonitz's (2011) study show that female students gave significantly higher SET scores than male students independent of course type or instructor gender. Overall, the findings of Bonitz's (2011) study suggest that students' individual differences can bias SET scores, which poses a threat to the validity of the usefulness of student evaluations.

Although prior literature related to student evaluations of teaching has attempted to explore various student characteristics and contextual factors and potential biasing variables that may affect the legitimacy, validity and reliability of student evaluations, most of these studies have been conducted in the Western countries. The results of such studies might not be generalizable to such countries as the Gulf Countries which has a different social, cultural and educational setting. As mentioned above, Badri et al conducted their study on the UAE.

3. Methodology

3.1 The Instrument

The objective of this paper is to explore the determinants of student's evaluations of teaching. For this purpose, a student's evaluation of teaching (SET) form was constructed based on the official SET form developed by the Measurement and Evaluation Center at the Public Authority for Applied Education and Training, Kuwait. The developed SET consists of three parts. Part 1 contains seven questions obtaining demographic data and other background information about participating students. Part 2 contains thirty-two specific questions covering a students' rating of the instructor's presentation skills, time management, evaluation methods, fairness, class preparation and relationship with students. Responses on part 2 questions were based on a five-point Likert scale, with student answers ranging from 1 "*do not agree at all*" to 5 "*totally agree*." Part 3 contains one question that requires student to report the expected final grade at the end of the course. The developed SET content was validated by faculty members of the Public Authority for Applied Education and Training. In addition to the SET that was administrated to the students, a separate questionnaire was administrated to each instructor under evaluation containing five questions related to the instructor characteristics and course taught. The pilot study found that students easily understood the questions and had no difficulty in completing the SET in a reasonable period of time.

3.2 Data Collection

The population examined in this study is consisted of students attending accounting classes at the College of Business Studies, the Public Authority for Applied Education and Training in the fall and spring semesters of 2013-2014. The SET was administered by the researchers during accounting classes. Seven hundred and twenty SET forms were distributed in 21 various accounting classes offered to students. Of the 720 SET administered,

678 responses were considered appropriate for statistical investigation.

3.3 Analysis Methods

To identify the determinants of student's evaluations of teaching, data collected from the SET were analyzed using descriptive statistics, independent sample t-tests, univariate analysis of variance (ANOVA) tests, and multiple regression analyses. Based on the existing empirical literature, seven independent variables were identified to explore the determinants of student's evaluations of teaching. The regression model is as follows:

$$SET = \beta_0 + \beta_1\,(Expected\ Grade) + \beta_2\,(Student\ Gender) + \beta_3\,(Student\ Age) + \beta_4\,(Class\ Size) + B_5\,(Course\ Level) + \beta_6\,(Faculty\ Experience) + \beta_7\,(Faculty\ Gender) + \varepsilon \qquad (1)$$

Where

       SET = the student's evaluation of teaching score

       Expected Grade = the expected grade at the end of the course

       Student Gender = gender of the student

       Student Age = student age

       Class Size = number of students in the class

       Course Level = principle, for accounting major, or for non-accounting major

       Faculty Experience = years of experience

       Faculty Gender = gender of faculty

4. Results and Discussion

Table 1 outlines descriptive statistics for SET scores. Panel A indicates that the mean (median) SET score was 3.85 (3.95), with a minimum score of 1 and a maximum of 5. Panel B of Table 1 outlines the frequency distribution of SET scores. The statistics show that 10% of SET scores are between 1 and 1.99. Twenty-two percent of SET scores between 2 and 2.99, 20% scores between 3 and 3.99, and 37% between 4 and 4.99. Only 9% of the SET scores are five. Thus, the frequency distribution of SET scores reveals a noticeable variation in SET scores.

Table 1. Descriptive statistics for student's evaluation of teaching scores (SET)

Panel A: Descriptive Statistics for SET Scores								
Dependent Variable	N	Mean	Median	Standard Deviation	Min.	Max.	Skewness	Kurtosis
SET	678	3.85	3.95	1.09	1	5	−0.76	−0.58
Panel B: Frequency Distribution of SET Scores								
SET Range	Frequency		Percentage			Cumulative Percentage		
1-1.99	71		10			10		
2-2.99	163		24			34		
3-3.99	135		20			54		
4-4.99	250		37			91		
5	59		9			100		
Total	119		100					

Table 2 outlines a description of students and faculty's demographics. It reveals that the sample consisted of 181 male students (26.7%) and 497 female students (73.3%). Among the 678 students, 226 (33.3%) students expected to achieve an "A" grade at the course, 178 (26.3%) expected to achieve a "B" grade, 162 (23.9%) expected a "C" grade, while 46 (6.8%) expected a "D". Only 66 students (9.7%) expected an "F" grade at the course. Table 2 shows that 30.6% of the students examined were between 18 and 20 years old, 49.4% were between 21 and 23 years old, 9.6% were between 24 and 26 years old, 8.3% were between 27 and 29 years old and 2.1% were 29 years or older.

In addition, Table 2 outlines the frequency distribution of class size. It reveals that 22.9% of students examined were on a class size that ranges from 20 to 30 students, 46.3% were on a class size that ranges from 31 to 40 students, whereas 30.8% of students were on a class size that ranges from 41 to 50 students. Furthermore, the frequency presented on Table 2 shows that 52.2% were enrolling on an accounting principle course, 42% were enrolling on an accounting major course, while 5.8% were enrolling on an accounting course that is required for

non-accounting students.

Table 2. Description of students and faculty's demographics

Variable	Frequency	Percentage
Expected Grade		
A	226	33.3
B	178	26.3
C	162	23.9
D	46	6.8
F	66	9.7
Total	*678*	*100*
Student Gender		
Male	181	26.7
Female	497	73.3
Total	*678*	*100*
Age		
18-20	208	30.6
21-23	335	49.4
24-26	*65*	9.6
27-29	*56*	8.3
Above 29	14	2.1
Total	*678*	*100*
Class Size		
20-30	155	22.9
31-40	314	46.3
41-50	209	30.8
Total	*678*	*100*
Course Level		
Principle	354	52.2
For accounting major	285	42.0
For non-accounting major	39	5.8
Total	*678*	*100*
Faculty Experience		
Less than 1 year	58	8.6
1 to 5 years	296	43.6
5 to 10 years	154	22.7
More than 10	170	25.1
Total	*678*	*100*
Faculty Gender		
Male	*423*	*62.4*
Female	*255*	*37.6*
Total	*678*	*100*

With respect to faculty's demographics, the frequency distribution of faculty experience presented on Table 2 shows that 8.6% of the instructors included in this study have less than 1-year experiences, 43.6% have between one and five years experience, 22.7% have between five and ten years, while 25.1% have more than ten years teaching experience. Furthermore, Table 2 reveals that 62.4% of the instructors were male, while 37.6 were female.

Table 3 outlines the results of t-tests used to determine whether there were significant variations in SET scores with respect to student gender and faculty gender. The results outlined in Table 3 show that there were no significant variations in SET scores from male students ($M = 3.75$) and females ($M = 3.87$). Similar insignificant differences were observed in SET scores between male faculty ($M = 3.88$) and female faculty ($M = 3.78$).

Table 3. T-test for differences with respect to student's evaluation of teaching (SET)

Category	Variable	Mean	Standard Deviation	T-value	Significant level
Student Gender					
	Male	3.75	1.19	−1.26	None
	Female	3.87	1.05		
Faculty Gender					
	Male	3.88	1.09	1.05	None
	Female	3.78	1.08		

A one-way analysis of variance (ANOVA) test was used to investigate variations in SET scores with respect to expected grade, student age, class size, course level, and faculty experience. Untabulated results show significant variations ($p < 0.01$) in SET scores across expected grades. Similar significant differences ($p < 0.01$) were observed in SET scores across the student age categories ($p < 0.01$), faculty experience categories ($p < 0.01$). The results show that similar significant differences were also observed in SET scores across course level ($p < 0.01$) and class size ($p < 0.01$) categories.

Table 4 present Pearson's correlation and Spearman's rank correlation among the dependent and independent variables. The correlation matrix presented in Table 4 shows that no pair-wise correlation coefficient exceeds 0.70, suggesting that multicollinearity is unlikely to be a serious problem in interpreting the multiple regression results (Gujarati, 2003).Variance inflation factors (VIF) were also examined and found to be well within acceptable limits.

Table 4. Bivariate correlations among dependent and independent variables

Variable	SET	Expected Grade	Student Gender	Student Age	Class Size	Course Level	Faculty Experience	Faculty Gender
SET	1	0.64**	0.01	0.15**	−0.05	0.01	−0.06	−0.05
Expected Grade	0.63**	1	0.14**	0.10**	0.04	0.17**	0.13**	−0.04
Student Gender	0.05	0.13**	1	−0.08*	−0.26**	0.30**	0.32**	0.47**
Student Age	0.20**	0.15**	−0.14**	1	−0.05	−0.13**	−0.16**	0.003
Class Size	−0.08*	0.05	−0.22**	−0.06	1	0.40**	0.32**	−0.42**
Class Level	0.03	0.17**	−0.30**	−0.16**	0.41**	1	0.58**	−0.39**
Faculty Experience	−0.10**	0.06	0.18**	−0.18**	0.32**	0.61**	1	−0.31**
Faculty Gender	−0.04	−0.06	−0.47**	−0.02	−0.45**	−0.30**	−0.24**	1

Note. *, ** Correlation is significant at 0.05 and 0.01 levels, respectively (two-tailed). The upper-right diagonal presents Spearman's correlation and the lower-left diagonal presents Pearson's correlation of variables.

Table 5 outlines the multivariate regression analysis results. The table reveals that the multiple regression model is significant ($p < 0.000$, F = 77.143). Determinants identified as likely influencing student's evaluations of teaching in this study explain about 44% of the association between the student's evaluation of teaching (SET) scores and the potentially influential factors. According to the regression results presented in Table 5, the expected grade ($p < 0.01$), student age ($p < 0.01$), and course level ($p < 0.01$) variables are significantly and positively related to SET. In contrast, class size ($p < 0.01$) and faculty experience variables found to be significantly and negatively related to SET. Interestingly, the coefficient estimates of student gender and faculty gender variables are both negative, however their influence on SET scores are insignificant. In examining the influence of the identified variables influencing student's evaluations of teaching, the standardized coefficient beta presented in Table 5 reveals that expected grade has the strongest impact on SET scores, followed by faculty experience, followed by course level, followed by class size. The standardized coefficient beta shows that student age has the least effect on SET scores.

Table 5. Multivariate regression analysis results

$$SET = \beta_0 + \beta_1 (Expected\ Grade) + \beta_2 (Student\ Gender) + \beta_3 (Student\ Age) + \beta_4 (Class\ Size)$$
$$+ \beta_5 (Course\ Level) + \beta_6 (Faculty\ Experience) + \beta_7 (Faculty\ Gender) + \varepsilon$$

Variable	Unstandardized Coefficient		Standardized Coefficient	t-statistic	Significant
	B	Std. Error	Beta		
Intercept	2.102	0.370		5.687	***
Expected Grade	0.528	0.026	0.614	20.430	***
Student Gender	−0.103	0.098	−0.042	−1.049	None
Age	0.031	0.012	0.081	2.704	***
Class Size	−0.018	0.005	−0.134	−3.839	***
Course Level	0.342	0.105	0.189	3.244	***
Faculty Experience	−0.149	0.031	−0.238	−4.757	***
Faculty Gender	−0.096	0.088	−0.043	−1.094	None
N	R²	Adj.R²	F-statistic	p-value (F-statistics)	
678	0.446	0.440	77.143	0.000	

Note. *** indicates significant at the 0.01 level.

Our study confirms that expected grade is related to student evaluations. The results are consistent with other studies that expected grade positively affects student evaluations (For instance, Eiszler, 2002; Badri et al., 2006; Love & Kotchen, 2010), suggesting that students who assigned higher expected grade gave higher evaluations to faculty than students who assigned lower expected grades. Similarly, the findings support some prior research regarding the effects of course level on SET scores. The results suggest that higher SET scores are associated with higher courses levels, whereas lower SET scores are associated with introductory course levels. Badri et al. (2006) justify that observation by noting that senior students enjoy certain levels of maturity and experience with faculty to be more selective in courses enrollment.

Consistent with prior research that expected SET to be higher in small classes compared to larger classes (e.g., Kwan, 1999; Liaw & Goh, 2003; Bonitz, 2011), the results show that class size has inappropriately influenced students' judgments on teaching evaluations, suggesting that classes with small enrolment receive good teaching ratings, whereas large classes are associated with poor evaluation ratings.

As far as faculty experience, more experienced faculty is expected to promote better educational outcomes. The results show that faculty experience is significantly and negatively related to SET, suggesting that less experienced faculty receive significantly better SET than experienced faculty. One potential explanation for this finding is that the more experienced faculty may adhere more strictly to the curriculum and produce students with a deeper understanding of the material, whereas the less experienced faculty brings more energy and enthusiasm into the classroom, which may explain an inverse relation between age and teaching effectiveness (Carrell, 2010).

5. Conclusion

Given the importance and the heavy reliance on students' evaluations as a measure of teaching effectiveness, the aim of this study is to investigate the determinants of SET among students attending the College of Business Studies at the Public Authority for Applied Education and Training (PAAET), Kuwait. For this purpose, a SET form was constructed based on the official SET form developed by the Measurement and Evaluation Center at PAAET. The developed SET consisted of three parts. Multivariate regression analysis results clearly indicated that SET are significantly and positively biased by expected grade, student age and course level. In contrast, class size and faculty experience variables were found to be significantly and negatively related to SET. Interestingly, student gender and faculty gender variables were both negative; however, their influence on SET scores were insignificant. In contrast, student age had the least effect on SET scores. In examining the influence of the identified variables influencing student's evaluations of teaching, the findings reveals that expected grade has the strongest impact on SET scores, followed by faculty experience, followed by course level, followed by class size. The results show that student age has the least effect on SET scores.

The findings of our study are consistent with prior research and suggest that students who assigned higher expected grade gave higher evaluations to faculty than students who assigned lower expected grades. Similarly, the results suggest that higher SET scores are associated with higher courses levels, whereas lower SET scores are associated with introductory course levels. Furthermore, the results shows that class size has inappropriately influences students' judgments on teaching evaluations, suggesting that classes with small enrolment receive good teaching ratings, whereas large classes are associated with poor evaluation ratings.

In consistent with the conventional wisdom experienced faculty promotes better educational outcomes, the results show that faculty experience is significantly and negatively related to SET, suggesting that less experienced faculty receive significantly better SET than experienced faculty. One potential explanation for this finding is that the more experienced faculty may adhere more strictly to the curriculum and produce students with a deeper understanding of the material, whereas the less experienced faculty brings more energy and enthusiasm into the classroom which may explain an inverse relation between age and teaching effectiveness (Carrell, 2010).

The findings of this study make several important contributions. First, they raise concerns about the reliability and validity of the SET and their suitability for evaluation purposes. Given the increased emphasis on the use of SET as a measurement of teaching effectiveness in personnel decisions, the findings of this study raise a concern regarding faculty incentives to manipulate their gardening policies to receive good teaching evaluations, thus intensifying the problems of grade inflation. Administrators should devote more efforts to ensure a careful and complete understanding and interpretation of student evaluations if they want to effectively incorporate these student evaluations into faculty evaluation process. Second, the findings are helpful to the PAAET administrators in reviewing and interpreting SET scores. Third, the findings provide faculty with information about the potential determinants of student evaluations of teaching.

The primary limitation of this study was its sample size and diversity, as the data was collected from one business school. Future research may provide more generalizable results by expending the size and diversity of the sample. The most important conclusion from the results of this study is that expected grade had the strongest impact on SET scores. Further research is needed to explore the potential influence of student evaluations on the quality of education and the resultant grade inflation and lower academic performance. The variables used in the study were expected to explain differences in the levels of student evaluations of teaching. Future research might consider qualitative method that involves student and faculty interviews.

References

AACSB. (2015). *Eligibility Procedures and Accreditation Standards for Business Accreditation.* Retrieved from http://www.aacsb.edu/en/accreditation/standards/2013-business/

Algozzine, B., Beattie, J., Bray, M., Flowers, C., Gretes, J., Howley, L., et al. (2004). Student Evaluation of College Teaching: A Practice in Search of Principles. *College Teaching, 52*(4), 134-141.

Badri, M. A., Abdulla, M., Kamali, M. A., & Dodeen, H. (2006). Identifying Potential Biasing Variables in Student Evaluation of Teaching in a Newly Accredited Business Program in the UAE. *International Journal of Educational Management, 20*(1), 43-59. http://dx.doi.org/10.1108/09513540610639585

Bonitz, V. S. (2011). *Student Evaluation of Teaching: Individual Differences and Bias Effects.* (Unpublished PhD., Iowa State University, Ann Arbor).

Campbell, J., & Bozeman, W. (2007). The Value of Student Ratings: Perceptions of Students, Teachers, and Administrators. *Community College Journal of Research and Practice, 32*(1), 13-24. http://dx.doi.org/10.1080/10668920600864137

Carrell, S. E. (2010). Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors. *Journal of Political Economy, 118*(3). http://dx.doi.org/ 10.3386/w14081

Centra, J. A. (1993). *Reflective Faculty Evaluation.* San Francisco: Jossey-Bass.

Chulkov, D., & Van Alstine, J. (2012). Challenges in Designing Student Teaching Evaluations in a Business Program. *International Journal of Educational Management, 26*(2), 162-174. http://dx.doi.org/10.1108/09513541211201979

Clardy, A. (2003). Integrating Course and Instructional Evaluation with Learning History Approach. *Journal of the Scholarship of Teaching and Learning, 3*(3), 9-30.

Eiszler, C. F. (2002). College Students' Evaluations of Teaching and Grade Inflation. *Research in Higher Education, 43*(4), 483-501.

Gujarati, D. N. (2003). *Basic Econometrics.* New York: McGraw-Hill.

Heine, P., & Maddox, N. (2009). Student Perceptions of the Faculty Course Evaluation Process: An Exploratory Study of Gender and Class Differences. *Research in Higher Education Journal, 3,* 1-10.

Helterbran, V. R. (2008). The Ideal Professor: Student Perceptions of Effective Instructor Practices. *Attitudes and Skills. Education, 129*(1), 125-138.

Johnson, V. E. (2002). Teacher Course Evaluations and Student Grades: An Academic Tango. *Change, 15*(3), 9-16. http://dx.doi.org/ 10.1080/09332480.2002.10554805

Kwan, K. (1999). How Fair are Student Ratings in Assessing the Teaching Performance of University Teachers? *Assessment & Evaluation in Higher Education, 24*(2), 181-195. http://dx.doi.org/10.1080/0260293990240207

Liaw, S., & Goh, K. (2003). Evidence and Control of Biases in Student Evaluations of Teaching. *International Journal of Educational Management, 17*(1), 37-43. http://dx.doi.org/10.1108/09513540310456383

Love, D. A., & Kotchen, M. J. (2010). Grades, Course Evaluations, and Academic Incentives. *Eastern Economic Journal, 36,* 151-163. http://dx.doi.org/10.1057/eej.2009.6

Millea, M., & Grimes, P. W. (2002). Grade Expectations and Student. *Evaluation of Teaching College Student Journal, 36*(4), 582-590.

Morgan, D. A., Sneed, J., & Swinney, L. (2003). Are Student Evaluations a Valid Measure of Teaching Effectiveness: Perceptions of Accounting Faculty Members and Administrators. *Management Research News, 26*(7), 7-32. http://dx.doi.org/10.1108/01409170310783565

Pounder, J. S. (2007). Is Student Evaluation of Teaching Worthwhile? An Analytical Framework for Answering the Question. *Quality Assurance in Education: An International Perspective, 15*(2), 178-191. http://dx.doi.org/10.1108/09684880710748938

Robin, W. (1998). New Research Casts Doubt on Value of Student Evaluations of Professors. *Chronicle of Higher Education, 44*(19), A12.

Sauer, T. M. (2012). *Predictors of Student Course Evaluations* (Unpublished PhD., University of Louisville, Ann Arbor).

Seldin, P. (1993). How Colleges Evaluate Professors: 1983 Versus 1993. *AAHE Bulletin, 12*, 6-8.

Seldin, P. (Ed.). (1999). *Building Successful Teaching Evaluation Programs.* In Changing Practices in Evaluating Teaching: A Practical Guide to Improved Faculty Performance and Promotion/Tenure Decisions. Bolton, MA: Anker.

Sproule, R. (2002). The Underdetermination of Instructor Performance by Data from the Student Evaluation of Teaching. *Economics of Education Review, 21*(3), 287-294. http://dx.doi.org/10.1016/S0272-7757(01)00025-5

Stapleton, R. J., & Murkison, G. (2001). Optimizing the Fairness of Student Evaluations: A Study of Correlations between Instructor Excellence, Study Production, Learning Production, and Expected Grades. *Journal of Management Education, 25*(3), 269-291.

Whitwortha, J., Barbara, A. P., & Randalla, C. H. (2002). Factors That Affect College of Business Student Opinion of Teaching and Learning. *Journal of Education for Business, 77*(5), 282-289. http://dx.doi.org/10.1080/08832320209599677

Wilson, R. (1998). New research casts doubt on value of student evaluation of professors. *The Chronicle of Higher Education, 16,* A12-14.

Yunker, P. J., & Yunker, J. A. (2003). Are Student Evaluations of Teaching Valid? Evidence from An Analytical Business Core Course. *Journal of Education for Business, 78*(6), 313-317. http://dx.doi.org/10.1080/08832320309598619

Copyrights