

Students' Perceptions of Student Evaluation of Teaching (SET) Process

By

¹Ale J. Hejase, ¹Rana S. Al Kaakour, ¹Leila A. Halawi, and ²Hussin J. Hejase

¹School of Business, Lebanese American University, Beirut, Lebanon

²Faculty of Business and Economics, American University of Science and Technology
Beirut, Lebanon

Abstract

Researchers have mixed views about Student Evaluation of Teaching (SET) as means to evaluate teaching where some agreed and others viewed SET as being biased. This study aims to measure students' perceptions of the effectiveness and appropriateness of the evaluation process in Lebanon. A survey questionnaire was administered to students from five Lebanese universities. Findings revealed that students were positive and perceived the evaluation process as effective and appropriate to evaluate teaching. Students identified students' perceptions, instructors' behavior, and course characteristics as variables that may impact the process. Results and implications were discussed for future research.

Keywords: Gender bias, students' evaluation of teaching, students' perceptions, teaching effectiveness, Lebanon.

1. Introduction

Student Evaluation of Teaching (SET), also referred to as student ratings, has been an area of interest for many researchers (Murray, Rushton, & Paunonen, 1990; Marsh & Bailey, 1993; Centra, 2003; Isely & Singh, 2005; Surratt & Desselle, 2007; Young, Rush, & Shaw, 2009; Heine & Maddox, 2009; Kozub, 2010). In recent years, SET has been established in many countries worldwide such as US, UK, Canada, Australia, as well as many European countries, but it is still a humble experiment in the Arab world's private and public universities. However, in Lebanon, universities have established their own SET forms.

Although SET is applied in many universities, and many studies have investigated the factors that affect students' ratings of their professors, students' participation in the teaching evaluation process is still a controversial issue. Some studies indicated positive validity and reliability (Marsh, 1987; Marsh & Bailey, 1993; Centra, 2003; Thornton, Adams, & Sepehri, 2010) while others showed that, due to many factors affecting students' assessment, SET does not adequately measure teaching effectiveness (Hamermesh & Parker, 2003; Kidd & Latif, 2004; Isely & Singh, 2005; Weinberg, Hashimoto, & Fleisher, 2009; Brockx, Spooren, & Mortelmans, 2011).

In light of the above findings, this study was devised to shed the light on the experience of universities in Lebanon in the field of Student Evaluation of Teaching. Such a study demands designing a survey questionnaire administered to a selected sample of students from various universities in Lebanon. The data gathered was then analyzed using the appropriate statistical techniques in order to answer the following questions:

- What are students' perceptions of the SET process?
- What is the impact of students' expectations of grades on SET scores?
- What is the average response rate to fill out SET forms?
- Are there differences in students' ratings based on professors' gender?
- How are course workload and students' ratings related?
- Is there any relationship between course difficulty and students' ratings?
- Did professor's appearance and attractiveness affect student's ratings?

- Are SETs valid measures of course effectiveness?
- What are the factors that most influence students' opinions?

As SET is becoming central in the teaching evaluation process in many universities in Lebanon, this subject could not be ignored. A study that can shed light on SET experience in Lebanon is needed because it has implications on SET practices. Findings of the study may contribute to the improvement of SET systems currently used in Lebanon by various universities. Research implications may help the Lebanese universities in developing clear and well-designed SET forms and therefore finding ways to increase students' participation in completing the surveys.

2. Literature Review

A large body of research exists in the area of SET. Some studies showed that SETs are valid measures of teaching effectiveness and are unaffected by variables identified as potential biases to the evaluation (Marsh & Bailey, 1993; Centra, 2003; Thornton et al, 2010). While other researchers viewed SETs as being inadequate measures of teaching effectiveness and suggested that SETs are biased by many variables (Basow, 1995; Isely & Singh, 2005; Weinberg et al, 2009). In addition, many faculty members complain that SETs are not reliable, they do not have any meaning, arguing that students will give favorable scores with higher grades and less workload. The following literature review will help the development of the necessary concepts to carry out the current research.

A study completed at the American University of Beirut (AUB) in 2005, addressed faculty and students perceptions of SET. Results showed that 66% of students have a positive view of the SET process. 81% of faculty did not agree that SET results should be used for taking decisions related to promotion and salary. Furthermore, faculty believed that students' ratings are affected by the course workload and that instructors may change their behavior to receive higher scores (ICE Report, 2005). While Surratt & Desselle (2007) found that students viewed SET as appropriate (92.5%) and necessary (95.5%) but admitted that the faculty members receiving the best evaluations were not always the most effective teachers (50.4%). Most students indicated a willingness to complete the Teaching Effectiveness Questionnaire (TEQ) when given the opportunity (80%) but expressed frustration that their feedback did not appear to improve subsequent teaching efforts. In addition, students acknowledged that the professors' personality affected their TEQ responses (51%), and even influenced their decision on whether to complete the TEQ (57.3%).

Heine and Maddox (2009) found that female students are more serious with the faculty course evaluation process than did their male counterparts and females perceived the evaluation process as more important than males in the surveyed sample. Moreover, male students reported a more negative view of the evaluation process than did female students and they believed that the higher the grade they expected, the higher the ratings on SET. Also, they believed that professors changed their behavior at the end of the semester in order to receive higher ratings.

Concerning the factors that may affect students' ratings of their instructors, Thornton et al (2010) studied "the impact of students' expectations of grades and perceptions of course difficulty, workload, and pace on faculty evaluations" and concluded that SET is not affected by grading, workload, or pace. While, Isely and Singh (2005), who addressed the impact of grades on SET, revealed that when students expect higher grades, they tend to give more favorable SET scores, that better students tend to provide less favorable SET, and that the instructor receives higher SET scores once the expected grade increases relative to cumulative GPA. As for Kidd and Latif (2004), they found a statistically significant and positive correlation between students' expected grades and course evaluation score. Similarly, Weinberg et al (2009) inferred that SET scores were positively correlated with current grades but unrelated to learning. Moreover, they found no evidence of a relation between learning and evaluations. Greenwald

and Gillmore (1997), who studied the effect of the instructor's grading leniency on SET ratings, found that the courses that gave higher grades were better rated by students.

Furthermore, Hamermesh and Parker (2005) compared students' ratings of 94 professors based on their good looks and appearance versus their scores received on the courses they have taught and found that SET scores increased by 1 point for the professors who had been rated among the most beautiful/handsome. Abrami, Leventhal, and Perry (1982), who studied the effect of the instructor's personality on student ratings of instructor, argued that "*instructional ratings should not be used in decision making about faculty promotion and tenure because they are affected by the instructor attributes with charismatic and enthusiastic faculty receiving more favorable student ratings regardless of how well they know the subject matter*". While, Jones (1989) examined the influence of teacher's personality on SET ratings and found that "*students' perceptions of the teacher's personality is very significantly related to their ratings of teaching effectiveness*".

Kozub (2010) found that students' ratings are affected by the course type and the instructors' characteristics. Students tend to give higher ratings for elective courses as well as for the instructor's appearance/attractiveness. Additionally, he found a significant correlation between the instructor's gender and students' evaluation of teaching with male instructors receiving lower evaluations than their female counterparts and that the instructors' age was unrelated to students' ratings. As for the interest in the course content, it was strongly related to the overall evaluation of teaching effectiveness. In another study, Darby (2006) found that the elective courses are rated higher than the required courses. Murray et al (1990) investigated the effect of teacher personality traits on student instructional ratings and found that "*the rated teaching effectiveness varied substantially across different types of courses for a given instructor, and that teaching effectiveness in each type of course could be predicted with considerable accuracy from colleague ratings of personality*".

Furthermore, Young et al (2009) revealed that "*gender bias plays a role in students' views of effective teaching in terms of how students evaluate pedagogical and content characteristics*". Bachen, McLoughlin, and Garcia (1999) as well as Basow (1995) asserted that female students rated female instructors higher than male instructors across five teaching dimensions, whereas the ratings given by male students were not affected by the instructor's gender.

In spite of the large body of research conducted to measure the effectiveness of SET process, the literature shows that the majority of studies have been conducted outside Lebanon, with only one study published by the American University of Beirut (AUB). However, this study was dedicated to measure solely the perceptions of AUB students and instructors without sharing the opinions and perceptions of other Lebanese students. The lack of studies in Lebanon was mainly the stimulating force to go forward with the current research.

3. Methodology

The current research is exploratory in nature. It uses a survey questionnaire divided into five sections. Section one includes seven statements which addressed students' opinions regarding the format and content of the SET questionnaire used as means to evaluate professors in their universities. Section two includes seven statements which explore issues related to the overall students' perceptions of the SET process. Section three consists of eleven statements aimed at identifying students' seriousness when completing the SET questionnaire and the reasons that drive them to complete such questionnaire. Section four has seven statements aimed at identifying whether students' ratings were affected by the instructor's personality, traits, appearance, and position, by course workload, or by students' expectations of grades. All four sections used a 5-level Likert scale criteria such that, SD: Strongly Disagree; D: Disagree; N: Neutral; A: Agree; and SA: Strongly Agree. The fifth section is about demographics and uses multiple choice, dichotomous, and open-ended questions.

Sampling Procedure

Researchers distributed 500 questionnaires to a convenient sample of university level students. However, 418 valid questionnaires were received from students from five different universities in Lebanon. These universities are the American University of Beirut (AUB), the Lebanese American University (LAU), the American University of Science and Technology (AUST), the Lebanese International University (LIU), and Haigazian University (HU). Respondent students were chosen based on their willingness to participate. Questionnaires were handed in classrooms, libraries, and university cafeterias. The final response rate was 83.60% which is considered high and suitable for the purpose of the research. Disqualified questionnaires included incomplete and wrongly filled questionnaires.

4. Results and Findings

Validity and Reliability

In order to test the validity, an exploratory factor analysis (EFA) using principal components factor analysis (PCA) with varimax rotation was performed on each predefined multi-item construct. From the factor loadings produced by the rotations, those loadings that are greater than 0.5 are considered significant (Hair, Anderson, Tatham, & Black, 1998; Changing Minds, 2012). By examining the rotated component matrix and following an iterative approach, the items that load heavily on more than one factor were dropped. The iterative process continued until a meaningful factor structure (a component matrix of one component only) was obtained. This was performed for each section separately. As a result, using factor analysis on each of the first four sections demonstrated homogeneity of the Likert scales due to the fact that only one significant component was extracted that reflects a unitary attitude which is the core requirement for construct validity. Table 1 includes the details of the loadings for each of the factors for the four sections where the single unitary component was extracted.

Table 1: Loadings of the Unitary Factor for Each Section of the Survey

Component Matrix ^a		Component Matrix ^a		Component Matrix ^a		Component Matrix ^a	
	Component		Component		Component		Component
	Section I		Section II		Section III		Section IV
Item_2	.724	Item_11	.725	Item_16	.687	Item_28	.843
Item_3	.735	Item_12	.556	Item_17	.713	Item_29	.822
Item_6	.719	Item_13	.702	Item_18	.628	Item_30	.854
Item_7	.535	Item_14	.783	Item_19	.673	Item_31	.847
				Item_24	.683	Item_32	.800

Extraction Method: Principal Component Analysis.

a. 1 component extracted.

In regards to reliability, an assessment of the internal consistency of each survey set of items was performed, essentially assessing whether all the items belonging to one set were measuring the same thing by using Cronbach's alpha technique. Reliability increases when the alpha value approaches 1. An alpha value of 0.8 or above is regarded as highly acceptable for assuming homogeneity of items (Burns & Burns, 2008), while an alpha value that is greater than 0.7 is considered appropriate even though this value could be as low as 0.6 for exploratory research (Hair et al, 1998; Nunally, 1978). The resultant Cronbach's alphas were 0.631, 0.644, 0.689, and 0.897 for sections I, II, III, and IV respectively which are appropriate as measures of internal reliability for the attitude scale in each section.

Demographic Analysis

Respondents were 53.8% females and 46.2% males and 78.2% of them were within the 18 to 22 years age group. As for the educational level, 89.2% of the respondents were undergraduate students and only 10.5% were graduate students. Participants belonged to five universities such that: 20.3% from AUB, 20.6% from LAU, 21.3% from LIU, 19.1% from AUST, and 18.7% from HU.

Moreover, results show that 39.7% of the students have completed 61 to 90 credits, 29.4% have completed 30 to 60 credits, 14.1% have completed less than 30 credits, 6.5% have completed 91 to 120 credits, and only 0.7% of students have completed more than 120 credits. These results reveal that the majority of students have completed the SET questionnaire many times and are familiar with the SET process.

Descriptive Analysis

Jamieson (2004) stated clearly that Likert scales fall within the ordinal level of measurement. Moreover, her paper emphasizes the fact that the responses in Likert scales cannot have equal intervals between the pairs of adjacent Responses. A response to Jamieson's article was published by Pell (2005) where the conclusion was that it is acceptable in many cases to consider Likert scales' responses as interval levels of measurement, in particular when the data is of appropriate size and shape. This same argument is supported by (Burns and Burns, 2008, p. 475); where they agree that many attitude investigators do consider Likert scales to be interval levels of measurements especially when their sample is large and randomly selected. On this basis the Likert scale was treated as an interval scale thus allowing the calculations of means and standard deviations. Tables 2 to 5 show the means, standard deviations, and results of the four sections of the questionnaire.

Table 2: Descriptive analysis of 11 items in section 1

Survey Items	Overall Response, Mean (Std Deviation)	SD	D	N	A	SA
Section I - SET format and content						
1. The SET Questionnaire administered in my university is an effective and appropriate mean for the evaluation of teaching	3.43 (0.968)	3.1	12.2	36.4	35.4	12.9
2. The SET Questionnaire is well designed and it uses specific and clear items	3.7 (0.833)	2.2	6.9	19.9	59.8	10.5
3. The items included in the SET questionnaire are within my ability of judgment	3.84 (0.909)	1.2	6.2	24.2	43.3	24.6
4. SET Questionnaire is too long so that I feel bored when completing it	3.56 (1.055)	3.6	13.4	25.6	38.3	18.9
5. The items included in the SET questionnaire do not cover all evaluation criteria	3.32 (1.034)	2.6	20.6	31.6	30.4	13.4
6. The items included in the SET questionnaire are relevant to evaluate what is addressed in the classroom	3.66 (0.872)	2.2	7.4	25.6	51.7	12.7
7. The rating scale used in the SET questionnaire is understandable	3.89 (0.906)	1.2	6.2	21.5	45	26.1

Results from Table 2 show that 48.3% of the respondents perceived the SET questionnaire as being effective and appropriate (Item 1), 70.3% said that SETs are well designed and clear (Item 2), 67.9% believe they are within their ability of judgment (Item 3), 64.4% felt they are relevant to evaluate what is addressed in the classroom (Item 6), and 71.1% accepted that a SET uses understandable rating scale (Item 7). However, 57.2% of the respondents perceived the SET questionnaire as being too long (Item 4) and 43.8% manifested that the questionnaire does not cover all evaluation criteria (Item 5).

Table 3: Descriptive analysis of 11l items in section 2

Survey Items	Overall Response, Mean (Std Deviation)	SD	D	N	A	SA
Section II - Students' perceptions of the SET process						
8. SET feedback is being considered and adopted by instructors to improve their teaching efforts	3.24 (1.076)	6	16.7	37.6	25.6	13.6
9. I believe that students have the ability to judge their instructors appropriately	3.7 (0.944)	2.4	10.3	18.7	52.2	16.5
10. I believe that the instructors consider the SET scores to make improvements in their teaching and courses	3.37 (1.071)	4.3	16.5	33	29.9	16
11. Some professors will give lower grades as a result of poor SET scores	3.12 (1.169)	9.8	21.8	25.6	31.1	11.2
12. Higher SET scores do not necessarily mean most effective teaching	3.61 (0.962)	2.9	9.6	27	44	15.8
13. Some professors change their behavior and teaching practices in order to receive more favorable SET scores	3.37 (1.012)	2.9	18.9	28.2	37.8	11.7
14. It is possible that the professor may retaliate on the final exam after receiving poor SET scores	3.32 (1.143)	7.9	14.4	30.6	30.6	15.8

In addition, Table 3 shows that 68.7% of the students stated that they have the ability to judge their instructors (Item 9), 59.8% did not believe that SET scores reflect effective teaching (Item 12), and 45.9% of them believed that instructors consider SET scores to make improvements in their teaching and courses (Item 10), 49.5% believed that instructors change their behavior in order to receive more favorable SET scores (Item 13) and even retaliate on the final exam after receiving poor SET scores (Item 14). Moreover, 46.4% of the students believed that some professors will give lower grades as a result of poor SET scores (Item 11).

Table 4: Descriptive analysis of 11l items in section 3

Survey Items	Overall Response, Mean (Std Deviation)	SD	D	N	A	SA
Section III - Students' response rate and seriousness						
15. I complete the SET form every time I am given the opportunity to do so	3.54 (1.066)	5	10	30.1	35.6	19.1
16. I tend to complete the SET questionnaire only because it is required by the university	3.52 (1.037)	1.4	20.8	17.5	43.8	15.8
17. I tend to complete the SET forms only for courses that I am interested in	3.27 (1.233)	7.4	24.9	19.4	29.4	18.7
18. I tend to complete the SET Questionnaire when I find the instructor's performance is especially well	3.31 (1.119)	4.8	22.2	23.9	33.5	14.6
19. I don't feel free to write my comments for fear of being identified and losing grades	2.86 (1.276)	18.4	22.7	23	24.6	10.5
20. I use SET questionnaire as a mean to indicate suggestions and to identify instructors' weaknesses for improvement	3.66 (0.973)	2.4	10.8	22.7	45.5	17.7
21. I tend to take the SET seriously	3.48 (1.075)	6.5	11	24.6	42.8	14.6
22. When completing the SET questionnaire, I give fair and accurate ratings based on what I have learned	3.77 (0.906)	1.9	7.2	22.2	49.5	19.1
23. I tend to complete the SET Questionnaire when I find the instructor's performance is especially poor	3.55 (1.105)	3.8	15.3	24.9	34.2	21.8
24. When completing the SET Questionnaire, I feel I am wasting time	3.18 (1.169)	7.2	25.1	23.4	30.1	13.9
25. I feel comfortable giving a negative evaluation for a bad professor	3.74 (1.172)	5.7	11.2	17	35.6	30.4

Furthermore, Table 4 shows that 54.7% of the students tend to complete the SET questionnaire (Item 15) but 59.6% of them participate in the SET process only because it is a university requirement (Item 16). In addition, 57.4% of them stated that they are serious when participating in the SET process (Item 21), 68.6% give fair and accurate ratings (Item 22), 66% do not hesitate to give negative evaluation for a bad professor (Item 25), and 63.2% use the SET questionnaire as a mean to provide suggestions and to identify instructors' weaknesses for the sake of instructors' improvement (item 20). Also, an equal number of students (201 students forming 48.1%) agreed that the factors that drive them to participate in SET process are the instructor's performance (Item 18) and the course type (Item 17).

Table 5: Descriptive analysis of 11l items in section 4

Survey Items	Overall Response, Mean (Std Deviation)	SD	D	N	A	SA
Section IV - factors that may affect students' ratings						
26. My SET scores are influenced by the professor's personality and traits	3.18 (1.066)	6.2	20.8	31.1	31.6	9.6
27. I tend to give higher SET scores when I expect higher grades	2.88 (1.168)	12.2	29.7	23	26.8	7.9
28. My SET scores are influenced by the professor's gender	2.72 (1.458)	26.6	25.4	14.1	15.3	17.7
29. I tend to give higher SET scores in courses that require less workload	2.77 (1.242)	17	29.2	22	21.1	9.8
30. I tend to give higher SET scores for the professors who have good appearance	2.65 (1.251)	23	25.1	23	21.1	7.4
31. I believe that professors who taught required courses perform better than those who taught elective courses	2.84 (1.27)	17.2	26.6	20.6	23.9	10.5
32. I believe that a senior professor should not undergo SET evaluation as applied to junior (fresh graduate) professor	2.79 (1.342)	22.5	21.3	23.9	18.4	13.4

Finally, Table 5 shows that 41.9% of the students believed that SET scores are not affected by higher grades expectations (item 27), gender (Item 28, 52%), course workload (Item 29, 46.2%), professor's appearance (Item 30, 48.1%), type of course - required or elective (Item 31, 43.8%), and professor's rank (Item 32, 43.8%), and 41.2% of the students believed that students' ratings are influenced by the professor's personality and traits (item 26).

Chi-Square Crosstab Test to Assess Students' Responses Based on Their Gender and University

When considering the differences in students' responses based on gender, three significant differences emerged (p -value < level of significance taken as 5%). Female students were surer that the items included in the SET questionnaire are within their ability of judgment (item 3). In addition, they tend to give higher scores for professors who have good appearance (item 30) than did their male counterparts in the sample. Gender differences suggest also that female students have a higher tendency than male students to believe that professors who taught required courses perform better than those who taught elective courses (item 31).

Investigating differences between students' responses by university revealed significant differences on all survey items, i.e. there is a significant relation or dependency between the respondents and their universities. Such dependencies cannot be identified using a non-parametric test such as the Chi-Square, but can be easily found using post-hoc tests with a one way ANOVA.

One Way ANOVA Analysis: To Assess Differences in Students' Responses Based on Their University

In order to identify the nature of the dependencies between responses and among universities, a One Way ANOVA was performed to assess if the means of the Likert scales for every item under each university

can be considered statistically the same. The One Way ANOVA revealed significant differences among the universities for all the items of the survey (all p-values of the corresponding F-tests were in the range 0.000 to 0.011). To understand where these differences were significant, multiple comparisons of item means for the different universities were performed using the Scheffe post-hoc test. For each survey item, the multiple comparisons tables showed the difference in means between the two groups (Mean-Difference) in addition to the significance of this difference in means (sig. column). Details of the post-hoc test implied the following:

Students' perceptions of the effectiveness, format, and content of SET questionnaire

Results showed significant differences in responses among university groups. AUST students perceived the SET process as being more effective and appropriate (Item 1) than did students in LAU, LIU, and HU. Additionally, AUST and LIU students had a more positive view of the clarity of items included in the SET questionnaire (Item 2) than did students in LAU and HU. In regards to LIU students, they perceived SET items as being more relevant to evaluate what is addressed in the classroom (Item 6), more understandable (Item 7), and they were more certain that these items are within their ability of judgment (Item 3) than did their counterparts in LAU and HU. Moreover, LIU and AUST students were less comfortable with SET questionnaire length (Item 4), and LIU students believed that SET questionnaire (Item 5) do not cover all evaluation criteria more than did HU students.

Students' perceptions of the SET process

AUST students perceived SET feedback as being more adopted by instructors than did the students in the other universities (item 8 and item 10). In addition, LIU and AUST students had a more negative view of the SET process than did students in AUB, LAU and HU in that instructors will give lower grades as a result of poor students' ratings (item 11) and that they will retaliate on the final exam after receiving poor SET scores (item 14). Additionally LIU students had a more negative view in that higher SET scores do not necessarily reflect effective teaching (item 12) than did AUST and HU students and that some professors change their behavior in order to receive more favorable ratings (item 13) than did their counterparts in the other universities.

Students' response rate and seriousness

Results revealed that AUST had the highest response rate (Item 15). On the other hand, LIU students tend to participate in the SET process because it is a requirement (Item 16). Moreover, they are more affected by the instructor's performance (items 18 and 23) and their interest in the course (item 17) when deciding to participate in the SET process and they are less free to write their comments (item 19) than did students in the other universities.

In regards to students' seriousness (item 21), the results revealed that LIU students are more serious than AUB and AUST students. As for AUST students, their ratings tend to be less fair and accurate (item 22) than those of LAU, LIU, and HU students. And finally, LAU students feel more comfortable giving a negative evaluation for a bad professor (item 25) than did LIU students.

Factors that may affect students' ratings

Results showed that AUST students' ratings are less affected by the professor's personality and traits when compared to students' ratings in the other universities contrary to AUB and LAU students' ratings, which are more affected by the aforementioned factor (Item 26). In addition, AUST students' ratings are less affected by higher grades expectations (Item 27) contrary to LIU students' ratings, which are the mostly affected by this factor.

Concerning LIU students, their ratings are also more affected by the professor's gender (item 28), course workload (item 29), professor's appearance (item 30), course type (item 31), and professor's rank (item 32) than did their counterparts in the other universities.

4. Conclusions and Implications

This study was devised to shed the light on Students' Evaluation of Teaching (SET) experience in Lebanese universities and has attempted to measure students' perceptions of SET process. Valuable information was revealed.

Findings showed that the majority of students have a positive view of the evaluation process in terms of format and content of SET questionnaire. They perceived the questionnaire as being an effective and appropriate means to evaluate instructors, well designed and clear, relevant to assess what is addressed in the classroom, within their ability of judgment, and uses understandable rating scale. These findings are consistent with the results obtained from previous studies (ICE, 2005; Surratt & Desselle, 2007). On the other hand, students complained that the questionnaire is too long and that it does not cover all evaluation criteria.

As for students' perceptions of the outcomes of the SET process, results showed that the majority of students do not trust SET ratings and have a negative view in terms of instructor's behavior toward the evaluation process. They reported that SET scores do not necessarily reflect effective teaching as reported by other studies (Surratt & Desselle, 2007), and that instructors may change their behavior in order to receive more favorable scores and may retaliate on the final exam after receiving poor ratings. On the other hand, Lebanese students have a positive view towards the use of SET outcomes to improve professor's future teaching efforts whereas other studies reported that SET feedback do not appear to improve teaching efforts (ibid).

Concerning students' decision to participate in SET process and their seriousness, the findings suggested that, on the average, 60% of the students tend to complete SET surveys only because these are required by the university, which is in contradiction with the findings of previous studies (ibid). However, they stated that they are serious when they complete the survey, they give fair and accurate ratings, do not hesitate to give negative evaluation for bad professors, and participate in the evaluation process in order to indicate suggestions and to identify weaknesses for improvement. Additionally, Lebanese students' participation in the evaluation process increases when they find the instructor's performance especially poor while Surratt and Desselle (2007) reported that Duquesne University students admitted that their decision to participate in the evaluation process is affected by the professor's personality.

Moreover, students believed that their ratings are not affected by professor's gender, professor's appearance, course type, professor's rank, grades, and course workload. The last two findings support views expressed by Thornton et al (2010). While the first and third findings are not consistent with the results of the studies that reported a significant dependency between SET scores and instructor's gender (Basow, 1995; Bachen et al, 1999; Young, Rush, & Shaw, 2009; Kozub, 2010), and the studies that indicated a significant correlation between SET ratings and the course type (Murray et al, 1990; Kozub, 2010).

In addition, students reported that their ratings are affected by professors' personality and traits which is supported by the previous studies (Murray et al, 1990; Hamermesh & Parker, 2005; Kozub, 2010).

When considering the differences in students' responses, the key finding was that the responses to all survey items depend on the specific university in question. Other influencing factors include students' decision to participate in the evaluation process, their seriousness, and the factors that may affect their ratings for example, the professor's personality, higher grades' expectations, professor's gender, course workload, professor's appearance, course type, and professor's rank.

As for gender differences, female students perceived SET survey as being more relevant to the evaluation and more understandable than did their male counterparts. In addition, they are surer of their ability to

judge their instructors. These findings are consistent with the results obtained from the study devised by Heine and Maddox (2009).

5. Recommendations

This research has attempted to study the effectiveness of students' Evaluation of Teaching from students' perspectives. In general, the findings provided a partial support that students bring a fair amount of accuracy to the evaluation process and take the process seriously despite the misperceptions of some. Therefore, universities need to motivate students and convince them that their opinions and participation in the SET process are valuable and essential to improve future teaching efforts.

There is only one study that has addressed faculty and student perspectives on student teaching evaluations in Lebanon (ICE Report, 2005). It is important to note that the results of the current research will provide exploratory findings that can be used by other researchers, Middle Eastern or others; consequently, cross-cultural comparisons could be performed. Moreover, another contribution of the current study is its stimulating effect that might lead others to test effectiveness of the SET process.

However, the researchers had two limitations. The sample surveyed in this study is limited to Lebanese students in five universities therefore the results cannot be generalized to all Lebanese students. In addition, instructors' perceptions and opinions towards the SET process were not examined and included in the current study.

Other implications from the current research stress that the evaluation process is complicated and the evaluation survey differs from one university to another. Therefore, the challenge for future research is to continue with the study of the effectiveness and validity of the SET process. There are many ways to address this subject, for instance, a content analysis of SET survey instruments completed by students for the past years coupled with relevant statistical tests could provide valuable information pertaining to the evaluation process. This could be done by each university and results will contribute to the improvement of SET questionnaires as well as finding ways to encourage student participation in the evaluation process because their feedback is valuable.

References

- Abrami, P. C., Leventhal, L. and Perry, R. P. (1982). Educational seduction. *Review of Educational Research*, 52(3): 446-464.
- Bachen, C. M., McLoughlin, M. M. and Garcia, S. S. (1999). Assessing the role of gender in college students' evaluations of faculty. *Communication Education*, 448(3): 193-210.
- Basow, S. A. (1995). Student evaluations of college professors: When gender matters. *Journal of Educational Psychology*, 87(4): 656-665.
- Brockx, B., Spooren, P. and Mortelmans, D. (2011). Taking the grading leniency story to the edge: The influence of student, teacher, and course characteristics on student evaluations of teaching in higher education. *Educational Assessment, Evaluation and Accountability*, 23(4): 289-306.
- Burns, R. B. and Burns, R. A. (2008). *Business research methods and statistics using SPSS*. London, SAGE.
- Centra, J.A. (2003). Will teachers receive higher student evaluations by giving them higher grades and less course work? *Research in Higher Education*, 44(5): 495-518.
- Changing Minds. (2012). Validity. Retrieved July 24, 2012, from http://changingminds.org/explanations/research/design/types_validity.html

- Darby, J.A. (2006). The effects of the elective or required status of courses on student evaluations. *Journal of Vocational Education and Training*, 58(1): 19-29.
- Greenwald, A. G. and Gillmore, G. M. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology*, 89(4): 743-751.
- Hair, J., Anderson, R., Tatham, R. and Black, W. (1998). *Multivariate data analysis*. Upper Saddle River, NJ: Prentice Hall.
- Hamermesh, D. S. and Parker, A. M. (2005). Beauty in the classroom: Professors' pulchritude and putative pedagogical productivity. *Economics of Education Review*, 24(4): 369-376.
- Heine, P. and Maddox, S. (2009). Student perceptions of the faculty evaluation process: An exploratory study of gender and class differences. *Research in Higher Education Journal*, 3: 1-10.
- ICE Report. 2005. *Faculty and Student Perspectives on Student Teaching Evaluations*. Retrieved from the American University of Beirut.
- Isely, P. and Singh, H. 2005. Do higher grades lead to favorable student evaluations? *Journal of Economic Education*, 36(1): 29-42.
- Jamieson, S. 2004. Likert scales: how to (ab)use them. *Medical Education*, 38: 1212-1218.
- Jones, J. 1989. Students' ratings of teacher personality and teaching competence. *Higher Education*, 18, 551-558.
- Kidd, R. S. and Latif, D. A. 2004. Student evaluations: Are they valid measures of course effectiveness? *American Journal of Pharmaceutical Education*, 68(3): 1-5.
- 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.
- Marsh, H. W. 1987. Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11(3): 253-388.
- Marsh, H. W. and Bailey, M. 1993. Multidimensional students' evaluation of teaching effectiveness. *Journal of Higher Education*, 64(1): 1-18.
- Murray, H. G., Rushton, P. J. and Paunonen, S. 1990. Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology*, 42(2): 250-261.
- Nunally, J. 1978. *Psychometric* (2nd edition). New York: McGraw Hill.
- Pell, G. 2005. Use and misuse of Likert scales. *Medical Education*, 39: 970.
- Surratt, C. K. and Desselle, S. P. 2007. Pharmacy students' perceptions of a teaching evaluation process. *American Journal of Pharmaceutical Education*, 71(1): 1-7.
- Thornton, B., Adams, M. and Sepehri, M. 2010. The impact of students' expectations of grades and perceptions of course difficulty, workload, and pace on faculty evaluations. *Contemporary Issues in Education Research*, 3(12): 1-5.
- Weinberg, B. A., Hashimoto, M. and Fleisher, B. 2009. Evaluating teaching in higher education. *Journal of Economic Education*, 40(3): 227-261.
- Young, S., Rush, L. and Shaw, D. 2009. Evaluating Gender Bias in Ratings of University Instructors' Teaching Effectiveness. *International Journal for the Scholarship of Teaching and Learning*, 3(2): 1-14.