Students' Evaluation of the Teaching Process at Lebanese Universities: An Assessment of the Faculty's Perceptions

Ale J. Hejase [1*] • Hussin J. Hejase [1] • Anwar M. Tarhini [2] • Joumana Younis [2]

[1] Faculty of Business Administration, Al Maaref University, Beirut, Lebanon
[2] Faculty of Economics and Business Administration, Islamic University of Lebanon, Beirut, Lebanon
hhejase@mu.edu.lb

Abstract: Lebanon, like other countries, has a special interest in conserving the integrity, quality, and fairness of its higher education, especially that it serves approximately 200,000 students registered in the scholastic year 2017-2018; these students injected millions of dollars into the private institutions. Students are seen as customers whose satisfaction and comfort jump to the front and prevail as the main doctrine that governs the educational process. The student-as-customer concept is a fundamental factor that plays a major role in educational institutions, even the traditional non-profit ones that basically depend on tuition fees incurred by the number of enrolled students in order to cover their inflated budgets; hence, the need to have satisfied stakeholder —students. This orientation has made the Student Evaluation of Teaching (SET) a key criterion to evaluate the teaching process; especially since SET is currently used to assess and improve teaching and learning processes. Faculty members who are considered by the students, non-lenient, serious, strict, and highly demanding are put under pressure, if not penalized, irrespective if it is fair or not, hence leading to the former being frustrated. This research provides an insight into the frustration state of the faculty members at certain Lebanese universities due to the importance which is given to SET and its effects on the educational procedure and their careers per se.

This research is exploratory and quantitative using a survey questionnaire distributed to 146 faculty members. Data is analyzed using SPSS software. Findings reveal that the majority of the said faculty members have a negative view of the SET evaluation process in terms of format, and the content of the corresponding questionnaires; results are seen as being ineffective and inappropriate means of evaluation. The SET evaluation process is complicated and the assessment survey differs from one university to another. Outcomes of this paper serve policymakers at the higher education institutions when setting the institution's mission and vision and purpose; it may direct the continuous efforts to create agreed upon teaching and learning quality standards.

1. Introduction:

Researchers around the globe have mixed opinions about the implementation of the Student Evaluation of Teaching (SET) as its being considered as a basic criterion to the appraisal of the teaching process. A review of researches indicates that the topic of student ratings of university teaching has been very popular in the past century. Brandenburg and Remmers in 1927 wrote about the Purdue rating scale for faculty members. Heilman and Armentrout (1936), in an early research done at Colorado State College of Education in 1935, reported that students do differ widely in rating a single faculty member; moreover, the faculty was rated very differently by their students on each of the evaluation traits. Since then, many investigators have studied faculty/student rating mechanisms and have presented significant results that do

relate reliability, validity, and effectiveness of SETs to different characteristics pertaining to the faculty themselves and to their respective students. In 1971, Costin, Greenough, and Menges presented an extensive review of related literature. In 1994, Stratton, Myers, and King concluded that SETs' results became the principal measure for salary and promotion decisions and thus the SETs mandatory adoption resulted in an increase in grades to students of about 11%. During the years after, the debate of student and faculty perceptions of SETs continued with the same thrust as before. Sojka, Gupta, and Deeter-Schmelz (2002) examined the differences in students' and faculty members' perceptions of SETs and concluded that faculty and student do have different perceptions of SETs; however, they found that neither party wanted to remove

this scale from the academic environment. The subject continued to attract the attention of numerous researchers from all around the world. In Ireland, Surgenor (2013) states that SET process is not to be only implemented to collect data but proposes the inclusion of evidence and sources of information to avoid non-representative data that affect the decisions taken by the academic sector, such as promotions. In Australia, after tens of years of debates, Alaudinn and Kifle (2014) concluded that "*SET instrument currently used fails to provide a valid measure of teaching quality as it does little to measure the extent of students' actual learning*" (p. 156). Likewise, in the far east, specifically in Hong Kong, where Chan, Luk and Zeng (2014) concentrated on the fairness and validity of SETs and concluded that faculty members have diverse attitudes towards them, where only 39% of the faculty members believe that SETs reflect their teaching and quality abilities, and that faculty members believe that they can attain high SET scores upon spoon-feeding their students and lowering the standards of their courses. Finally, in Lebanon, there is modest literature as to how Lebanese faculty members perceive the SET; according to the researchers of this study, there exist four publications, including one thesis, one conference proceedings, and two research papers; the first publication is a single descriptive, narrow, private and unpublished study performed at the American University of Beirut (2005). The two papers are written by Hejase et al. (2013) and (2014); in the first one, the authors reported the students' perception of the SETs, while the second concluded that SET evaluations are tied with the students' characteristics such as their gender, GPA, age, number of credits completed and institution. As for the conference proceedings, Hejase et al. (2015) provided insight into the faculty's perceptions of SETs.

Therefore, with SETs being one of the prime measures used to assess and improve teaching and learning, it will continue to incur controversy and criticism as its use. That's why this research aims at contributing to the current existing SET literature by exploring the attitudes of faculty members at Lebanese universities. This paper is divided into six sections. The first is the introduction; the second is a literature review of the topic, followed by chapters on methods and tools. The fourth is about results and findings; the fifth includes a discussion section and ends with conclusions and recommendations.

2. Literature Review:

Heilman and Armentrout (1936) contend that the purpose of their study, administered to 2115 students, is to provide the students with some means to judge their faculty members on ten traits, namely, interest in subject, sympathetic attitude towards students, fairness in grading, liberal attitude, presentation of subject-matter, sense of humor, self-confidence, personal peculiarities, appearance, and the degree of stimulating intellectual curiosity. The results obtained show that students vary widely in rating

any single teacher as seen in the standard deviations which are significantly large. Moreover, this 1936 study highlighted some of the known debates, for example:
 a) There is no decline in teaching skills as faculty members accumulate years of experience.
 b) There is no correlation between the severity of grading and the ratings.
 c) The student's gender differences and their influence on ratings are negligible.

More research was carried out whose details were reported by Costin et al. (1971) offering an extensive literature review and presentation of most of the works done prior to 1971 on the subject of student ratings of teaching. This work which was published in twenty-five pages was later summarized by the same authors in only three pages and did appear in 1973 (Costin, Greenough and Menges, 1973). The findings of these aforementioned papers reflect the fact that students' ratings can provide reliable and valid information on the quality of courses and instruction, and that students' ratings are tied more to the quality of subject-matter than to class entertainment. Moreover, the study attributed the relatively high correlation between ratings and grades to greater interest in the course rather than seeing it as a "reward effect". Noticeably, in all the aforementioned pioneer studies, little was mentioned about faculty's opinion on the student ratings.

In 1973, Villard continued to question the validity of SETs; he, specifically, asserts that "*The more formal the use made of SET scores, the more rapidly will the average level at which courses are taught drift downward*" (p. 50). Additionally, he raised questions on matters such as the manipulation of the SETs by giving leniently graded tests prior to dates of administering the SET; he concluded that "*The more formal the use made of SET scores, the more rapidly will average grades drift upwards*" (p. 50). Two years later, Kipps (1975) reexamined the findings of previous SET researchers and came to the conclusion that student course evaluations may contribute positively to the increase in the faculty learning output or simply the students' accomplishments. Again, nothing was said about what faculty themselves think about SETs.

Faculty behavior in response to SETs was remarkably treated in the work of Stratton, Myers, and Kung (1994). This study cited many distinguished SET works that were done during the seventies and eighties, and highlighted questions and issues such as: Can faculty members 'buy' or incur higher SET scores by lowering the efforts required from the student body? Stratton et al. (1994) indicated that the grades related to economics courses suffered 11% inflation upon introducing the mandatory SETs at the University of Akron. This fact was tested by Eiszler (2002) who studied 983,491 student evaluations between the spring of 1980 and the fall of 1999; he concluded that across the 40 semesters, the semesters in which faculty received higher SET ratings were the

semesters in which they awarded higher percentages of A and A⁻ grades. Later, McPherson, Jewell, and Kim (2009) also found that "faculty members can 'buy' better evaluation scores by inflating students' grade expectations" (p. 37). Once more, nothing was raised about how faculty members perceive SETs.

By the end of the second millennium, the amount of research devoted to faculty perceptions of SETs continued to be comparatively little. On this track, Schmelkin et al. (1997) concluded that contrary to the anecdotal literature cited in their paper, their results do not portray the fact that a great deal of resistance to SETs exists on part of the faculty. Indeed, this research showed that in the eyes of faculty, SETs are useful for both academic formative and summative purposes.

The fact that SETs are being diverged from being a push for improvement of teaching effectiveness to a means for determining faculty merit, promotion, and tenure, was considered and dealt with in the work of Simpson and Siguaw (2000). In their research, to answer the question about the heavy reliance on SETs by many universities, they emailed a questionnaire to 543 members of the Academy of Marketing Science; only 53 responded (9.6% response rate), thus making their research exploratory in nature. Their general results indicate that SETs are too affluently used as a sole means for evaluating teaching performance. Likewise, they express serious concerns about the abilities and knowledge-base of students to evaluate their faculty members. Similarly, their results show that 48.1% of the respondents believe that SETs are somewhat inaccurate (42.3% believe that they are accurate); in general, respondents perceive SETs as being a problematic assessment that encourages the lowering of the academic standards.

The similarities and differences between faculty and student perceptions of SETs is treated by Sojka et al. (2002), where the disturbing finding indicates that on one side students do not believe that SETs cause faculty to be more lenient in grading while on the other side faculty members think that students give higher scores to easy going and entertaining faculty members. On the other hand, McPherson's (2003) research, that included 987 economics classes over 18 consecutive semesters at the University of North Texas, concludes that "a principal finding is that there is no strong evidence that SET scores are 'contaminated' by faculty members attempting to 'buy' better SET scores by raising grade expectations" (p. 15). To overcome the aforementioned dilemma, Engelland (2004) presents guidelines and advice to faculty in order to manage the SET process and achieve effective teaching. The study reveals that teaching effectiveness goes beyond SET scores and requires on-going assessment that is not tied to the end of the semester where faculty may manipulate the scores by offering the students inducements, grading leniency, pre-evaluation favorable announcements, and manipulated class expectations or

student achievements. Engelland's (2004) suggestions for improving teaching effectiveness are depicted in Exhibit 1.

Exhibit 1: *Suggestion for Improving Teaching Effectiveness*

1. A well-rounded assessment process that really determines if students have learned the intended subject material;
2. A student self-assessment mechanism to measure self-satisfaction with own performance; and,
3. Adoption of SETs to determine student's satisfaction with the quality of teaching.

Source: Engelland (2004).

In a more recent study, Balam and Shannon (2010) assessed and compared the opinions of both faculty and students, 34 faculty members, and their 968 students. Findings revealed that faculty members believe that SET results are invalid and unreliable, SETs are not meant to improve instruction, and SETs are a popularity contest where winners are those that are more friendly and humorous; while students supported the classical myths such as they were qualified to judge their faculty members, and grades received are highly correlated with ratings of faculty members. A similar comparative study investigated the attitudes of 71 lecturers and 137 students (Rosemarin, 2010) and three principal results were reached. The first one states that both groups agree that the knowledge of the lecturer is the most important criterion to consider in the evaluation. The second result indicates that the use of audio-visual class techniques is to be given the lowest importance in the evaluation process. The third result states that both groups marked a high correlation between student understanding and their ratings, in addition to conflicting learning goals whereby students want to be prepared for the future while faculty members want their students to possess enhanced thinking.

Furthermore, the issue of how faculty members respond to SETs was studied taking into consideration the faculty members' genders, positions and rank differences (Kogan, Schoenfeld-Tacher, & Hellyer, 2010). The outcomes of the study reveal that minimal differences exist in the faculty responses based on rank and position while gender played a dominant role with females being more negatively impacted by the SET results as compared to their male counterparts.

Finally, a nice mechanism of faculty assessment was proposed as an alternative to the sole use of SET scores (Turpen, Henderson, & Dancy, 2011). Turpen et al. conclusions are delineated in Exhibit 2.

Exhibit 2: *Turpen et al. (2011): Principales Conclusions*

Four principal conclusions are presented:
1. Faculty members are more positive towards their own methods of teaching assessment;

2. Institutions continue to base their assessments of the teaching process on SETs which are in the point of view of faculty members questionable;
3. Faculty have their own methods to evaluate their teaching process; and,
4. Neither faculty members nor institutions make much use of the quantity of the current research and findings that deal with the assessment domain.

In what relates to Lebanon, a limited study on the perceptions of SETs was executed at The American University of Beirut (2005) where more than half of the 145 faculty members who participated in the study believed that faculty members change their teaching style to receive higher ratings, and 40% asserted that ratings do determine what is addressed in class. Furthermore, the study revealed that the majority of faculty believed that high demands of students reflect negatively on their scores, around 50% of the respondents believed that good faculty members get high SET evaluations.

2.1. Research Problem:

Lebanon, among other countries, is a place where institutions of higher education have mushroomed to such an extent that, during the past twenty years, the number of higher education institutions increased from around fourteen to more than forty-nine, including one public university (Higher Education, 2018); among these are the newly emerging institutions owned by families or individuals seeking financial profits. The aforementioned newly established higher education institutions capitalize on the Ministry's policy which states, as Hejase and Alaeddine (2017) contend, the "pillars upon which the Ministry of Education and Higher Education (MEHE) has founded its philosophy are: education based on equal opportunity, quality education to build a knowledge society, education to contribute to social integration and economic development" (p. 27).

The aforementioned MEHE philosophy has helped "prepare higher numbers of high school graduates to join the ranks of higher education institutions, keeping the demand for higher education in Lebanon robust, and leading to the establishment of more universities" (ibid). In general, high education institutions are profit organizations that served in 2009-2010, 180,850 students (Yaacoub & Badre, 2012), and in 2017-2018, 200,100 students (Dr. Ghada Chehimi, personal communication April 9, 2018), resulting in the injection of millions of dollars into the pockets of these institutions' owners. Students are seen as customers whose satisfaction and comfort jump to the front and prevail as the main doctrines that govern the educational process. The student-as-customer concept has expanded to cover even the traditional non-profit institutions that do rely on tuition fees and consequently on students' satisfaction in order to cover their inflated budgets. This customer-oriented educational process has given SETs a deeper importance, and has led to penalizing non-lenient, serious, strict, and highly demanding faculty, and putting them under the institution pressure driven by the customers' opinions.

The customer declaration has been studied extensively in the literature (Muncy, 2008; Vuori, 2013; Mark, 2013a) but a new view of customers has emerged where the old slogan "customers are always right" is considered outdated and is being updated by matters related to getting the customers engaged in a partnership with their suppliers and thus making them partly accountable of their own satisfaction. In fact, unless customers (students) perform their "co-production" roles effectively, the expected outcomes will not be realized (Kotze' & du Plessis, 2003). Actually, students cannot be viewed anymore as passive recipients but should be part of the teaching environment, i.e. co-producers of the services they receive (Mark, 2013b). Furthermore, dealing with students as customers empowers them to the extent that they can blame their universities and their faculty members for their personal failures. Truthfully, the highest objective of the higher education institution should be the career success of the students (Winer, 1999).

Currently, in Lebanon as in many other places, universities excel at satisfying their students. Nonetheless, the meaning of satisfaction and its corresponding measurement need further attention. Alves and Raposo (2009) cite the dimensions proposed by different researchers about the satisfaction of students in higher education; they initially conclude that in most of the reviewed studies, the satisfaction of students is measured by only one variable "*quality of the education service*" (p. 207). However, they conclude their work by showing that with an internal reliability of 0.93, "the construct satisfaction can be measured by three indicators being: level of global satisfaction, level of correspondence to expectations, and the level of correspondence to the student's current necessities/wishes" (p. 216). Unfortunately, SETs are considered the main tool to measure student satisfaction, and the literature is inundated with studies that have highlighted the extensive use of SETs as feedback measures related to the learning process (Engelland, 2004; Zabaleta, 2007; Turpen et al., 2011; Lidice & Saglam, 2013; Alauddin & Kifle, 2014). Visibly, most faculty members have rejected this aforementioned claim; in fact, a recent news briefing reported that faculty members at a reputable American university have rejected bonuses based on the results of student evaluations (Wiely Periodicals, 2009).

Thus, this paper aims to provide a concise answer to the aforementioned frustrations by surveying faculty members as to their perceptions of SETs that according to many are becoming the excuse and because which endangers their academic careers. Considering that students in Lebanese universities are customers and accepting that these universities excel at satisfying their

students; and, having SETs as the principal means to measure customer satisfaction, the current study proposes the following hypotheses:

1. Faculty members have a negative attitude towards SETs.
2. Faculty members believe that SETs are not useful.
3. SETs are not the main tool to measure teaching effectiveness.
4. Students are not able to make accurate SET judgments.
5. SETs are a double-edged sword used by the administration as per their convenience.

Examining the aforementioned hypotheses from the point of view of the faculty members, we will be able to determine a collection of beliefs that Lebanese faculty members carry against or in favor of the SET mechanism.

3. Methodology:
The researchers are independent and assume the role of objective analysts; a fact that justifies the research philosophy adopted or positivism. However, the research approach is deductive, whereby a concept is assessed based on a structured survey questionnaire developed as a tool to collect and analyze primary as well as secondary data. Also, the research is exploratory and comparative in nature; the researchers have sought descriptive analysis first; then, tried to create relationships among variables in order to explain the concept under study.

3.1. Sample
The data for the study was collected during the fall semester of 2014 form a convenient sample distributed across nine Lebanese universities. Questionnaires were sent to 240 faculty members; 146 responses were received (response rate 60.83%).

3.2. Survey Questionnaire Design
The survey questionnaire consists of four sections. This questionnaire was mainly developed based on former questionnaires and surveys used in other studies related to faculty members' perceptions of students' evaluations of the teaching process (Schmelkin, Spencer, & Gellman, 1997; Simpson & Siguaw, 2000; Nasser & Fresko, 2002; Sojka, Gupta, & Deeter-Schmelz, 2002; Morgan, Sneed, & Swinney, 2003; The American University of Beirut, 2005; Zabaleta, 2007; Balam & Shannon, 2010; Kogan, Schoenfeld-Tacher, & Hellyer, 2010; Chan, Luk, & Zeng, 2014; Hejase, Hejase, & Al Kaakour, 2014).

The questionnaire starts with a paragraph that explains its purpose and is divided into four parts. The first three parts are composed of close-ended questions, using a 5-point Likert scale, extending from "Strongly Disagree" to "Strongly Agree". Part I deals with the faculty's attitude to current students' evaluations of the teaching of the

course, its 36 items cover most of the aspects and ideas that relate to the faculty's attitudes to the SETs and similar evaluation tools. Part II of the questionnaire covers nine items whose main purpose is to get feedback on issues, thoughts, and feelings related to the usefulness of the students' course evaluations. Part III includes three close-ended items plus another close-ended item with three categorical answers, in addition to two open-ended questions: "*Provide examples of the changes you have made to improve your SET scores*" and "Personal Opinions" about SETs. Finally, Part IV deals with demographic and professional information related to the respondent's institution/school, academic rank (if applicable), degree earned, marital status, religion, rank, position, age and years of experience.

3.3 Data Analysis
All responses were processed using SPSS software —Statistical Product and Service Solutions, an IBM product acquired in 2009 (Hejase and Hejase, 2013, p. 58). Descriptive statistics were performed including frequency and percentage distributions data tables. Moreover, Chi-square and Factor analysis were performed to study relationships between variables that may add value to the findings of the research.

4. Results:
4.1. Participants
Respondents are faculty members of nine Lebanese universities: 26.7% at The Lebanese American University (LAU); 17.8% at the Antonine University; 17.1% at the Lebanese International University LIU; 11.6% at the American University of Science and Technology (AUST); an equal percentage of 8.9% at the American University of Beirut (AUB) and Islamic University in Lebanon; and, the remaining including three universities; 8.9% at Kaslik University, University of Saint Joseph USJ, and Sagesse University. Likewise, 58.9% of the respondents belong to the Business Faculty; 17.1%, 11%, and 8.2% belong respectively to the Engineering & Architecture, Liberal Arts, and Arts and Sciences Faculties.

The dominant academic rank is Assistant Professor (30.8%) followed by faculty members (25.3%), Associate Professor (12.3%), Lecturer (18.5%), Full Professor (1.4%), and the remaining percentage is the staff. Furthermore, 50% of the respondents hold a Doctoral Degree, 38.4% a Master Degree, and 11.6% had no answer. 57.5% of the respondents are married, and males marginally dominated the gender characteristic with 53.4%. Moreover, the percentage of full-timers is 50.7% against part-timers who amounted to 43.2%. As for age and years of experience, the sample indicated an average age of 38.24 years with a standard deviation of 8.553 years (median=36 years) and average years of experience of 10.43 years with a standard deviation of 7.133 years (median=9 years). The non-traditional "*Religious*

Preference" question that for many looked unethical (27.4% did not answer and 2.7% indicated that it is not a good question) was purposely introduced to the questionnaire demographics to assess the Lebanese faculty's perception of such a delicate issue that is very sensitive and irritable within the Lebanese society.

In effect, 28.8% of the respondents declared their Christianity, 28.1% their Islam, and 27.4% abstained from answering the question, and 10.3% asserted that they do not have any religious preference. In fact, the authors are not aware of any published study that tackles the question of religious bias as a factor that affects students' evaluations; however, some researchers did examine the issue of racial bias as a factor in students' evaluations. For example, the work of Smith (2007, p. 6), where the findings indicate that African American faculty members received lower ratings than that of the Caucasian faculty. In fact, the lower students' ratings of the African American faculty are troublesome since these ratings affect these faculty's members merit increases and careers.

4.2. Attitudes

The percentages of responses made by all faculty members as to their attitudes to SETs are reported in Tables 1 and 2. For simplicity to judge the overall degree of agreement or not [SD and D] were grouped to mean disagreement as well as [A and SA] were grouped to mean agreement. The negative perceptions of SETs are shown in Table 1.

Table 1. *Negative perceptions of SETs*

(1) 54.7% of respondents believe that students do not treat course evaluations seriously (15.8% neutral),
(2) 44.1% thinks that students do not have enough knowledge to judge the quality of instruction (13.1% neutral),
(4) 41.4% judges that faculty members who demand a lot from their students get low evaluations (25.5% neutral),
(6) 57.2% considers that SETs measure how nice an instructor is (26.2% neutral),
(8) 57.9% believes that one does not have to be a good instructor to get positive evaluations (11.7% neutral),
(9) 60.3% ascertains that students' ratings are not an acceptable criterion for granting salary increases (21.2% neutral).
(10) 45.2% concludes that SETs are an unacceptable criterion for granting tenure (28.5% neutral),
(11) 50% agrees that students' ratings are not an acceptable criterion for granting promotion (28.5% neutral),
(12) 46.6% declares that students write comments only when they feel negative about the instructor (15.1% neutral),

(15) 45.9% emphasizes that students give better ratings to faculty members that teach less demanding courses (18.5% neutral),
(16) 39.6% believes that students base their course ratings on how entertaining a professor is (29.9% neutral),
(17) 48.9% agrees that students evaluate professors based on character rather than on teaching ability (23.4% neutral),
(21) 39.9% indicates that students' ratings can seriously jeopardize the career of a professor (31.5% neutral),
(27) 40.4% points out that students tend to agree among themselves as to the evaluations of their faculty members (29.5% neutral), and
(28) 76% states that unhappy students show their feelings about a class by writing bad evaluations of the instructor (15.1% neutral).

On the other hand, there was some positive attitude towards the SETs. Table 2 depicts these responses.

Table 2. *Positive attitude towards the SETs.*

(26) 58.5% of the respondents disagree on that a tenured professor with low SET scores receives insulting letters and minimal pay increments (35.9% neutral),
(3) 56.5% of the respondents believe that good faculty members get high evaluations (24.1% neutral),
(7) 44.1% agrees that students' ratings correctly reflect the quality of instruction (28% neutral),
(13) 68.3% does not agree that administering teacher and course evaluations every semester is a waste of time (20.7% neutral),
(14) 49.3% opposes the idea that faculty member tends to reduce their course requirements in order to get favorable ratings (26.4 neutral),
(18) 53.4% believes that students evaluate faculty member based on how much they have learned in the course (24% neutral),
(19) 48.9% specifies that students evaluate faculty member based on the latter teaching ability (23.4% neutral),
(23) 50.7% does not accept that students' ratings reduce faculty's morale and job satisfaction (28.5% neutral),
(29) 49.3% refuses the idea that SET systems do not encourage good teaching (30.1% neutral),
(31) 51.4% of the respondents refuse the idea that SETs, as an evaluation tool, kills high-quality teaching (27.1% neutral),
(32) 49.3% of the respondents reject that the gender of a student affects the SET scores (34.2% neutral),
(34) 39.7% refuse that the time of the day a course is offered affects SETs results (27.4% neutral),

(35) 48.6% accept that the level of a course, being sophomore or higher, affects SET scores (17.1% neutral),
(36) 44.8% refuse that the rank of the professor influences the SETs scores (29.5% neutral).

4.2.1. Attitudes towards the Usefulness of SETs

The percentages of responses made by all faculty members on their attitudes towards the usefulness of SETs are reported next. The results show that 61.4% of the respondents believe that SET ratings are a good basis for improving teaching (13.1% neutral), and 40.8% believes that faculty members make significant changes to their courses' contents based on the SET scores (27.5% neutral). Moreover, 42.9% does not accept that SET ratings undermine student-faculty member relations (28.9% neutral), while 61.2% of the respondents agree that SET ratings help faculty members improve their treatment of students (21.8% neutral). Also, 68.7% agrees that SET scores help faculty members focus on the weak points of their teaching methodology (13.2% neutral). In addition, the results reveal that 45.5% of the respondents agree that if SETs were given at an earlier point in the semester, they would use the students' feedback to improve immediately (28.7% neutral). Also, 37.6% refuses the idea that, in general, the current evaluations do not provide any useful information related to teaching (36.2% neutral). Moreover, 56% believes that students don't take the evaluation process seriously enough to provide meaningful feedback to faculty's performance (19.6% neutral), and 62.5% believes that SETs should be used only to provide formative feedback for the faculty members (16.7% neutral).

4.3. Research Questions Analysis

The percentages of responses depicted in Table 3 can help build answers to three of the aforementioned research questions which are:
1. SETs are not the main tool to measure teaching effectiveness;
2. Students are not able to make accurate SET judgments; and
3. SETs are a double-edged sword used by the administration as per their convenience.

The results show that 44.8% of the respondents refuse their peers' evaluation (14.7% neutral), 44.1% refuses that students cannot make accurate judgments (16.1% neutral), and 42.8% agrees that low SETs are a double-edged sword that is used by administration as per their convenience, ignoring the scores as supportive comments and observing these results as detractors (35% neutral).

Table 3. *Percentage results for the Hypotheses 3, 4 and 5*

Attitude	D	N	A
1. Professors' colleagues with excellent publication records and expertise are better qualified to evaluate their peers' teaching effectiveness.	44.8	14.7	40.5
2. Students are not able to make accurate judgments until they have been away from the course and possibly away from the university for several years.	44.1	16.1	39.8
3. Low student evaluations are a double-edged sword that is used by the administration as per their convenience, ignoring them as supportive comments and observing these results as detractors.	18.6	35	42.8

4.4. Descriptive Statistics

The questions presented in the research instrument follow a 5-point Likert scale with the following coding: 1 for Strongly Disagree, 2 for Disagree, 3 for Neutral, 4 for Agree, and 5 for Strongly Agree. On considering the 36 items related to faculty members' attitudes towards SETs, we notice that Tables 4, 5 and 6 present the mean for each item. In addition, the means are compared to the test value of 3 (the Neutral case) in order to test if the means are significantly larger than 3; if they are, then this means that the agreement side of the response is emphasized. All significant tests (5% level of significance) are depicted in Tables 4, 5 and 6.

Table 4. *Strong agreement & statistically significant faculty members' attitudes towards SETs*

Statement	Mean ≥ 3 & Stat. Signif. ≤ 5%
Unhappy students show their feelings about a class by negatively evaluating for the faculty member;	Mean = 3.89
Students evaluate the faculty member based on their teaching ability	Mean = 3.53
The size of the class affects the students' ratings	Mean = 3.53
Good faculty members faculty member gets high course evaluations	Mean = 3.46
Students' ratings measure how nice a faculty member is?	Mean = 3.46
One does not have to be a good faculty member in order to get positive evaluations	Mean = 3.35
Students evaluate faculty members based on how much they have learned in the course	Mean = 3.34
Students evaluate faculty member based on the latter's character rather than on their teaching ability	Mean = 3.24
The level of the course (freshman, sophomore, junior, senior, or graduate) affects students' ratings	Mean = 3.19
Students' ratings correctly reflect the quality of instruction	Mean = 3.17

Table 4 shows strong agreement and at the same time statistically significant faculty attitudes towards SETs.

Table 5. Marginal Agreement of faculty members' attitudes towards SETs

Statement	Mean ~ 3 & Stat. Signif. ≥ 5%
Faculty members who demand a lot from their students get low evaluations	Mean = 3.14, p=.110
Students' ratings can seriously jeopardize the career of a faculty member	Mean = 3.14, p=.105
Students tend to agree among themselves regarding the evaluations of their faculty member	Mean = 3.14, p=.100
Students give better ratings to faculty members that teach less demanding courses	Mean = 3.12, p=.176
Students' ratings encourage faculty members to be lenient in grading	Mean = 3.11, p=. 207
Students base their course ratings on how entertaining a faculty member teacher is	Mean = 3.08, p=.349
Students do not have enough knowledge to judge the quality of instruction	Mean = 3.06, p=.477
Whether students are majoring or non-majoring in a course affect their ratings of the teacher faculty member	Mean = 3.05, p=.563
Required courses generally get lower evaluations than elective courses	Mean = 2.99, p=.857
The time of the day the course is offered affects students' ratings	Mean = 2.92, p=.326
Students write comments only when they feel very positively about the instructor	Mean = 2.86, p=.286

Table 5 shows marginally but statistically not significant faculty's attitudes towards SETs.

Table 6. Rejected & statistically significant faculty members' attitudes towards SETs

Statement	Mean ≤ 3 & Stat. Signif. ≤ 5%
Students' ratings pose a threat to academic freedom	Mean = 2.79
Most students treat course evaluations seriously	Mean = 2.73
Students' ratings are an acceptable criterion for granting tenure	Mean = 2.68
Faculty members tend to reduce their course requirements in order to get favorable ratings	Mean = 2.67
The academic policy of using SETs as an evaluation tool is killing high-quality teaching	Mean = 2.65
Students' ratings reduce faculty's morale and job satisfaction	Mean = 2.62
SET systems do not encourage good teaching	Mean = 2.62
An untenured professor who does not score well on SETs has to start looking for a new job	Mean = 2.61
Students' ratings are an acceptable criterion for granting promotion	Mean = 2.60
It is research performance that identifies a good faculty member teacher not the teaching method	Mean = 2.55
The rank of the faculty member teacher (instructor, assistant professor, associate professor, professor) affects students' ratings	Mean =2.53
The gender of the student affects the SET's scores	Mean = 2.51
Students' ratings are an acceptable criterion for granting salary increases	Mean = 2.45
A tenured professor with low SET scores receives insulting letters & minimal pay increments	Mean = 2.32
Administering teacher & course evaluation every semester is a waste of time	Mean = 2.23

Table 7. *Descriptive statistics of questions related to the usefulness of SETs*

Descriptive Statistics (test value = 3)	Mean	Std. Dev.	N	p-value (2-tails)
Students' ratings are a good basis for improving teaching.	3.39	1.016	145	.000
Faculty make significant changes to course content based on students' evaluations.	3.10	.955	142	.221
Students' ratings undermine student-faculty member relations.	2.82	1.001	142	.031
Students' ratings help faculty members improve their treatment of their students.	3.46	.912	142	.000
In general, students' ratings help faculty members to focus on the weak points of the teaching methodology of a course.	3.55	.930	144	.000
If SETs were given at an earlier point in the semester, I [faculty member] would use the students' feedback immediately	3.26	1.005	143	.002
In general, the current evaluations do not provide any useful information.	2.89	.954	141	.188
Students don't take evaluations seriously enough to provide meaningful feedback to faculty.	3.45	1.086	143	.000
SETs should be used only to provide formative feedback for the faculty member.	3.52	.961	144	.000

Furthermore, considering the nine items related to the usefulness of SETs, Table 7 presents the mean and standard deviation for each item. In addition, the means are compared to the test value of 3 (the Neutral case) in order to test if the means they are significantly larger than 3. Table 7 shows that there are two statements; one is marginal but rejected and the other is lower than the mean and also rejected. All other statements are accepted by the sample and are statistically significant.

As for the research hypotheses 3, 4 and 5 presented earlier, testing is based on H_0: $\mu \geq 3$; and H_a: $\mu < 3$, in that if the p-value $> \alpha = 5\%$, the null hypothesis is accepted or otherwise rejected an alternative hypothesis is chosen. Table 8 shows the following results: For hypothesis 3, "SETs are not the main tool to measure teaching effectiveness", the question presented is: "*Faculty member' colleagues with excellent publication records and expertise are better qualified to evaluate their peers' teaching effectiveness*"; the mean response is 2.89, which when tested, a p-value = .245 $> \alpha = 5\%$ is gotten, and the null hypothesis, meaning that colleagues with excellent publication records and expertise are better qualified (at 5% level of significance) to evaluate their peers' teaching effectiveness is accepted.

For hypothesis 4, "Students are not able to make accurate SET judgments", the question presented is: "*Students are not able to make accurate judgments until they have been away from the course and possibly away from the university for several years*"; the mean response is 2.98, which when tested, a p-value =.823 $> \alpha = 5\%$ is attained and the null hypothesis, meaning that students are not able to make accurate SET judgments, cannot be rejected at 5% level of significance.

Furthermore, for hypothesis 5, "SETs are a double-edged sword used by administration as per their convenience", the question presented is: "*Low students' evaluations are a double-edged sword that is used by administration as per their convenience, ignoring them as supportive comments and observing them as detractors*"; the mean response is 3.29 with p-value =.001 $< \alpha = 5\%$ and the null hypothesis is rejected, meaning that we do not accept the fact that low students' evaluations are a double-edged sword that is used by administration as per their convenience, ignoring them as supportive comments and observing them as detractors.

4.5 Factor Analysis

A principal factor analysis (Principal Components) with subsequent rotation (Varimax with Kaiser Normalization) was conducted on the 9 items of the questionnaire pertaining to the "Usefulness of course evaluation". Many correlations were in excess of 0.3, and both the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO=0.7) and Bartlett's Test of Sphericity (Approx. chi-square 227.196, p-value=0.000) produced criteria that supported the application of PCA.

Table 8. Descriptive statistics of questions related to hypotheses 3, 4 and 5.

Descriptive Statistics (test value = 3)	Mean	Std. Dev.	N	p-value (2-tails)
Faculty member' colleagues with excellent publication records and expertise	2.89	1.15	143	.245
Students are not able to make accurate judgments until they have been away from	2.98	1.12	143	.823
Low Student evaluations are a double-edged sword that is used by the administration as per	3.29	1.02	140	.001

Communalities varied between 0.496 and 0.749, and three factors were deemed important for the analysis. Following rotation, factor 1 was loaded on four items that reflect the "Feedback Informative Dimension of SETs to Faculty Members" and accounted for 30.891% of the variance exemplified by the two highest loading items: "*In general the current evaluations do not provide any useful information*" and "*Students don't take evaluations seriously enough to provide meaningful feedback to faculty*". Factor 2 was loaded on 5 items and accounted for 15.540% of the variance. It was labeled" Help Dimension of SETs to Faculty Members" and was represented by the three highest loading items: "*Faculty make significant changes to course content based on students' evaluations*", "*Students' ratings are a good basis for improving the teaching*", and "*Students' ratings help faculty members improve their treatment of students*". The third factor accounted for 12.648% of the variance and was loaded on two items, suggesting it was measuring the "faculty members Hope and Expectation Dimension"; this dimension was represented by the highest loading item, namely, "*If SETs were given at an earlier point in the semester, I (faculty member) would have used the students' feedback immediately*". Tables 9 and 10 show the corresponding factor analysis' results.

Likewise, another principal factor analysis (Principal Components) with subsequent rotation (Varimax with Kaiser Normalization) was conducted on the 36 items of the questionnaire pertaining to the "Attitudes toward course evaluation". Many correlations were in excess of 0.3, and both the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO=0.693) and Bartlett's Test of Sphericity (Approx. Chi-Square 1816.160, p-value=0.000) produced criteria that supported the application of PCA. Communalities varied between 0.532 and 0.790, and ten factors were deemed important to the analysis.

Table 9. *Total Variance Explained for the 9 Usefulness of SETs factors*

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.780	30.891	30.891	2.780	30.891	30.891	2.003	22.254	22.254
2	1.399	15.540	46.431	1.399	15.540	46.431	2.002	22.246	44.500
3	1.138	12.648	59.078	1.138	12.648	59.078	1.312	14.578	59.078
4	.862	9.575	68.653						
5	.815	9.058	77.711						
6	.703	7.816	85.527						
7	.513	5.695	91.222						
8	.412	4.583	95.805						
9	.378	4.195	100.000						
Extraction Method: Principal Component Analysis.									

Table 10. *The loadings of the 9 usefulness of course evaluation items on three factors*

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.	Component		
	1	2	3
In general, the current evaluations do not provide any useful information.	.739		
Students don't take evaluations seriously enough to provide meaningful feedback to faculty.	.708		
SETs should be used only to provide formative feedback for the faculty member.	.615		
Students' ratings undermine student-faculty member relations.	.561	.489	
Faculty make significant changes to course content based on students' evaluations.		.694	
Students' ratings are a good basis for improving the teaching.		.681	
Students' ratings help faculty member improve their treatment of students.		.667	
If SETs were given at an earlier point in the semester, I (faculty member) would have used the students' feedback immediately.			.863
In general, students' ratings help professors to focus on the weak points of a course.		.487	.543

Following rotation, factor 1 accounted for 19.343% of the variance and was loaded on 8 items, suggesting it was measuring the "Instrumental dimension" and was represented by the highest 3 loading items: "*Students evaluate faculty member based on how much they have learned in the course", "Students evaluate faculty member based on their teaching ability", and "Good faculty member get high course evaluations*". Similarly, factor 2 was loaded on 3 items and accounted for 10.715% of the variance. It was labeled "Job performance" and was represented by the 3 high loading items: "*Students' ratings are an acceptable criterion for granting promotion", "Students' ratings are an acceptable criterion for granting salary increases", and "Students' ratings are an acceptable criterion for granting tenure*". Factor 3 was loaded on 7 items and accounted for 6.869% of the variance. It was labeled "Utility Dimension of SETs to Faculty Members" and was represented by the highest 3 loading items: "*It is research performance that identifies a good faculty member not the teaching method", "Academic policy of using SET as an evaluation is killing high-quality teaching", and "Students' ratings reduce faculty's morale and job satisfaction*". Moreover, factor 4 was loaded on 5 items and accounted for 5.461% of the variance.

It was labeled "Ego-defense" and was represented by the highest three loading items: "*The time of the day the course is offered affects students' ratings", "Level of course (freshman, sophomore, junior, senior, graduate) affects students' ratings", and "Whether students are majoring or non-majoring in a course affect their ratings*".

In addition, factor 5 was loaded on 5 items and accounted for 5.110% of the variance. It was labeled "Trust in SETs" and was represented by the highest three loading items: "*Students' ratings measure how nice a faculty member is", "Faculty members tend to reduce their requirements to get favorable ratings", and "Students base their course ratings on how entertaining a faculty member is*". Furthermore, the sixth factor was loaded on 2 items and accounted for 4.552% of the variance. It was labeled "Reliability of SETs" and was represented by the highest two loading items: "*The size of the class affects the students' ratings" and "Unhappy students show their feelings by writing negative evaluation*".

The seventh factor accounted for 4.462% of the variance and was loaded only on two items, suggesting it was measuring the "Job security", and was represented by the high loading items: "*An untenured faculty member who does not score well on the SETs, has to start looking for a job" and "A tenured faculty member with low SET scores receives insulting letters & minimal pay increments*".

Likewise, factor 8 was loaded on 3 items and accounted for 3.474% of the variance. It was labeled "Play to the audience feeling" and was represented by the highest loading item: "*Students' ratings encourage faculty members to be lenient in their grading*". On the other hand, factor 9 was loaded on a single item and accounted for 3.322% of the variance. It was labeled "Gender influence" and was represented by the unique loading item: "*The gender of the student affects (faculty member's) SET scores*".

Lastly, factor 10 was loaded on 3 items and accounted for 2.869% of the variance. It was labeled "Validity sentiment" and was represented by the high 3 loading items: "*Students do not have enough knowledge to judge the quality of instruction*", "*faculty members who demand a lot from their students get low evaluations*", and "*Students tend to agree among themselves regarding the evaluations*". Tables 11 and 12 show the corresponding factor analysis results.

Table 11. *Total Variance Explained for the 36 Attitudes toward Course SETs*

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.964	19.343	19.343	6.964	19.343	19.343	3.999	11.108	11.108
2	3.858	10.715	30.059	3.858	10.715	30.059	3.116	8.657	19.764
3	2.473	6.869	36.928	2.473	6.869	36.928	2.931	8.142	27.906
4	1.966	5.461	42.388	1.966	5.461	42.388	2.738	7.605	35.511
5	1.840	5.110	47.499	1.840	5.110	47.499	2.181	6.059	41.570
6	1.639	4.552	52.051	1.639	4.552	52.051	2.074	5.760	47.330
7	1.606	4.462	56.513	1.606	4.462	56.513	1.860	5.166	52.496
8	1.251	3.474	59.987	1.251	3.474	59.987	1.848	5.134	57.630
9	1.196	3.322	63.309	1.196	3.322	63.309	1.550	4.306	61.936
10	1.033	2.869	66.179	1.033	2.869	66.179	1.527	4.242	66.179
11	.970	2.696	68.874						
12	.927	2.575	71.450						
13	.876	2.434	73.884						
14	.772	2.146	76.029						
15	.752	2.088	78.117						
16	.677	1.882	79.999						
17	.639	1.774	81.773						
18	.612	1.700	83.473						
19	.593	1.647	85.120						
20	.534	1.484	86.604						
21	.528	1.466	88.070						
22	.479	1.332	89.402						
23	.468	1.299	90.701						
24	.413	1.148	91.849						
25	.396	1.101	92.950						
26	.380	1.056	94.006						
27	.329	.914	94.920						
28	.310	.862	95.782						
29	.288	.800	96.582						
30	.246	.683	97.265						
31	.229	.636	97.901						
32	.208	.577	98.478						
33	.157	.437	98.915						
34	.149	.415	99.330						
35	.140	.390	99.720						
36	.101	.280	100.000						

Extraction Method: Principal Component Analysis.

Table 12. *The loadings of the 36 SET attitude items on ten factors.*

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.	Component									
	1	2	3	4	5	6	7	8	9	10
Students evaluate faculty member based on how much they have learned in the course	.743									
Students evaluate faculty member based on their teaching ability	.709									
Good faculty member gets high course evaluations	.699									
Students' ratings correctly reflect the quality of instruction	.681									
Students evaluate faculty member based on character rather than on the teaching	-.635				.406					
One does not have to be a good faculty member in order to get positive evaluations	-.563				.456					
Most students treat course evaluations seriously	.525									
Students' ratings are an acceptable criterion for granting promotion		.827								
Students' ratings are an acceptable criterion for granting salary increases		.819								
Students' ratings are an acceptable criterion for granting tenure		.815								
It is research performance that identifies a good faculty member not the teaching			.733							
Academic policy of using SET as an evaluation is killing high quality teaching			.710							
Students' ratings reduce faculty's morale and job satisfaction			.666							
Students' ratings pose a threat to academic freedom			.651							
Administering teacher & course evaluation every semester is a waste of time			.511		.					
SET systems do not encourage good teaching			.503							
Students write comments only when they feel very positive about the instructor			.402							
The time of the day the course is offered affects students' ratings				.759						
Level of course (freshman, sophomore, junior, senior, graduate) affects students' ratings				.704						
Whether students are majoring or non-majoring in a course affects their ratings				.584						
The rank of the faculty member affects students' ratings				.558						
Students give better ratings to faculty member that teach less demanding courses				.453						
Students' ratings measure how nice a faculty member teacher is					.781					
Faculty members tend to reduce their requirements to get favorable ratings					.608					
Students base their course's rating on how entertaining a faculty member teacher is	-.445				.600					
The size of the class affects the students' ratings						.753				
Unhappy students show their feelings by writing negative evaluations						.681				
A tenured faculty member with low SET scores receives insulting letters …							.783			
An untenured faculty member who does not score well on the SETs has to start looking for a job							.735			
Students' ratings encourage faculty to be lenient in their grading								.720		
Students' ratings can seriously jeopardize the career of a faculty member.								.640		
Required courses generally get lower evaluations than elective courses.								.447		
The gender of the student affects the faculty members' SET scores.									.798	
Students do not have enough knowledge to judge the quality of instruction.										.792
Faculty member who demand a lot from their students get low evaluations.										.548
Students tend to agree among themselves as to the evaluations										.547

4.5.1 Perceptions of respondents by gender, academic rank, status, and institution

The Instrumental Dimension

The factor analysis part shows that the *Instrumental dimension* is represented by the dominant seven items, being questions 1, 3, 7, 8, 17, 18, and 19 from the questionnaire, the section corresponding to 'Attitudes toward Course Evaluation'. The highest loading corresponds to item number 18, "*Students evaluate faculty member based on how much they have learned in the course*" with a load of 0.743. On carrying a chi-square test for the gender dependency on the aforementioned item, number 18, the outcome is 0.696 for the value of the Pearson chi-square with p=0.952, which indicates a significant independence between the item and the gender. In fact, apart from the gender, the academic rank and status (part-time or full-time) of faculty member shows independency with Pearson chi-square 14.797 (p=0.788) and 4.995 (p=0.288), respectively. However, the chi-square test shows a dependency at 10% level of significance between the main instrumental dimension item and the type of institution, being an old university or a newly established university (Chi-Square=14.820, p=0.063); it is clear that the responses obtained from the newly established universities' faculty members (those from AUST and LIU) tend to agree more with the instrumental dimension when compared to those obtained from their old universities' colleagues (AUB and LAU).

The Job Performance Dimension

The factor analysis shows that the *Job performance dimension* is represented by three items, being questions 9, 10, and 11, from the questionnaire section corresponding to 'Attitudes toward Course Evaluation'. The highest loading corresponds to item number 11, "*Students' ratings are an acceptable criterion for granting promotion*" with a load of 0.827. On carrying a chi-square test for the gender dependency on the aforementioned item, number 11, the outcome is 8.704 for the value of the Pearson chi-square with p=0.069, which indicates a significant dependency at 10% level of significance between the item and the gender. Furthermore, the academic rank of the faculty member shows independency with Pearson chi-square 18.956 (p=0.525) and dependency at 10% level of significance for the status of instructor with a chi-square equal to 8.623 (p=0.071). Additionally, the chi-square test shows a dependency at 5% level of significance between the *Job performance dimension* item and the type of institution, being an old or a newly established university (chi-square= 20.229, p= .010); it is clear from the responses obtained that faculty members at old universities, in comparison with their colleagues at the newly established universities, strongly oppose the idea that SETs play a major role in their academic careers.

The Utility Dimension

The factor analysis concludes that the *Utility dimension* is represented by seven items, namely questions 12, 13, 22, 23, 29, 30 and 31, from the questionnaire section corresponding to 'Attitudes toward Course Evaluation'. The highest loading corresponds to item number 30, "*It is research performance that identifies a good teacher not the teaching method*" with a load of 0.733.

On carrying a chi-square test for the gender dependency on the aforementioned item, number 30, the result obtained is 2.30 for the value of the Pearson chi-square with p=0.681, which indicates a statistically significant independence between the item and the gender. Moreover, no statistically significant relationships are found between the *Utility dimension* and the respondents' characteristics of the faculty's academic rank (chi-square=17.015, p=0.652), status (chi-square=7.656, p=0.105), and type of university (chi-square=10.844, p=0.211).

The Ego-Defense Dimension

The factor analysis part shows that the *Ego-defense dimension* is represented by five items, being questions 15, 33, 34, 35, and 36, from the questionnaire section corresponding to 'Attitudes toward Course Evaluation'. The highest loading corresponds to item number 34, "*The time of the day the course is offered affects students' ratings*" with a load of 0.759.

On carrying a chi-square test for the gender dependency on the aforementioned item, number 34, the result obtained is 3.962 for the value of the Pearson chi-square with p=0.411; this indicates a significant independence between the item and the gender. Moreover, no significant relationships are found between the *Ego-defense dimension* and the faculty members' academic rank (chi-square=25.41, p=0.186), and being on a full-time/part-time basis (chi-square=1.706, p=0.79). In addition, no dependency is detected between the *Ego-defense dimension* and the type of university; being old or newly established (chi-square=9.5, p=0.302).

The Trust Dimension

Along the same aforementioned approach, the factor analysis part demonstrates that the *Trust dimension* is dominated by three items, being questions 6, 14, and 16, from the questionnaire section corresponding to 'Attitudes toward Course Evaluation'. The highest loading corresponds to item number 6, "*Students' ratings measure how nice a faculty member is*" with a load of 0.781.

On carrying a chi-square test for the gender dependency on the aforementioned item, number 6, the outcome is 1.485 for the value of the Pearson chi-square with p=0.829; this indicates a significant independence between the item and the gender. Additionally, the academic rank of the faculty member shows independency with Pearson chi-square 11.362 (p=0.936), while the academic status of faculty member shows a dependency at 10% level of significance with Chi-square= 8.735 (p=0.068). Moreover, with chi-

square equal to 11.362 (p=0.936), the test shows independency between the main trust dimension and being a faculty member at an old or newly established high education institution.

The Reliability of SETs Dimension

The factor analysis performed within this current research concludes that the *Reliability of SETs dimension* is represented by two items, being questions 24 and 28, from the questionnaire section corresponding to 'Attitudes toward Course Evaluation'. The highest loading corresponds to item number 24, "*The size of the class affects the students' ratings*" with a load of 0.753. On carrying a chi-square test for the gender dependency on the aforementioned item, number 24, the outcome is 2.245 for the value of the Pearson chi-square with p=0.691; this indicates a significant independence between the item and the gender. Furthermore, the academic rank of the faculty shows independency with Pearson chi-square 22.253 (p=0.327), also the academic status of the institution shows independency with a chi-square equal to 6.884 (p=0.142). Additionally, the chi-square test shows a dependency at 5% level of significance between the main reliability of SETs dimension item and the type of institution (chi-square= 22.790, p= .004); it is clear from the responses obtained that faculty members at newly established universities oppose item 24 more than their peers at old established universities.

The Job Security Dimension

The factor analysis part shows that the *Job Security dimension* is represented by two items, being questions 25 and 26, from the questionnaire section corresponding to 'Attitudes toward Course Evaluation'. The highest loading corresponds to item number 26, "*A tenured faculty member with low SET scores receives insulting letters and minimal pay increments*" with a load of 0.783.
On carrying a chi-square test for the gender dependency with the aforementioned item, number 26, the result obtained is 2.739 for the value of the Pearson chi-square with p=0.434; this indicates a significant independence between the item and the gender. Moreover, no significant relationships are found between the *Job security dimension* and the respondents' academic rank (chi-square= 13.107, p=0.594). A significant dependency at 10% level of significance is detected between item 26 and the academic status of the faculty member (full-time or part-time) where chi-square is 6.471 (p= .091). Additionally, no dependency is detected between the *Job Security dimension* and the type of university (chi-square= 4.878, p=0.560).

The Play to the Audience Feeling Dimension

The factor analysis part shows that the *Play to the audience feeling dimension* is represented by three items, being questions 5, 20 and 21, from the questionnaire section corresponding to 'Attitudes toward Course

Evaluation'. The highest loading corresponds to item number 20, "*Student ratings encourage faculty to grade easier*" with a load of 0.720. On carrying a chi-square test for the gender dependency on the aforementioned item, number 20, the outcome is 0.436 for the value of the Pearson chi-square with p=0.979; this indicates a significant independence between the item and the gender. Indeed, apart from the gender, the academic rank and status (part-time or full-time) of the faculty member show independency with Pearson chi-square 22.617 (p=0.308) and 3.612 (p=0.461), respectively. However, the chi-square test shows a dependency at 5% level of significance between the main *Play to the audience feeling dimension* item and the type of institution (chi-Square= 17.218, p=0.028); it is clear that the responses obtained from newly established university's faculty members tend to agree more with the *Play to the audience feeling dimension* when compared to their colleagues at old-established universities.

The Gender Influence Dimension

The factor analysis part shows that the *Gender influence* dimension is represented by only one item, being question 32, from the questionnaire section corresponding to 'Attitudes toward Course Evaluation': "*The gender of the student affects my SET's score*s". On carrying a chi-square test for the gender dependency on the aforementioned item, number 32, the outcome is 2.901 for the value of the Pearson chi-square with p=0.525; this indicates a significant independence between the item and the gender of the respondent. In fact, this result implies that faculty members of both genders do not believe that student gender may influence their SET scores. Likewise, the academic rank and status of faculty members show independency with Pearson chi-square 12.482 (p=0.898) and 5.617 (p=0.230), respectively. Additionally, the chi-square test shows another independency between the gender influence dimension item and the type of institution (chi-square= 12.776, p= 0.120).

The Validity Sentiment Dimension

The factor analysis part confirms that the *Validity sentiment dimension* is represented by three items, being questions 2, 4 and 27, from the questionnaire section corresponding to 'Attitudes toward Course Evaluation'. The highest loading corresponds to item number 2, "*Students do not have enough knowledge to judge the quality of instruction.*" with a load of 0.792. On carrying a chi-square test for the gender dependency on the aforementioned item, number 2, the outcome is 1.745 for the value of the Pearson chi-square with p=0.783; this indicates a significant independence between the item and the gender. Likewise, the academic rank of faculty members shows independency with Pearson chi-square 18.813 (p=0.534), but the faculty member's status shows a dependency at 5% level of significance with the considered

item (chi-square= 10.006, p=0.040). Additionally, the chi-square test shows another independency between the *Validity sentiment dimension* item and the type of institution (chi-square= 3.589, p= 0.892).

The Feedback Informative Dimension

The factor analysis performed on the nine items corresponding to Usefulness of Course Evaluations show that the *Feedback informative dimension* is represented by four items, being questions 3, 7, 8 and 9. The highest loading corresponds to item number 7, "*In general the current evaluations do not provide any useful information.*" with a load of 0.739. On carrying a chi-square test for the gender dependency on the aforementioned item, number 7, the outcome is 2.862 for the value of the Pearson chi-square with p=0.581; this indicates a significant independence between the item and the gender. Furthermore, both of the academic rank and academic status of the faculty show independency with item 7 as seen by the resulting Pearson chi-square values 25.809 (p=0.172) and 5.946 (p=0.203), respectively. Additionally, the chi-square test shows a dependency at 10% level of significance between the main *Feedback informative dimension* item and the type of institution (chi-square= 14.554, p= .068); it is clear from the responses obtained that faculty members at newly established universities support the *Feedback informative dimension* more than their colleagues at old universities.

The Help to Faculty Dimension

Again, the factor analysis performed on the nine items corresponding to Usefulness of Course Evaluations show that the *Help to faculty members dimension* is represented by three items, being questions 1, 2 and 4. The highest loading corresponds to item number 2, "*Faculty makes significant changes to course content based on students' evaluations*" with a load of 0.694. On carrying a chi-square test for the gender dependency with the aforementioned item, number 2, the outcome is 2.554 for the value of the Pearson chi-square with p=0.635; this indicates a significant independence between the item and the gender. Indeed, apart from the gender, the academic rank and status (part-time or full-time) of faculty member show independency with Pearson chi-square 19.789 (p=0.471) and 7.557 (p=0.109), respectively. Moreover, the chi-square test shows another independency between the main *Help to faculty members dimension* item and the type of institution (chi-square= 9.586, p=0.295).

The Hope and Expectancy Dimension

Similarly, the factor analysis performed on the nine items corresponding to Usefulness of Course Evaluations show that the *Hope and Expectancy Dimension* is represented by two items, being questions 5 and 6. The highest loading corresponds to item number 6, "*If SETs were given at an earlier point in the semester, I would use the students'*

feedback immediately" with a load of 0.863. On carrying a chi-square test for the gender dependency with the aforementioned item, number 2, the outcome is 1.869 for the value of the Pearson chi-square with p=0.760; this indicates a significant independence between the item and the gender. In addition to gender, the academic rank and status (part-time or full-time) of faculty show independency with Pearson chi-square 22.957 (p=0 .291) and 5.946 (p=0. .225), respectively. Moreover, the chi-square test shows another in independence between the main *Hope and Expectancy Dimension* item and the type of institution (chi-square= 7.092, p=0.527).

Here, it is worth summarizing all the previous chi-square tests by indicating that there are minimal differences within the faculty members' responses as to rank and position which really coincides with the findings and results revealed in the work of Kogan et al. (2010).

4.6 Respondents' opinions on who should receive the SET results?

The questionnaire on faculty members' perceptions of SETs presents different options to answer the question: *Who should directly receive the results of course evaluation?* These options are course faculty member, students, administration, students and administration, faculty member and administration, and all of the faculty, students and administration. The collection of answers obtained from the respondents is summarized in Table 13.

Table 13. *Who should directly receive the results of course evaluation?*

		Frequency	%	Valid %	Cumulative %
Valid	Course faculty member only	62	42.5	**51.2**	51.2
	Students only	1	.7	.8	52.1
	Administrators only	12	8.2	9.9	62.0
	Faculty members, students & administrators	13	8.9	10.7	72.7
	Faculty members & administrators	32	21.9	**26.4**	99.2
	Students & administrators	1	.7	.8	100.0
	Total responses received	121	82.9	100.0	
	Missing responses	25	17.1		
Total respondents		146	100.0		

%: Percent.

The frequencies presented in Table 13 clearly demonstrate that more than 50% of the received responses tend to exclusively name the course's instructor as the sole destination where the results of SETs should go. In the second place, 26.4% of the respondents' name both the course's instructor and administrators as the parties who should receive SET results.

Respondents' Comments

The questionnaire "Faculty Perceptions of SETs" included two spaces for two open-questions allowing the respondents to include the personal changes they have made to improve their SET scores and their own opinions related to SETs. In what relates to changes made, the following was quoted in Exhibit 3.

Exhibit 3. *Modifications instituted by faculty members*

Action	Remark
Adding videos to a certain chapter, or focusing more on certain subject sections…	Students' advise for improvement via SETs
Adding problem-solving and Q/A sessions	Students' advise for improvement via SETs
Give less reading material	No change on SET scores
Bring to class more real-life examples	Applying theory to practice
Improving relations with students by being nice, listen to their feedback, have a positive attitude	Timing around SET period
Adding solved problems sessions	Avoiding what students call mismatch between exams and class contents
Removing events, adding office hours and changing the style	Creating positive influence with students
Making classes more interactive and encouraging students to share concerns	Improve SET scores
Performing a mid-semester evaluation	Showing readiness to add changes
Diluting the material, easier exams, similar exams every term, and high raises	Influencing students' opinions

The aforementioned proposed changes to improve SET scores mainly revolve around course load adjustment, extra exam examples, diluted materials, easy exams, high raises, similar exams every term, change of style, and try to be nice. Indeed, with such changes put in action, one of the respondents' position led to the disappointing comment: "*I feel a bit like I'm training monkeys*".

Exhibit 4 depicts direct comments highlighted as personal opinions on the SET evaluations.

Exhibit 4. Faculty members' personal opinions about SETs

- *Institutions should use a normalized SET score where a student's evaluation of a faculty member is to be manipulated by the expected course grade.*
- *I feel that my institution gives too much importance to making students happy; lots of good faculty member have been fired simply because a student with a lot of money had complained.*
- *SET questions must be modified or adjusted based on the level of students (freshman, sophomore, junior, or senior).*
- *Faculty members need to be aware of students' evaluations immediately after they are processed in order to implement any needed changes.*
- *SET scores are essential, but Peer Observation Reports better represent the reality.*
- *Negative SETs are used by administrators to nail those faculty members they are not on good terms with…*
- *SET scores should be considered as indicators rather than exact measures of faculty members' performances.*
- *SET evaluations are important because some faculty member is not doing a good job.*
- *SETs are good because students should express their views. Just don't know how true or valid and how serious students take it.*
- *With all due respect to SETs, I believe that SET scores will never be fair in a country like Lebanon.*
- *Some students rate randomly without reading the items. Many of these unserious students put punctuation marks like "!? /; ^" as answers. Their objective is to simply access their grades. The fact is that at my institution, completing a SET is essential to having access to the final grades.*
- *SETs are highly affected and biased due to the few numbers of students who complete them. Moreover, some students rate their faculty members based on their grades instead of rating the teaching ability or quality of instruction.*
- *I think that both faculty and administrators should work together to set the evaluation criteria. Sometimes they change them between semesters and then faculty members don't know what they are judged on. They also think that teaching should be judged by having peers sit in the class and evaluate the sessions.*
- *Some students' evaluation doesn't make sense. For example, two similar questions are answered differently (No congruence in answers).*
- *I believe that students' evaluations are based on grades rather than the quality of teaching.*
- *Students' ratings are very subjective and universities usually force their students to do these*

> *evaluations; this fact makes the evaluations unreliable data.*
> - *The SET evaluates the student's seriousness more than the tutor's.*
> - *Quality students produce quality in SET's scores.*

Source: Hejase et al. (2015).

It is noticed from the aforementioned written comments that at least twelve of them carry a negative view of SETs and a clear disillusionment is embedded in a respondent's comment: "*SET scores will never be fair in a country like Lebanon*".

5. Discussion

This research presented five main hypotheses that are deeply answered within the context of data analysis. In what follows, each hypothesis (test details presented earlier, in a section of descriptive statistics) will be presented in conjunction with the results reached.

5.1 Faculty members have a negative attitude towards SETs

The research instrument presented a section titled "Attitudes toward Course Evaluation" which included 36 items that, after factor analysis, have been condensed into ten dimensions as per the following factors:

The "*Instrumental dimension*" mainly represented by item 18: "*Students evaluate faculty members based on how much they have learned in the course*". The mean response is 3.34 with a p-value =.000 < α = 5%; hence, the null hypothesis is rejected, meaning that the fact that students evaluate faculty based on how much they have learned in the course is accepted. This result contradicts the first hypothesis of negative attitude towards SETs. Indeed, this conclusion agrees with the findings of Schmelkin et al. (1997) in that SETs are a useful means for both academic formative and summative purposes. Likewise, Rosemarin (2010, p. 234) concludes that there is a high positive correlation between the level of students' understanding of the learned material and their evaluation of the corresponding faculty members.

The "*Job performance*" represented by item 11: "Student ratings are an acceptable criterion for granting promotion". The mean response of this item is 2.60 with a p-value =.000 < α = 5%, so, the null hypothesis is not accepted, meaning that the fact that in general, the faculty's responses present a negative attitude towards using SETs as a promotion criterion is accepted. Here, it is worth mentioning that Table 1 indicates that 50% of the respondents Strongly Disagree or Disagree with item 11, 28.5% of the responses are Neutral and only 21.5% of the respondents Agree or Strongly Agree with the item. Indeed, Turpen et al. (2011) raised similar issues and concluded that SETs results remain questionable. In her study Zakka (2009, p. 243) report that her respondents agreed in that SET results are not a fair weighing

mechanism to be used for promotion and tenure decisions as per the results of this present study.

The "*Utility dimension of SETs to faculty members*" represented by item 30: "It is research performance what identifies a good faculty, not the teaching methodology". The mean of this item is 2.55 with a p-value =.000 < α = 5%, so, the null hypothesis is rejected, meaning that the fact that no negative attitude towards SETs is accepted; this implies that faculty members do not recognize SET as a significant tool to evaluate their teaching. The work of Costin et al. (1971, p. 530) clearly conclude that the claim that teaching effectiveness and research productivity go hand in hand, is a frail one. Moreover, according to Costin et al., students raised a counterclaim in that faculty are often so busy in the "publish or perish" symptom that they pay little attention to their teaching duties.

The "*Ego-defense*" was highly loaded on item 34: "The time of the day the course is offered affects students' ratings". The mean of this item is 2.92 with a p-value =.326 > α = 5%, so, the null hypothesis is accepted, and the fact that no negative attitude towards indicating that there are other issues involved in SETs that may bias results and consequently are used by faculty members as defense mechanisms against the evaluation results is not accepted.

The "*Trust in SETs*" came up to be highly loaded on item 6: "Student ratings measure how nice a faculty member is". The mean of this item is 3.46 and with a p-value =.000 < α = 5%, so, the null hypothesis is rejected, meaning that the fact indicating the sarcasm faculty members carry towards SETs in that they do measure the "niceness" of an instructor as seen from a biased student point of view is accepted. Sojka et al. (2002) reached similar results and concluded that students do give higher SET scores to easy and entertaining faculty members.

Under the "*Reliability of SETs*" dimension which is highly loaded on item 24, "The size of the class affects the students' ratings"; the mean is 3.53 with a p-value =.000 < α = 5%, so, the null hypothesis is rejected, meaning that the fact which leads the faculty members to fear the reliability of SETs is also rejected. The work of Nowell (2007, p. 51) concludes that class size and faculty member's SET rating are inversely related. Similar conclusions have been reached by Balam and Shannon (2010) who report that in fact, SETs are invalid and unreliable. In contrast to the attained results of this study, the pioneer work of Heilman and Armentrout (1936, p. 215) concluded that class size, among many other factors like severity of grading, sex of the teacher, and maturity of the rater as evidenced by the collected data, have no effect on the ratings.

As for the "*Job security*" dimension which is represented by item 26: "A tenured faculty member with low SET scores receives insulting letters and minimal pay increments"; the mean is 2.32 with a p-value =.000 < α = 5%, so, the null hypothesis is rejected, indicating that faculty members do not agree with the fact that SETs may affect the permanent status of their job.

The "*Play to the audience feeling*" dimension is represented by item number 20, "Students' ratings encourage faculty to be lenient in their grading"; the mean of this item is 3.11 with a p-value =.207 > α = 5%, so, the null hypothesis is accepted, indicating that faculty members go along with the fact that SETs do encourage faculty members to comply with the students; whims. Again, it is worth mentioning that Balam and Shannon (2010) do not hesitate to write that SETs are a popularity contest for the faculty members where the winners are those who are more friendly and humorous. Likewise, Stratton and his collaborators (1994) found that grades were inflated by 11% after implementing the SET mechanism. Similarly, Villard (1973) noticed that as SET scores go higher, grades tend to drift upwards as the level at which a course is taught slides downwards. In the same way, Eiszler (2002, p. 498) conclude, after having done a longitudinal research study that the semesters in which faculty members got higher SET ratings are the semesters in which they were awarded higher percentages of A and A- grades. Additionally, Nowell (2007, p. 54) report that faculty is able to 'buy' higher SET ratings by giving higher grades. Once more, the findings of Ewing (2012, p. 150) indicate that no matter the estimation procedure, there is a significant positive correlation between evaluation scorers and students' expected grades.

The "*Gender influence*" dimension is represented by item 32, from the questionnaire section corresponding to 'Attitudes toward Course Evaluation': "The gender of the student affects the SET scores". The mean of this item is 2.51 with a p-value =.000 < α = 5%, so, the null hypothesis is rejected, indicating that faculty members do not agree with the fact that a student's gender may affect their SET scores. A contradictory result was reported by Basow (1995, p. 661) and Basow and Martin (2012, p. 40) who indicate that faculty gender appears to interact with the gender of the rater to influence the SET scores.

Finally, the last dimension under 'Attitudes toward Course Evaluation' is "*Validity sentiment*" which is heavily loaded on item number 2: "Students do not have enough knowledge to judge the quality of instruction". The mean of this item is 3.06 with a p-value =.477 > α = 5%, so, the null hypothesis is rejected, indicating that faculty members do question the judging ability of students and consequently the validity of the SET scores. Such results did show up in Simpson & Siguaw's work (2000) where it is reported that 48.1% of the faculty respondents believe that SETs are somehow inaccurate, while only 42.35 believes that they are really accurate, and 92.3% of the respondents report that SET ratings are important. In addition, Surgenor's study (2013, p. 368) concludes that most of the faculty respondents regard validity-related concerns to be a considerable barrier to SETs. A study conducted at The Lebanese American University (Zakka, 2009, p. 243) conclude that SETs are not a valid tool for measuring instruction.

In summary, for the first hypothesis, faculty members presented a negative attitude in seven of the aforementioned dimensions, and only in three dimensions, did they present a positive attitude; these three dimensions are Instrumental, job security, and gender influence.

5.2 Faculty members believe that SETs are not useful.

The research instrument presented a section titled "Usefulness of Course Evaluation" which included nine items that, after factor analysis, are condensed into three dimensions as per the following factors:

The "*Feedback informative dimension*" is represented mainly by item number 7 of the nine items corresponding to the usefulness of course evaluations, "In general the current evaluations do not provide any useful information". The mean of this item is 2.89 with a p-value =.188 > α = 5%, so, the null hypothesis is accepted, indicating that faculty members do agree with the fact that their SET scores do not provide useful information. In the same way, Turpen and her collaborators (2011, p. 374) indicate that SETs are questionable and even faulty.

The "*Help to dimension*" is represented by item number 2, from the nine items corresponding to the usefulness of course evaluations, "Faculty makes significant changes to course's content based on students' evaluations". The mean of this item is 3.10 with a p-value =.221 > α = 5%, so, the null hypothesis is accepted, indicating that faculty members do use their SET scores to modify their course's contents. Similarly, the study performed at AUB (American Univeristy of Beirut, 2005) concluded that 50% of the 145 respondents believe that faculty members change their teaching style to receive higher SET ratings. The work of Wilson and Ryan (2012, p. 27) conclude that students often share valuable feedback that permits faculty members to adjust their courses in some way or another. On the other hand, besides benefiting faculty members, Chan and her collaborators (2014, p. 282) conclude that faculty members are aware that lowering their courses' standards, particularly on students' requests, will contribute to achieving higher SET scores.

Likewise, the factor analysis performed on the nine items corresponding to the usefulness of course evaluations shows that the "*Hope and Expectancy Dimension*" is mainly represented by item number 6, "If SETs were given at an earlier point in the semester the faculty member would use the students' feedback immediately". Again, the mean of this item is 3.26 with a p-value =.002 < α = 5%, so, the null hypothesis is rejected, indicating that faculty members do have expectations as to the immediate tangible benefits that SETs could provide to them.

In general, the results about the usefulness of SETs are not so pessimistic; in fact, they coincide with many of the published works, for example, the work of Schmelkin et al. (1997, p. 588) whose conclusion is: "Contrary to what might have been deduced from anecdotal literature, the

results of this study do not portray a great deal of resistance to student ratings in general nor to their use…".

5.3 SETs are not the main tool to measure teaching effectiveness.

This hypothesis was answered through the question: "*Faculty's colleagues with excellent publication records and expertise are better qualified to evaluate their peers' teaching effectiveness*". The mean response for this question is 2.89 with a p-value =.245 > α = 5%, so, the null hypothesis is accepted, indicating that faculty members give less importance to SETs as a mechanism for the evaluation of teaching effectiveness. In fact, such conclusion was reached by many of whom we cite Engelland (2004, p. 44) who wrote: "Assessment is an on-going process and should not be relegated to an end-of-semester activity". Likewise, Chang McKeachie, & Lin (2010, p. 217) concluded that in particular, colleagues from the same department may suggest teaching strategies that can help in dealing with issues that need improvement. The work of Alauddin and Kifle (2014, p. 164) implied that SETs are inadequate to measure teaching effectiveness in that faculty members are able to manipulate these ratings by engaging in less than scholarly teaching practices. Similar results were obtained by Zakka (2009, p. 243) who concluded that SETs are not the only method to evaluate teaching effectiveness and other methods like peer review are to be considered.

5.4 Students are not able to make accurate SET judgments.

This hypothesis is considered under the question: "*Students are not able to make accurate judgments until they have been away from the course and possibly away from the university for several years*"; the mean response achieved is 2.98 with a p-value =.823 > α = 5%, so, the null hypothesis is accepted. Thus, the hypothesis that students are not able to make accurate SET judgments cannot be rejected at 5% level of significance. This finding contradicts the study done by Zakka (2009, p. 243), where it was concluded that students do not possess the maturity and knowledge to answer certain questions like those related to the knowledge of the instructor of the subject matter.

5.5 SETs are a double-edged sword used by the administration as per their convenience.

The last hypothesis was dealt with through the question: "*Low student evaluations are a double-edged sword that is used by administration as per their convenience, ignoring them as supportive comments and observing them as detractors*"; the mean response of the answers is 3.29 with a p-value =.001 < α = 5%, so, the null hypothesis is rejected, indicating that in fact, SETs are a double-edged sword used by administrators as per their own convenience, however depending on the university's

culture and the set purpose by administration on how to deal with SETs.

6. Conclusion:

This paper is the first contemporary research in Lebanon to address the point of view and assess the attitude of faculty members towards SETs, involving several universities. Its first merit emanates from the aforementioned fact; in addition, results and outcomes enrich the current scarcity of information about the topic. The aim is to explore the attitudes of faculty members teaching within selected Lebanese universities. The empirical assessment and findings enlighten interested stakeholders as to the extent of faculty's frustration due to the observed university practices in using SETs. This research, with its extensive results, leads to the conclusion that the majority of faculty members at Lebanese universities have very serious reservations against the SET outcomes. In fact, the major findings of this research, agree with the findings of many other researchers like those in the work of Balam and Shannon (2010), where they concluded that "Professors were more likely to discredit students' ratings as a valid and reliable source of effective teaching. Faculty, did, however, believe that student ratings could be useful in improving instruction" (p. 128). Likewise, McMartin and Rich (1979, p. 150) concluded that the majority of faculty seem to take a "wait and see" attitude toward SETs, and that faculty's opinions are divided based on their view of the validity of the SET instrument. Similarly, the different opinions of this current research go in parallel with the recent findings of Sojka and her collaborators (2002, p. 48) who conclude that even though there are controversial issues surrounding SETs, yet they remain useful and their elimination is undesirable. This latter fact is clearly shown in the results of this study; results that were previously asserted by Barnett and Mathews (1997, p. 350) who concluded that SETs affect instructional practices in a sizable and desirable way. In fact, SETs are tied to the quality of students as one respondent commented: "*Quality students produce quality in SET scores*". The main problem is that universities continue to use SETs in their administrative decisions that obviously affect their faculty. A fact confirmed by Johnson (2002) who contends that "the opportunities for faculty to manipulate their teaching style in general and grading policies in particular in order to enhance their evaluations really tend to increase (p. 16). But doubted by McPherson (2003) who asserts that "A principal finding is that there is no strong evidence that SET scores are 'contaminated' by faculty members attempting to 'buy' better SET scores by raising grade expectations" (p. 15). Ultimately, SETs should not be the potential minefield for faculty members; they ought to be reshaped to reward teaching that really challenges students to think outside of the box (Schneider, 2013, p. 133).

Researchers in this paper recommend universities not to consider SETs as unique tools to assess faculty's promotions, rewards or others based on the apparent concerns manifested herein, but to make sure first that SET results are unbiased and investigate awkward results which may not represent actual behavior within the classrooms. In fact, McPherson et al. (2009) have found that their research results suggest that university departments may usefully consider "adjusting rankings to account for factors that can be manipulated by faculty members to their advantage (especially expected grade)" within SET scores and "factors that are beyond the control of the instructor (for example, race, and gender)" (p. 48). Further, McPherson et al. assert "that such adjustment to rankings can lead to statistically significant changes in SET score rankings, a result that has clear implications for promotion, tenure, and salaries" (ibid).

The knowledge from this study will guide university policy makers in formulating and adjusting policies to promote sound and fair use of SET results and foster the faculty members self-improvement efforts accompanied by an adequate reward system, a fact that will boost both the students' and the faculty members' performance which will, in turn, promote the effective achievement of learning outcomes. Furthermore, this study serves as an eye-opener to other researchers as it provides useful guides concerning empirical results emanating from several universities. Also, this research work serves as a reference material for further research on this field of study. In addition, it serves as a study material for both faculty members and students interested in this topic. Moreover, as the study guides policymakers in making and adjusting policies to achieve a more effective performance; it will help the university to grow results in an enhanced standard of teaching and learning of the current students and encourages the recruitment of new students.

7. Acknowledgment

Corresponding Author:
Hussin J. Hejase, Ph.D.
Faculty of Business Administration, Al Maaref University, Beirut, Lebanon
E-mail: hhejase@mu.edu.lb

References:
1. Alauddin, M., & Kifle, T. (2014). Does the student evaluation of teaching instrument really measure instructor's teaching effectiveness? An econometric analysis of students' perception in economics courses. *Economic Analysis and Policy*, *4*(2), 156-168.

2. Alves, H., & Raposo, M. (2009). The measurement of the construct satisfaction in higher education. *The Service Industries Journal*, *29*(2), 203-218.

3. The American University of Beirut. (2005). *Faculty and Student Perspectives on Student Teaching Evaluations.* Beirut: American University of Beirut.

4. Balam, E. M., & Shannon, D. M. (2010). Student Ratings of College Teaching: A Comparison of Faculty and Their Students. *Assessment & Evaluation in Higher Education*, *35*(2), 209-221.

5. Barnett, C. W., & Matthews, H. W. (1997). Student Evaluation of Classroom Teaching: A Study of Pharmacy Faculty Attitudes and Effects on Instructional Practices. *American Journal of Pharmaceutical Education*, *61*, 345-350.

6. Basow, S. A. (1995). Student Evaluations of College Professors: When Gender Matters. *Journal of Educational Psychology*, *87*(4), 656-665.

7. Basow, S. A., & Martin, J. L. (2012). Bias in Student Evaluations. In M. E. Kite, *Effective Evaluation of Teaching* (pp. 40-49).

8. Brandenburg, G. C., & Remmers, H. H. (1927). The Purdue Scale for Instructors. *Educational Administration and Supervision* (13), pp. 399-406.

9. Chan, C. K., Luk, L. Y., & Zeng, M. (2014). Teacher's perceptions of student evaluations of teaching. *Educational Research and Evaluation*, *20*(4), pp. 275-289.

10. Chang, T.-S., McKeachie, W., & Lin, Y.-G. (2010). Faculty Perceptions of Teaching Support and Teaching Efficacy in Taiwan. *High Education*, *59*, 207-220.

11. Costin, F., Greenough, W. T., & Menges, R. J. (1971, December). Student Ratings of College Teaching: Reliability, Validity, and Usefulness. *Review of Educational Research*, pp. 511-535.

12. Costin, F., Greenough, W. T., & Menges, R. J. (1973). Student Ratings of College Teaching: Reliability, Validity, and Usefulness. *The Journal of Economic Education*, 51-53.

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

14. Engelland, B. T. (2004). Making Effective Use of Student Evaluations to Improve Teaching Performance. *Journal for Advancement of Marketing Education*, *5* (Winter), 40-46.

15. Ewing, A. M. (2012). Estimating the Impact of Relative Expected Grade on Student Evaluations of Teachers. *Economics of Education Review*, *31*, 141-154.

16. Heilman, J. D., & Armentrout, W. D. (1936, March). The rating of College Teachers on Ten Traits by Their Students. *Journal of Educational Psychology*, *27*(3), pp. 197-216.

17. Hejase, A.J. & Hejase, H.J. (2013). *Research Methods: A Practical Approach for Business Students*, (2nd edition). Philadelphia, PA: Masadir Inc.

18. Hejase, Ale J., Al Kaakour, Rana S., Halawi, Leila A. & Hejase, Hussin J. (2013). Students' Perceptions of Student Evaluation of Teaching (SET) Process. *International Journal of Social Sciences and Education 3*(3), 565-575.

19. Hejase, Ale J., Hejase, Hussin J., & Kaakour Al, Rana S. (2014). The Impact of Students' Characteristics on their Perceptions of the Evaluation of Teaching Process. *International Journal of Management Sciences 4*(2), 90-105.

20. Hejase, A.J., Hejase, H.J., & Tarhini, A.M. (2015, November). Faculty Perceptions of Student Evaluation of Teaching Process. In *proceedings of the 3ème Colloque International Evaluation de la qualité des formations dans l'enseignement supérieur. Université Islamique du Liban*, November 30-December 1st, 2015. Beirut, Liban. DOI: 10.13140/RG.2.1.1895.3047. Retrieved April 1, 2018, from https://www.researchgate.net/publication/305719838_Faculty_Perceptions_of_Student_Evaluation_of_Teaching_Process

21. Hejase, Hussin J., & Alaeddine, Ali H. (2017). Fostering Academic Relations between Lebanon and the Islamic Republic of Iran. *The Journal of Middle East and North Africa Sciences, 3*(7), 22-38.

22. Higher Education. (2018). *Universities in Lebanon (in Arabic)*. Retrieved April 11, 2018, from http://www.higheredu.gov.lb/arabic/privuniv/personal_univ.html

23. Johnson, V. E. (2002). Teacher Course Evaluations and Student Grades: An Academic Tango. *Chance*, *15*(3), 9-16.

24. Kipps, P. H. (1975). The Use of Course Evaluation Scores to Influence Teaching and Research Activities. *The Journal of Economic Education*, 93-98.

25. Kogan, L. R., Schoenfeld-Tacher, R., & Hellyer, P. W. (2010). Student Evaluations of Teaching: Perceptions of Faculty Based on Gender, Position, and Rank. *Teaching in Higher Education*, *15*(6), 623-636.

26. Kotze', T. G., & du Plessis, P. J. (2003). Students as "co-producers" of education: a proposed model of student socialization and participation at tertiary institutions. *Quality Assurance in Education*, *11*(4), 186-201.

27. Lidice, A., & Saglam, G. (2013). Using students' evaluations to measure educational quality. *Procedia - Social and Behavioral Sciences*, 1009-1015.

28. Mark, E. (2013a). Student Satisfaction and the Customer Focus in Higher Education. *Journal of Higher Education Policy and Management*, *35*(1), 2-10.

29. Mark, E. (2013b). Students Are Not Products. They Are Customers. *College Student Journal*, *47* (3), 489-493.

30. McMartin, J. A., & Rich, H. E. (1979). Faculty Attitudes toward Student Evaluation of Teaching. *Research in Higher Education*, *11* (2), 137-152.

31. McPherson, Michael A. (2003, May). *Revisiting the Determinants of Student Evaluation of Teachers Scores*. Retrieved April 9, 2018, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=410934

32. McPherson, Michael A., Jewell, R. Todd, & Kim, Myungsup (2009). What Determines Student Evaluation Scores? A Random Effects Analysis of Undergraduate Economics Classes. *Eastern Economic Journal, 35*, 37–51.

33. 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), pp. 17-32.

34. Muncy, J. A. (2008). The Orientation Evaluation Matrix (OEM): Are Students Customers or Products. *Marketing Education Review*, *18*(3), 15-23.

35. Nasser, F., & Fresko, B. (2002). Faculty Views of Student Evaluation of College Teaching. *Assessment & Evaluation in Higher Education*, *27*(2), pp. 187-198.

36. Nowell, C. (2007). The impact of Relative Grade Expectations on Student Evaluation of Teaching. *International Review of Economics Education*, *6*(2), 42-56.

37. Rosemarin, S. (2010). The Evaluation of Lecturers and Students as Perceived by Both. *Gifted Education International*, *27*, 226-236.

38. Schmelkin, L. P., Spencer, K. J., & Gellman, E. S. (1997). Faculty Perspectives on Course and Teacher Evaluations. *Research in Higher Education*, *38*(5), 575-592.

39. Schneider, G. (2013). Student Evaluations, Grade Inflation, and Pluralistic Teaching: Moving from Customer Satisfaction to Student Learning and Critical Thinking. *Forum for Social Economics*, *42*(1), 122-135.

40. Simpson, P. M., & Siguaw, J. A. (2000). Student Evaluations of Teaching: An Exploratory Study of the Faculty Response. *Journal of Marketing Education*, 199-213.

41. Smith, B. P. (2007). Student Ratings of Teaching Effectiveness: An Analysis of End-of-Course Faculty Evaluations. *College Student Journal*, *41*(4), 1-14.

42. Sojka, J., Gupta, A. K., & Deeter-Schmelz, D. R. (2002). Student and Faculty Perceptions of Student Evaluations of Teaching: A Study of Similarities and Differences. *College Teaching*, *50*(2), pp. 44-49.

43. Stratton, R. W., Myers, S. C., & Kung, R. H. (1994). Faculty Behavior, Grades, and Student Evaluations. *Journal of Economic Education*, pp. 5-15.

44. Surgenor, P. W. (2013). Obstacles and opportunities: addressing the growing pains of summative evaluation of teaching. *Assessment & Evaluation in Higher Education*, *38*(3), pp. 363-376.

45. Turpen, C., Henderson, C., & Dancy, M. (2011). Faculty Perspectives about the instructor and Institutional Assessments of Teaching Effectiveness. *Physics Education Research Conference* (pp. 371-374). American Institute of Physics.

46. Villard, H. H. (1973). Some Reflections on Student Evaluation of Teaching. *The Journal of Economic Education*, *5*(1), 47-50.

47. Vuori, J. (2013). Are Students Customers in Finish Higher Education? *Tertiary Education and Management*, *19*(2), 176-187.

48. Wiely Periodicals. (2009, March). Faculty rejects bonuses based on student evals. (R. Rosenberg, Ed.) *Dean & Provost*, p. 2.

49. Wilson, J. H., & Ryan, R. G. (2012). Formative Teaching Evaluations: Is Student Input Useful? In M. E. Kite, *Effective Evaluation of Teaching* (pp. 22-29).

50. Winer, L. (1999, April 2). The pursuit of customer satisfaction ruins schools. *Marketing News*, p. 11.

51. Yaacoub, N., & Badre, L. (2012). *Education in Lebanon*. Beirut: Central Administration of Statistics.

52. Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in Higher Education*, *12* (1), 55-76.

53. Zakka, J. S. (2009). End of Semester Student Evaluation of Teaching Effectiveness Questionnaires: An Indicator of Teaching Quality. *Oxford Journal, Special Issue: Emerging Trends and Issues in Management and Economics*, *8*(1), 233-246.