HAIGAZIAN UNIVERSITY

RELATION OF INSTRUCTIONAL, PEDAGOGICAL AND COURSE CONTENT CHARACTERISTICS TO STUDENT RATINGS OF UNIVERSITY INSTRUCTORS' TEACHING EFFECTIVENESS

By
SAMAR HADDAD

A thesis
submitted in partial fulfillment of the requirements
for the degree of Masters of Arts
to the Department of Education
of the Faculty of Arts and Sciences
at Haigazian University

Beirut, Lebanon
June 2010

Thesis Release Form

I, Samar Haddad,

☑ Authorize Haigazian University to supply copies of my thesis to libraries or individuals upon request.

☐ Do not authorize Haigazian University to supply copies of my thesis to libraries or individuals for a period of two years starting with the date of thesis defense.

Samar H

Signature

June, 21, 2010

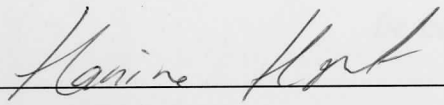
Date

HAIGAZIAN UNIVERSITY

Relation of Instructional, Pedagogical and Course Content Characteristics to Student Ratings of University Instructors' Teaching Effectiveness

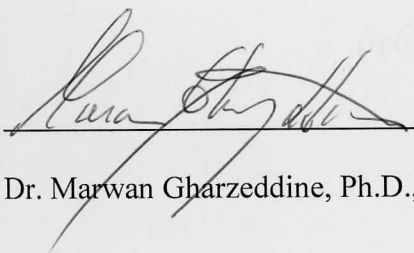
Samar Haddad

Approved by:



Dr. Hanine Hout, Ed. D., Advisor

Dr. Daoud Tawil, Ph.D., Reader



Dr. Marwan Gharzeddine, Ph.D., Reader

DEDICATION

To my husband, Baseel, whose love, warm personality and fine mind have made a wonderful contribution to my work

To my beloved son, Nicolas, for his great spirit of joy which brought a different meaning to my life

To the memory of my father, Kamil, the man whom I am proud to be called, daughter

To my mom, Wadiaa, for her unconditional love, support and encouragement

For my brothers and sisters who always looked high on me and believed in my potentials

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Hanine Hout, who guided me in the process of my thesis preparation, offered her professional advice, and encouraged me throughout. I appreciate all her efforts and willingness to go the 'Extra Mile' in time and space.

I am thankful to my two readers, Dr. Daoud Tawil and Dr. Marwan Gharzeddine who accepted to provide their advice and to contribute suggestions for my research.

Special thanks for Dr. Marwan Gharzeddine for his help and guidance in the research results and their analysis.

I am also thankful to Dr. Ghada Eid for her help in clarifying statistical data.

I would like to dedicate this thesis to my husband and to my family for their love and support throughout my work.

Honor and gratefulness be to God from whom guidance overwhelms.

TABLE OF CONTENTS

Page

ABSTRACT..... 1

1. INTRODUCTION..... 2

Background of the Study..... 3

Statement of the problem..... 8

- Student Characteristics and Student Evaluation of Instructors..... 8
- Instructor Characteristics and Student Evaluation of Instructors..... 11
- Course Content Characteristics and Student Evaluation of Instructors 12

Hypotheses..... 14

Professional Significance of the Study..... 14

Overview of Methodology..... 16

Limitations of the Study..... 17

Definition of Terminology..... 18

2. REVIEW OF LITERATURE..... 21

- Validity of Students' Ratings..... 22
- Relationship between Student Ratings of Instructors and Teaching Effectiveness..... 23
- Relationship between Student Ratings of Instructors and Students' Perceptions of Teacher Performance..... 26

- Relationship between Student Ratings of Instructors and Grades..... 29
- Relationship between Student Ratings of Instructors and Achievement..... 32
- Relationship between Student Ratings of Instructors and Gender..... 33
- Relationship between Student Ratings of Instructors and Class..... 35
- Relationship between Student Ratings of Instructors and Course Content Characteristics..... 36
- Relationship between Student Ratings of Instructors and Instructor Rank and Reputation 37
- Relationship between Student Ratings of Instructors and Distance from Teacher..... 37
- Relationship between Student Ratings of Instructors and Attendance..... 38

3. METHOD..... 40

4. RESULTS..... 44
- Descriptive Statistics..... 44
- Experimental Statistics..... 48

5. DISCUSSION..... 55

- Effect of Student Ratings on Instructors
  in Terms of Gender..... 55
- Effect of Student Course Content Characteristics
  on Student Ratings of Instructors..... 58
- Implications for the Previous Findings..... 59
- Effect of Student Course Required
  on Student Ratings of Instructors..... 60
- Effect of Student Level
  on Student Ratings of Instructors..... 62
- Effect of Student class size
  on Student Ratings of Instructors..... 63
- Effect of Student Major
  on Student Ratings of Instructors..... 63
- Limitations of the Study..... 64
- Conclusion and Further Research..... 65

REFERENCES..... 67

APPENDICES..... 75

Abstract

It is essential to determine if student ratings of instructors' teaching effectiveness reflect systematic bias due to instructor's gender. Students in five universities were asked to rate instructors in three factors: interpersonal characteristics, pedagogical characteristics and course content characteristics. Group differences were examined based on student gender, instructor gender, student level, and class size. Results indicated significant interactions between student gender and instructor gender in the course content characteristics but no differences were found in the interpersonal characteristics or pedagogical characteristics. Significant results were also found in the course content characteristics when the course was required. The findings were promising in the sense that they tended to eliminate gender bias in student evaluations of their instructors. The findings also suggested that administrators should not assume one gender to supply good or bad instruction, and they should not reward instructors on the basis of the instructor's gender but rather on individual course performance.

Relation of Instructional, Pedagogical, and Course Content Characteristics to Student Ratings of University Instructors' Teaching Effectiveness

Introduction

How do students view their instructors' teaching in higher education? Literally thousands of studies have discussed this issue, yet the question persists. As there has been great attempts to define good teaching, researchers have looked for differences in student evaluations based on student and instructor gender, type of class size, course, teaching effectiveness, students' perceptions of teacher performance, students' expected grades and students' achievement (Das & Das, 2001; Wigington, Tollefson & Rodriguez, 1989; Koon & Murray, 2002; Kember, Jenkins & Chi Ng, 2004; Millea & Grimes, 2002; Gallaher, 2004). Researchers have also examined students' evaluations of teaching in terms of instructor and student characteristics, with inconsistent results (Nasser & Fresco, 2006). The inconclusive nature of studies examining student gender, instructor gender, and student level, along with testing these characteristics individually, made this research focus on possible relationships among them. Drawing from Kulik's (2001), Marsh's (1987), Mckeachie's (1997) and Theall's & Franklin's (2001) findings that students' evaluations of teaching are reliable and valid measures of good teaching and based on the work of Young, Rush and Shaw (2009), this study examined interactions among instructor, course and student characteristics, and particularly emphasized the importance of instructor and student gender as well as student level and class size.

Background of the Study

Many higher education systems worldwide have expanded largely in the last decades, and have undergone wide and deep structural changes (Guri-Rosenblit Sebková & Teichler, 2007). The enormous increase in the number of students attending institutions of higher education in Lebanon was achieved due to the so-called "massification of higher education". According to Trow (1972), as reported by Davidovitch and Soen (2006), this massification was facilitated by the ideology of equality of educational opportunities. Consequently, the higher education system was transformed from a privilege for a few, to the right for all (Davidovitch & Soen, 2006). The massification of higher education is changing the cultural context of the institutes partaking in the system. It is now widely known that effective communication in the classroom is a crucial pre-requisite to the achievement of teaching and learning outcomes in higher education. Effective communication strategies in the classroom are important and student-teacher feedback is essential in fostering such an effective communication (Davidovitch & Soen, 2006).

Upon studying the history of higher education in Lebanon, the establishment of institutions of higher education can be divided into three time periods:

- From the beginning until 1950: The Foundation Period, when foreign missionaries established the two universities, American University of Beirut (AUB) (1866) and Universite Saint – Joseph de Beirut (USJ) (1875).
- From 1950 until the outbreak of the War in 1975: The Nationalization Period, dominated by the establishment of Lebanese University (LU) and Beirut Arab University (BAU).

- From 1975 until the present: The Civil War Period, and afterwards. During this long progression, it was normal that a number of forces interlocked, some in harmony and others in conflict (Higher Education in Lebanon, 2009).

According to statistical data in 1994-1995, there were about 6000 university teachers distributed almost equally between Lebanese University and the private institutions. More than 80% of them taught in the Greater Beirut area. As for the degrees, 77% of Lebanese University teachers and 48% of those at private institutions hold high degrees (Ph.D., specialization degrees in medicine...etc). 65% of the teachers are in theoretical studies (arts, humanities, and pure sciences) (Nizam, 2009). More than 140000 students are enrolled in Lebanese higher education institutions. Half of these students are in the Lebanese University, and the other half are distributed among 41 private higher education institutions. The rapid expansion of the student body in Lebanon as reported by UNESCO and World Bank statistics (2008) happened mainly in tertiary education. The percentage of the gross enrollment ratio increased from 28% to 48% (Male 44.5% and Female 51.4%) (Higher Education in Lebanon – A Brief Overview, 2009). According to the World Bank database, gross enrollment rate of tertiary education is 46.9 percent for male, 56.3 percent for female, and 51.6 percent for total (The World Bank 2009). Private enrollment share of tertiary education is 53.4 percent (Higher Education, 2009).

The huge expansion lately has been achieved by liberalizing the admittance standards in some higher education institutions. Some colleges admitted students who were being rejected by the so called "elite" universities. Other institutions decreased the fee per credit to allow the admittance of students from different socioeconomic status. This trend resulted in having more students per class and greater responsibility on the part of lecturers. These developments provided the reason to obtain feedback from students on the teaching practices of their instructors, the quality and clarity

of instruction and the nature of the relationship established by instructors with their students. The instructors are the providers of academic products and the students are the consumers, hence it is reasonable to ask the students to evaluate the products they receive.

Students' ratings of instruction have gained widespread use in colleges and universities and have been the focus and attention of much research and discussion on teaching effectiveness (Beran and Violato, 2005). Alean-Kirkpatrick et al. (1997) stated that there has been increased interest in monitoring quality in higher education. Many types of council such as The Higher Education Funding Council and The Higher Education Quality Council in the UK were created in 1992 to assess and audit the quality of higher education provision. It is now expected that these institutions create new methods to obtain students' views. For instance, the Swiss Federal Institute of Technology has urged the use of students' ratings through a standard questionnaire to assess didactic quality of teaching. Fee – paying students will constantly behave like consumers. Having a dialogue with students about what works is a method to improve the quality of teaching. Hence, students' opinions and their feedback about instructors are of vital importance (Narasimhan, 2001).

Student evaluation of college instruction has been a controversial issue. Some faculty members view students as convenient raters (e.g., Spooren & Mortelmans, 2006) whereas others are convinced that they are unqualified to provide valid assessment of teaching quality (e.g., Anastazi & Urbina, 1997). Despite these differing views, colleges and universities still collect data about the quality of teaching through questionnaires administered to students. The data that these surveys convey have four main purposes. Formative evaluation is used as feedback for instructional improvement, summative evaluation as administrative decisions for tenure and promotion, as an aid to students in course selections and for research in teaching for higher education (Nasser and Fresco, 2006).

As schools and universities work to improve the student achievement, it is important that high quality teachers and teaching be identified. Medley and Cocker (1987) report that previous studies of teacher evaluation have shown that typical evaluations are seen by both principals and teachers of little value. However, recent research has suggested that those evaluations that are specific and use multiple sources of data could lead to substantial improvements in the validity of teacher evaluation (Gallagher, 2004).

According to the Office of Institutional Research (2009), ratings of students, as stated by Anastazi & Urbina (1997), are inherently subject to two weaknesses; the "error of central tendency" and the "halo effect". They both tend to minimize discrimination among individuals and represent restrained estimates of effect. The error of central tendency results because most people tend to avoid the extremes in rating, so ratings tend to accumulate in the center of the scale. The "halo effect" refers to the tendency of the raters to be excessively influenced by a favorable or unfavorable general opinion of the person being rated and then lead to that opinion to tint all specific ratings. This causes raters to make less differentiation between the strengths and weaknesses of instructors or courses than reasonable.

Students' ratings of their instructor's techniques and their own learning have an acceptable validity. However, according to Cashin (1989), as reported in the Office of Institutional Research (2010), students are not qualified to judge a number of factors that characterize ideal instruction:

1. The appropriateness of the instructor's objectives.
2. The relevance of assignments or readings.
3. The degree to which subject matter and content was balanced and up-to-date.
4. The degree to which grading standards were unduly lax or severe.

Although these issues can form basic components of evaluation of teaching effectiveness, they may require alternatives to student ratings to address them. According to Cohen (1980), student ratings are valuable indicators of teaching effectiveness. They provide constructive information to help guide the improvement efforts of instructors, departments, and institutions. However, the highest increases in teaching effectiveness were found when instructors received not only feedback from student ratings, but a combination of ratings consultations and feedback (comprehensive program including a variety of evaluation tools and systematic faculty development) (Office of Institutional Research (2009).

Student evaluation of teacher performance has been an important but controversial tool in assessing the quality of teaching. Kulik (2001) states that the initial aim of student evaluations of teaching served two goals: mapping the quality of teaching and providing information and help instructors improve their teaching. Marsh (1987) and Mckeachie (1997) report that student ratings also used administrative decision making, helping and informing students for course selection, curriculum development, external quality care and research on teaching. Although the implementation of students' ratings of teaching was applauded in many faculties, not everyone is convinced of the reliability and utility of these evaluations (Spooren and Mortelmans, 2006).

Content validity of student ratings reflects how well the instrument's items represent a definite conceptualization of student ratings. Since the instructors themselves are the developers of the various student rating items, the items are considered to be good indicators of teaching effectiveness. In several validity studies performed by Murray (1984) and McKeachie (1979), a consensus has emerged that students' ratings are highly associated with supervisor, colleague, alumni and observer ratings. As reported by Murray (1984), Cohen's (1981) review of 41 such studies revealed that student achievement is related to student rating. That students learn more in

courses with better instruction, supplies the highest evidence of the validity of student ratings (Beran & Vilato, 2005). On the whole, there is an agreement that student evaluations provide consistent feedback of general areas of an instructor's strengths and weaknesses.

Despite this evidence of validity, researchers have argued that student ratings are biased by characteristics of the students conducting the evaluations, the instructors, and the courses being evaluated. Since student ratings have gained considerable weight in faculty merit, tenure and promotion decisions, it is important to determine the relationship between student ratings and characteristics of students, instructors and courses (Beran et al., 2005).

Statement of the Problem

Student Characteristics and Student Evaluation of Instructors

Several characteristics about students have been examined in relation to their evaluations of teaching effectiveness. One of the most studied student factors is class attendance. Some researchers suggested that students who attend class frequently tend to be more motivated and interested in the course than those who are less attendees, resulting in higher ratings. However, this assumption did not yield significant differences or interaction effects between students' sex (Davidovitch & Soen, 2006).

Additional student characteristics may be associated with student ratings. One of the most studied factors is the expected grade. Greenwald (2002), as reported by Beran and Vilato (2005), in a review of research on student ratings concluded that instructors who give high marks are likely to receive positive student ratings.

March and Rosh (2000), as stated by Beran et al. (2005) reported that a high workload is associated with higher student ratings. If effective teachers entail high workloads, then students who are challenged by these instructors may become more committed, hard working and learn more, hence, provide high ratings. Courses may be rated differently according to whether the student is registered in the program in which the course is offered. Theal and Franklin (2001), as noted by Beran and Vilato (2005), suggested that courses within the student program tend to be given low ratings (Beran & Vilato, 2005).

The relationship between student ratings and gender of students themselves has gained a great of importance. In a large study by Hancock, Shannon, and Trentham (1992), as reported by Young et al. (2009), the researchers found that female students rated their teachers higher than male students on most aspects of teaching effectiveness. In their study on "Student Evaluations of College Professors: Are Female and Male Professors Rated Differently?", Basow and Silberg (1987) found that male students gave female professors significantly less positive ratings than they gave male professors on all dependent measures. Male students also rated female professors significantly more negatively than did female students on Scholarship, Organization/Clarity, Dynamism/Enthusiasm, and overall teaching ability. However, they also found that female students rated female professors significantly more negatively than they rated male professors on Instructor-Individual Student Interaction, Dynamism/Enthusiasm, and overall teaching ability (Basow & Silberg, 1987). Other researchers (Summers, Anderson, Hines, Gelder, and Dean, 1996) studied undergraduate and graduate students' perceptions of course satisfaction in traditional courses. The results revealed that male students rated female instructors lower than did female students (Young et al., 2009).

In a review of the research on the link between student age and student level with evaluations of teaching effectiveness, Basow and Silberg (1987), in their study on undergraduate students, reported that the higher the student level, the higher was the teacher rating (Basow & Silberg, 1987). As noted by Young et al. (2009), Donaldson, Flannery, and Ross-Gordon (1993), reported comparative results from three different studies on graduate students. They found that adult graduate students identified different traits of teacher effectiveness than the undergraduate students. They found that graduate students were more likely than undergraduate students to mention instructor characteristics such as providing motivation and using a variety of teaching techniques. They also found differences in age group expectations. Older students were more concerned about relationship matters such as teachers who are dedicated and motivate their students whereas younger students were highly interested in traits that might enhance their own work such as being successful.

Crombie, Pyke, Silverthorn, Jones and Piccinin (2003) explored student perceptions of certain features of the academic environment. In particular, they studied the influence of several contextual aspects of the university classroom (i.e., class size and gender balance, discipline, instructor gender) on student perceptions of their own participation and of their instructor, as well as the influence of three individual factors: student gender, student self-perceived general activity level, and student age. Students categorized as active class participants considered themselves to raise their hands more frequently, interrupt more frequently, and interfere for longer periods of time as compared to those categorized as less active. Crombie et al. (2003) found that students who perceived themselves to be less active did not differ in their specific behavior patterns as a function of gender or discipline, whereas the male active students interrupted more, and those categorized as active who were in Arts/Social Science disciplines reported long and/or more frequent

interactions with instructors. Their findings revealed that active participators regarded their professors as more positive, as more personalizing, as stimulating more discussion, and they had a more positive impression of their professors overall than did students who perceived themselves as less active (Crombie et al., 2003).

Instructor Characteristics and Student Evaluation of Instructors

In investigating the influence of instructor's gender on student evaluations of college instructors, Young et al. (2009) found that students rated instructors of the same sex higher than instructors of the opposite sex (Young et al, 2009). Other researchers found that male faculty members were more likely to be chosen as best instructors more than females (Das & Das, 2001). Still other researchers, when studying the effect of gender on student ratings of their professor on items such as enthusiasm and acceptance, found that students rated male professors higher than female professors (Arbuckle et al., 2003). Smith, Yoo, Farr, Salmon, and Miller (2007) found that male and female students rated female instructors higher than male instructors (Smith et al., 2007). In relation to student ratings, instructor gender and other characteristics such as age, experience and academic rank have been also investigated extensively. For instance, in examining the influence of instructor gender on student evaluations, Basow and Silberg (1987) and Sandler (1991) found that female instructors are rated lower than male colleagues. Other researchers such as Basow and Distenfeld (1985), Feldman (1983) Goodwin and Stevens (1993) and Hancock, Shannon, and Trentham (1992) were unable to find evidence of gender differences. And still others, such as Feldman (2007), Bashen, McLouglin, and Garcia (1999) and Tatro (1995) found that college students rated female instructors higher than male instructors.

The setting in which the student evaluations take place may be important. Feldman (1993), as stated by Young et al. (2009) found that very little gender bias was found in classrooms in which extraneous variables (teacher expressiveness, student confusion of teaching style) were tightly controlled whereas a slight bias in favor of same gender preference took place in studies carried out in classrooms without such controls. Areola (2000), as reported by Young at al. (2009), in a summary of studies on gender bias found that the apparent bias may be due to courses that instructors are assigned to teach rather than instructor's gender (Young at al., 2009). Hence, from the aforementioned research on the relation of student characteristics and instructor characteristics to student evaluation of instructors, it was hypothesized that (1) male students rate male instructors higher than they rate their female instructors and (2) female students rate female instructors higher than they rate their male instructors.

Course Content Characteristics and Student Evaluation of Instructors

Arreola (1995), as reported by Beran and Vilato (2005), stated that student ratings are lower for required courses than for elective courses. Franklin (2001) indicated that physical sciences faculties tend to be given low ratings. It is possible, that 4-5 months in comparison to 1-2 months yield higher ratings as the longer duration allows for more student- teacher rapport to develop. The type of the course may also be an important factor. That is, lab or practicum courses that allow student experiences in applying newly learned knowledge and skills may lead to a sense of mastery which, in turn, may lead to feelings of satisfaction of the instructor.

Several additional characteristics may be related to student ratings. Courses offered in the fall and winter terms may receive lower ratings as students feel pressed to work under heavy course

load in comparison to courses offered in the spring and summer terms when students tend to enroll in fewer courses and have more holidays. The year in which the student takes the course may also be related to student ratings. When student ratings are first introduced to students as a means to formally evaluating their instructors, students may feel enthusiastic to express their feedback. As these rating instruments are continually administered, students may become fatigued and burdened by the repetition. Hence, they may provide lower ratings after several years of administering evaluation instruments (Beran and Vilato, 2005).

According to Feldman (1984), as reported by Young et al. (2009), students in large classes were more likely to rate teachers lower that students in small classes. He also found that teachers giving upper level courses tended to have higher ratings than those teaching lower level courses. In addition, Marsh and Baiely (1993) found that graduate level courses were given higher rates by student than undergraduate level courses. Therefore, it was hypothesized that graduate students rate instructors higher than undergraduate students and that students in smaller classes rate teachers higher than students in larger classes.

In summary, various student, instructor and course content characteristics might be related to student ratings of their instructors. The aim of the present study was to further explore the relative importance of these characteristics on Lebanese society and to test the existence of any cross cultural differences. We expected to find that students would rate same sex instructors higher than opposite sex instructors. We also expected that student level as well as class size would interact with instructor gender. More specifically, the following hypotheses were examined:

- **Hypothesis 1:** Male students rate male instructors higher than they rate their female instructors.
- **Hypothesis 2:** Female students rate female instructors higher than they rate their male instructors.
- **Hypothesis 3:** Graduate students rate instructors higher than undergraduate students.
- **Hypothesis 4:** Students in smaller classes rate instructors higher than students in larger classes.

Professional Significance of the Study

The purpose of the present study to examine students' evaluations of teaching based on student characteristics, instructor characteristics and course content characteristics. Much research has been performed to discuss the effect of student evaluation on instructors. Previous studies, as shown by literature review, examined these characteristics separately. This survey serves to study these factors in a combined way. It is expected that male and female students differ in their ratings of instructors, depending on instructor gender. Since we would be able to study interactions between student characteristics, instructor characteristics and course characteristics, it was hypothesized that these interactions would help further our understanding of student evaluation of their instructors and clarify previous findings. Hence, the results obtained in this study could have several purposes. First, they could be used to gather preliminary data about the effect of student characteristics, instructor characteristics and course characteristics on the students' ratings of their college instructor. Second, the findings could provide some data about how Lebanese students perceive their instructors' teaching. Third, they could be a foundation for developing university programs to help instructors become more aware of their teaching skills and enlighten them about the students' perceptions regarding their teaching habits. Fourth, the results could be helpful in motivating teachers to adopt new methods of instructional tools to enhance their teaching and the

students' educational opportunities. Finally, the findings could be useful to encourage further study in this aspect of education and pave the way for more research in this field.

Moreover, there have been no studies conducted in Lebanon on the topic of the relationship between student characteristics, course characteristics, instructor characteristics and students' evaluation of college instructors. This research is a pioneering research to gather foundational data on the relationship among the above mentioned factors. The main emphasis of this study was to examine students' evaluations of teaching based on certain student and instructor characteristics. We expected that male and female differ in their ratings of instructors, depending on instructor gender. Also, we expected to find differences between graduate and undergraduate students' ratings. Previous research, as shown by review of literature, has revealed the relationships between these factors without looking at interactions. Since we would be able to study interactions of student gender, instructor gender, student level and class size, it was hypothesized that these interactions would help further our understanding of any gender bias and would clarify previous research on the matter. Finally, since this was a research based on the work of Young et al., (2009) and since the Teacher Evaluation Scale (TES) the researchers used was based on content validity of previous research, it was expected that the results of the current study further validate the scale previously used. We expected to replicate the previous research on a Lebanese sample and test for significance of cross cultural differences in the findings that might exist. We also expected that, if American male students rated their male instructors higher than their female instructors, then Lebanese students (male or female) were, for sure, expected to rate their male instructors higher than female instructors since the Lebanese society tends to be a patriarchal one.

Overview of Methodology

Undergraduate and graduate students (n = 159) from five private universities in Lebanon, were asked to participate in this study. The sample was a convenient sample. The college instructors agreed to give permission to students to answer the questionnaire in 3-5 minutes mainly by the end of the class session. The instrument, the Teacher Evaluation Scale (TES) consisted of 25 items, which the students were asked to complete in addition to a number of demographic variables as well as data about student's gender, age, university level and major, instructor gender, course characteristics and class size. The nature of the study is quantitative and it mainly relied on the ANOVA and t-test for data analysis.

Limitations of the Study

A main limitation of this study was the fact that it relied completely on students' answers. Data collected to measure how the students evaluate their teachers was based completely on the students' responses and their own points of view. Regarding students' age group, experience, and environmental and circumstantial pressure, responses may be inaccurate and thus held untrue. Hence, a margin of error was expected in the results.

Another main limitation of the study was the fact of private verses public universities as well as demographics. The research was performed in only five universities in Lebanon. It did not include enough data from the public and private sectors. Data was limited to specific areas and it was not gathered from different areas in Lebanon such as the South. Hence, due to cost and time constraints the sample may not be as representative since it was only made up of 159 students and

a limited number of professors being evaluated. Therefore, findings could not be generalized to Lebanese population.

A third limitation was that which incorporate timing. The survey was conducted in the early spring time; the students most probably rated an instructor from the fall semester. Hence, the students' evaluations might not reflect their exact responses since a period of time had elapsed and this might have distorted the perceptions of their instructors. Thus, the margin of forgetfulness might lead to different results.

People often tend to remember extreme situations: be it bad or good. The idea that students were asked to rate an instructor within the context of a particular course could probably lead to students choosing to rate the professor of a course in which they received high grades compared to other courses, or to rate an instructor in which they had bad experience.

Students might rate an instructor they had for several courses. This might fall in the range of having predetermined ideas that are difficult to change and this in turn, might create an overlap of information where students might think the instructor acted in a certain way they were accustomed to where in fact he/she might have acted differently.

The number of male and female students as well as the number of male and female instructors might affect the results of the study. The researcher could not include participants from both genders equally.

Definition of Terminology

The following variables are defined in the way they have been used for the purpose of this study:

- **Student Evaluation of Instructor**: Student ratings are regularly used for summative faculty evaluation in colleges and universities. Previous Research has formed several rating instruments that researchers agree hold adequate levels of reliability and validity. It is agreed that rating instruments should 1) be constructed by measurement experts, 2) contain a limited number of high inference items, 3) be administered using standardized procedures, and 4) be subjected to systematic research within institutions. In effect, it has been strongly agreed that topics for discussion at current evaluation and development conferences tend to focus on alternative measures for the total evaluation process; i.e., peer evaluations, self- evaluation, and measures of student learning (Harris, 1982).

- **Student Characteristics**: Dunn and Dunn (1992), among those who conceptualized students' characteristics, define it in terms of learning styles as the way "each learner begins to concentrate on, process and remember new and difficult information" (p. 2). According to Gremli (1996), "an individual's learning style is the way that person begins to process, internalize and concentrate on new material" (p. 24) (Intime, 2001). Student differences with regard to gender may contribute a great deal to the significance that students place on certain aspects of effective teaching. McKeachie (1990), as reported by Young et al. (2009), suggested that effective teaching is dependent on the characteristics of the students themselves, as well as on the teacher's behavior. Researchers have also examined the relationship of student age and student level with evaluations of teaching effectiveness. Data about the student gender, age, level of study at university (undergraduate or graduate) and major were used for the current study.

- **Course Content Characteristics**: Classroom or course characteristics such as class size, course discipline, course level, or whether a course was required or an elective have been found to relate to students' evaluations of teachers (Young et al., 2009). With respect to course characteristics, two variables were examined by Nasser and Fresco (2006): course discipline and course length. Data about course type (major or elective) of any particular course were used in the current study.

- **Instructor Characteristics**: In relation to student ratings, instructor characteristics include gender and other characteristics such as age, experience, and academic rank (Young et al., 2009). In addition to instructors' predictions of student ratings, three variables related instructors' attributes were investigated: academic status, teaching experience, and academic degree. Some research has examined the relationship between student ratings and instructors' academic standing, level specialization, and research productivity. Centra (1993) revealed some evidence that very inexperienced instructors receive lower student ratings and that the accuracy of self-ratings seems to be related to experience (Nasser & Fresco, 2006).

- **Teaching Effectiveness:** According to the Office of Institutional Research (2009), researchers are in agreement that teaching effectiveness is a multidimensional construct. Both Centra (1993) and Braskamp and Ory (1994) identified six factors commonly found in student rating forms: (1) Course organization and planning (2) Clarity, communications skills (3) Teacher student interaction, rapport (4) Course difficulty, workload (5) Grading and examinations (6) Student self-rated learning. To insure that instructors receive ratings on all of the appropriate dimensions, forms must differentiate among the various items and their scopes (Office of Institutional Research, 2009).

- **Class size**: Class size is defined as the number of students enrolled in the class as of the third week of the quarter (Bedard & Kuhn, 2005). In contrast to the literature examining class size effects on test-based outcomes, Bedard and Kuhn (2005) found a large, highly significant, and nonlinear negative impact of class size on student evaluations of instructor effectiveness that was highly vigorous to the inclusion of course and instructor fixed effects (Bedard & Kuhn, 2005).

Chapter 2

Review of Literature

Student evaluations of instruction are extensively used as a basis for personnel decisions and faculty development recommendations in post-secondary education today. Student ratings add a valuable component to the range of input for the evaluation of teachers (Scriven, 1995).

Darling-Hammond and others (1983), as reported by Barett (1986), define teacher evaluation as "collecting and using information to judge" (p. 1). They consider two types of evaluation: formative and summative. Formative evaluation is a tool used to improve instruction. Summative evaluation is a tool used to make personnel decisions. Both evaluation uses have gained much attention in recent literature as the teaching profession considers evaluation an integral part of staff development and the administration looks to evaluation data as evidence in accountability debates (Barett, 1986).

Braskamp and Ory (1994), as reported by Ahmadi & Cotton (1998), identify six factors commonly measured by student rating forms: (1) communication skills (2) rapport with students (3) course organization (4) student self-rated accomplishments (5) course difficulty and (6) grading and examinations. Marsh (1994) included nine factors in the Student Evaluation of Educational quality form: learning/ value, enthusiasm, organization, group interaction, individual rapport, breadth of coverage, exam/grades, assignments and workload (Ahmadi et al., 1998).

Student evaluation of teacher performance has been an important but controversial tool in assessing the quality of teaching. Kulik (2001) states that the initial aim of student evaluations of teaching served two goals: mapping the quality of teaching and providing information and helping instructors improve their teaching. Marsh (1987) and Mckeachie (1997) report that student ratings

are also used for administrative decision making, helping and informing students for course selection, curriculum development, external quality care and research on teaching. Although the implementation of students' ratings of teaching was applauded in many faculties, not everyone is convinced of the reliability and utility of these evaluations (Spooren and Mortelmans, 2006).

Validity of Students' Ratings

Scriven (1995) discusses nine potential sources of validity for student ratings of instruction.

- The positive and statistically significant correlation of student ratings with learning gains.

- The unique position and qualifications of the students in rating their own increased knowledge and comprehension.

- The unique position of the students in rating changed motivation (a) toward the subject taught; perhaps also (b) toward a career associated with that subject; and perhaps also (c) with respect to a changed general attitude toward further learning in the subject area, or more generally.

- The unique position of the students in rating observable matters of fact relevant to competent teaching, such as the punctuality of the instructor and the legibility of writing on the board.

- The unique position of the students in identifying the regular presence of teaching style indicators. Is the teacher enthusiastic; does he or she ask many questions, encourage questions from students, etc.?

- Relatedly, students are in a good position to judge--although it is not quite a matter of simple observation--such matters as whether tests covered all the material of the course.

- Students as consumers are likely to be able to report quite reliably to their peers on such matters of interest to them as the cost of the texts, the extent to which attendance is taken and weighted, and whether a great deal of homework is required--considerations that have little or no known bearing on the quality of instruction.

- Student ratings represent participation in a process often represented as "democratic decision making." The "best available alternative" line of argument (Scriven, 1995).

Supporters of the validity of students' evaluations argue that it is logical that students who indeed enjoy learning teaching and instruction be part of the quality care. Theall and Franklin (2001) state that no one other than the student is "as qualified to report what transpired during the term simply because no one else is present for as much of the term" (p. 48). Opponents consider the ratings of the students as "meaningless qualification" and have doubts regarding the validity of the student perceptions of teaching. However, extensive research by Penny (2003) and Marsh (1987) has been done to prove that the results of students' ratings supply evaluators with valid, reliable and valuable data concerning the effectiveness of teaching and that student evaluation is the only indicator for teaching effectiveness (Spooren and Mortelmans, 2006).

Different researchers reported several variables that effect students' opinions when evaluating teachers. Each variable is discussed in relation to student ratings of instructors.

Relationship between Student Ratings of Instructors and Teaching Effectiveness

Different results were conveyed on the relationship between student evaluation of teachers and teacher effectiveness. For instance, when using multiple outcomes to validate student ratings

of overall teacher effectiveness, Koon and Murray (2002) found a low correlation ($r = 0.41$) between final examination scores and mean student ratings of overall teaching effectiveness, another low correlation ($r = 0.21$) between section means for advanced course registrations and final examination scores, and a correlation lower than expected ($r = 0.43$) between mean ratings of amount learned an mean ratings of overall instructor effectiveness (Koon et al., 2002).

Obenchain, Abernathy, and Wiest (2002), also examined the reliability of students' ratings of faculty teaching effectiveness. Whereas previous studies looked for reliability across a group of students and over time, the comparison in this study examined the reliability of the individual student. However, this study found that individuals were not consistent in their evaluations and considered collective reliability measures to be giving faculty a false sense of security. They also found a significant difference between teacher education students' evaluations of teacher education courses and non-teacher education courses. Teacher education students rated teacher education faculty higher than non-teacher education faculty (Obenchain et al., 2002).

In their profile analysis on " Multidimensional Students' Evaluations of Teaching Effectiveness (SETE)" Marsh and Bailey's (1993) primary purpose was to determine whether or not there are individual differences in the SETE profiles that generalize across rating of the same instructor collected over a thirteen-year period. They studied systematic differences in SETE scores using three aspects: (1) Level (same instructor) (2) Parallelism (different instructors) and (3) Flatness (all instructors). They found the following: - Different ratings of the same instructor over a 13-year period of time. - Some instructors got systematically higher ratings in graduate-level courses, whereas other instructors got systematically larger ratings in undergraduate courses. - A very good consistency in the profiles obtained by the same instructor on different course offerings over a 13-year period. - Differences between instructors varied substantially depending upon the

specific Students' Evaluation of Educational Quality (SEEQ) scale that was being considered (Marsh et al., 1993).

Mintzes (1979) examined relationships between overt classroom teaching behaviors of college instructors and student ratings of teaching effectiveness. Relationships between teaching behaviors and student ratings were examined by simple, multiple, and canonical correlation procedures. Canonical correlation analysis showed a significant relationship between batteries of teaching behaviors and measures of teaching effectiveness. The highest multiple correlations were associated with the Clarity and Rapport items (Mintzes, 1979).

Heckert, Latier, Ringwald and Silvey (2006) investigated the relation of course, instructor, and student characteristics to student ratings of teaching effectiveness, both overall and within the dimensions of pedagogical skill, rapport with students, difficulty appropriateness, and course value/learning. They found that interest in the course content, expected grades, satisfaction with the time of day, and instructor sex were related significantly to all dimension of teaching performance. They also found that year in school and reason for taking a course were related to value ratings, and student sex and year in school were related to rapport ratings. Finally, their results suggested that background characteristics (course, instructor and student) were related most strongly to course value ratings and the overall evaluation, and least strongly to the pedagogical skill ratings (Heckert et al., 2006).

Spooren and Mortelmans (2006) also presented and discussed the results of research on student evaluation of teacher effectiveness. One of their main concerns was to study the correlation between grading and student evaluation of teaching and tracing other factors that more or less influence this relationship. Confirmatory factor analysis revealed the existence of an underlying

factor which they called "teacher professionalism" that influenced the student ratings of teachers. The results also revealed that there indeed exists a strong effect of students' grades in a course on student ratings of that course. The findings also showed significant effects of class size and course attendance on final grades of the course and of class size on overall grades. This suggested that students who usually attend the lectures that are given in smaller classes might have better learning outcomes (Spooren & Mortelmans, 2006).

Relationship between Student Ratings of Instructors and Students' Perceptions of Teacher Performance

Students' perceptions of teacher performance have also gained much interest in research. Students in part time courses in the Open University of Hong Kong were interviewed about their perceptions of good teaching and tutoring. The results revealed differing perceptions between those with reproductive conceptions of learning and students holding self determining ones. The former preferred didactic teaching but disliked interaction whereas the latter found that student-centered approaches more consistent with their perspectives of learning. The findings have implications for the evaluation of teaching. Students faced with teaching incompatible with their perceptions gave lower ratings of their teachers. With respect to the evaluation of teaching, the findings reveal an undermining of the validity of student ratings particularly on compulsory standard questionnaire (Kember, Jenkins & Chi Ng, 2004).

In her study that aimed at exploring faculty and students perspectives on students' evaluations, as well as identifying the usefulness and appropriateness of the ratings for evaluating teacher effectiveness in the American University of Beirut, El Hassan (2009) found that students

and faculty believe in the effectiveness and usefulness of the Instructor Course Evaluation (ICE) system with the need to overcome some negative consequences inherent in the university. She also concluded that students positively viewed the evaluations, the uses made of them, their consequences as well as the rating process. However, she found that faculty members showed lack of agreement in their responses on some items but overall, they were not negative regarding the student evaluation system. The faculty cited in their comments some "biases" in the system and recommended enhancement of the validity of the system (El Hassan, 2009).

Sojka, Gupta, and Deeter-Schmelz (2002) also examined students' perceptions of student evaluation of teaching and compared them to faculty perceptions on several dimensions (faculty, gender, rank, and status). The results indicated differences between faculty and student perceptions compared to student evaluations. Faculty were more likely to agree that students award easier, more entertaining instructors higher ratings, and that students do not take SET seriously. Students, on the other hand, were less likely to agree that student evaluation encourage faculty to grade more leniently, have an influence on faculty members' careers, or lead to changes in courses and/or teaching styles (Sojka et al., 2002).

On his study on student perceptions of teaching evaluations, Brown (2008) examined how students perceive official student evaluations of teaching (SET) and unofficial mid-semester evaluations (MSE). Results showed that students believed SETs are valid measures of teaching; however, they had doubts about whether students or instructors took these evaluations seriously. The students had very positive perceptions of MSEs and instructors who conduct them. In addition, completing a MSE positively affected perceptions of the instructor's responsibility, his commitment to teaching and his desire for the class to do well (Brown, 2008).

In their study on students' perceptions of their classroom participation and instructor as a function of gender and context, Crombie et al. (2003) explored student perceptions of certain features of the academic environment. In particular, they examined the influence of various contextual aspects of the university classroom (i.e., class size and gender balance, discipline, instructor gender) on student perceptions of their own participation and of their instructor, as well as the influence of three individual factors: student gender, student self-perceived general activity level, and student age. The findings suggested that students categorized as active class participants perceived themselves to raise their hands more frequently, interrupt more frequently, and intervene for longer periods of time as compared to those categorized as less active. Moreover, active participators regarded their professors as more positive, as more personalizing, as stimulating more discussion, and they had a more positive impression of their professors overall than did students who perceived themselves as less active. In terms of student gender effects, the results revealed that male students rated their own participation as significantly higher than did female students from the same classes. Males reported significantly higher levels of participation overall, significantly higher levels of interrupting, and significantly more and longer interactions with their instructor than did females (Crombie et al., 2003). Okpala and Ellis (2006) examined the perceptions of college students on teacher quality components, with exclusive focus on teacher qualifications. The results showed that the perceptions of college students on teacher quality components vary and that teaching skills, commitment to student learning, content knowledge and verbal skills are important attributes of a quality teacher (Okpala et al., 2006).

In her study "Traditional and Non-Traditional College Students' Descriptions of the "Ideal" Professor and the "Ideal" Course and Perceived Strengths and Limitations", Strage (2008) found that students have fairly well - bounded ideas of an ideal learning context, and of their own

strengths and limitations. She also found that the "ideal professor" appeared to be something of a cross between a stand-up comic, a serious scholar, and a readily accessible and caring mentor. The "ideal course" appeared to need to be complete with rigorous up to date and relevant content for some, fun and entertaining for others and all of these for yet others. Another interesting finding was that traditional-age students, and those coming directly from high school preferred a college environment that is essentially an extension of high school. Older and more experienced students were more concerned about securing adequate preparation for career and life after college (Strage, 2008).

Miley and Gonsalves (2003) studied students' perceptions of teaching to provide feedback for faculty members about their teaching habits. Data collection consisted of handing out index cards to students on which they were to write anonymously five annoying teaching habits that they noticed in their professors. The results indicated that students perceived disorganized teaching, taslking too fast for students to process information, lecturing in a monotone voice, and degrading students were professors' most annoying habits (Miley, 2003).

Relationship between Student Ratings of Instructors and Grades

To test the relationship between grade expectation and student evaluation of teaching (SET), Millea and Grimes (2002) found that male students submitted lower overall ratings of the course when the instructor of this course was female. The results also suggested that teachers can maintain high expectations in their courses without having a detrimental impact on evaluation scores. Their study went further to examine the impacts of the different components that determine a student's grade expectation (current earned grade and expectation about remaining graded work)

These results indicated that higher earned grades increased overall evaluations while pessimism about future graded work lowered the overall evaluation. Hence, while current earned grades played a positive role in determining SET scores, if students had negative attitudes toward their future exams and assignments, their SET scores tended to fall (Millea et al., 2002).

On a comparative study on different models explaining the relationship between instructor ratings and expected student grades, Wright and Palmer (2006) used four different models to test the different relationships between students and expected student grades and students evaluations of the quality of instruction with and without student motivation, ability, and amount learned as important variables. Statistical tests of the alternative models revealed a more complex model that incorporated student motivation and ability levels as factors affecting student evaluations of their instructors. For this set, the findings indicated that instructor ratings were not simply a function of expected grades, or simply a function of perceived amount learned but a function of motivation, ability, amount learned and grades (Wright & Palmer, 2006).

Dolmans, Luijk, Wolfhagen and Scherpbier (2006) investigated the influence of harsh grading by tutors on tutor performance rating by student. Students were asked to rate tutor performance after receiving their grades for professional behavior. To reflect tutors' harshness of grading students were asked to indicate whether they perceived their grades as too positive, adequate or too negative. Students were asked to rate the quality of feedback they received from the tutors with respect to grades. The results revealed three findings. Professional behavior grades that students perceived as too negative, adequate or too positive were associated with tutor performance ratings. Harshness of grading did not influence tutor performance ratings. Tutor ratings were predicted more effectively by the quality of feedback the tutors provided on grading than by the harshness of grading (Dolmans et al., 2006).

On an experimental study that examines the connections between the fairness of grade distributions, the fairness of grading procedures, and evaluations of the instructor, Tata (1999) found that the fairness of both the grade distributions and the grading procedures influenced students' evaluations of instructors. Fair grading procedures, however, influenced evaluations of the instructor only when the grade distributions were perceived as unfair -- that is, when the grades did not meet expectations. When grade distributions were perceived as fair, there was no significant difference in evaluations of the instructor among participants exposed to fair and unfair procedures (Tata, 1999).

Miley (2009) surveyed undergraduate students taking upper level psychology courses about their course and their grade expectations and concluded the following. Students believed that success in a course was measured by good grades rather than by mastery of new material. They wanted effort to play an important role, yet realized that effort was hard to evaluate. Students stated that they took courses because they were requirements for their discipline or the college, rather than because they were interesting or led to self-improvement. Survey results showed that faculty members were quite aware that students emphasized grades over mastery and new learning in their courses (Miley, 2009).

Addison and Warrington (2006) studied students' perceptions of course difficulty and ratings of their instructors. Their research dealt with the possible relationship between high grades and favorable student evaluation of instruction. They found that students who earned higher grades evaluated their teachers more favorably than did students who earned lower grades. They also found that students who thought the class was easier than expected evaluated the professor more favorably than did students who thought the course was harder than expected, who in turn evaluated the professor more negatively, regardless of grade earned (Addison et al., 2006).

Relationship between Student Ratings of Instructors and Achievement

Gallagher (2004) examined the validity of a performance-based, subject-specific teacher evaluation system by analyzing the relationship between teacher evaluation scores and student achievement. He used hierarchal linear modeling to estimate value added teacher effects which were then correlated with teacher evaluation scores in literacy, mathematics, language arts, and a composite measure of student achievement. Results showed a strong, positive and statistically significant relationship between teacher evaluation scores and student achievement in reading and a composite measure of teacher and student performance and a positive relationship in mathematics. He also used document analyses and interviews with teachers to explore factors affecting the relationship between teacher evaluation scores and student achievement across subjects. The findings suggested that the relationship was stronger in reading than mathematics (Gallagher, 2004).

Diseth (2007) studied the relationship between three factors: student's evaluation of teaching, teaching approaches to learning, and academic achievement. He posted that not only student's evaluation and perception of learning environment are considered to be essential predictors of student's approaches to learning but also examination grades as well. His study investigated the factor structure of an inventory measuring evaluation-perception of the learning environment. The findings suggested a relationship between approaches to learning and academic achievement and between student evaluation of teaching and evaluation-perception. The results also indicated causality in the relationship between these variables, such that evaluation-perception appears to cause approaches to learning (Diseth, 2007).

Yao, Weissinger and Grady (2003) studied "Faculty Use of Student Evaluation Feedback". They examined faculty formative use of end of semester student rating of instruction (SRI) feedback. The underlying assumption of the SRI feedback model was that faculty would use the feedback to improve their teaching. The results supported the assumption and revealed a positive correlation between feedback use and evaluation results and suggested that faculty use of SRI feedback resulted in improvement in teaching (Yao et al., 2003).

Relationship between Student Ratings of Instructors and Gender

To examine the gender bias in ratings of university instructors' teaching effectiveness, Young et al. (2009) invited students in five colleges to rate instructors on three factors: interpersonal characteristics, pedagogical characteristics and course content characteristics. They analyzed group differences based on student gender, instructor gender and student level. The results revealed that student gender and instructor gender played an essential role in how the students viewed good teaching. Student gender interacted with instructor gender for two factors of teaching effectiveness: pedagogical characteristics and course content characteristics. Students rated instructors of the same sex higher than instructors of the opposite sex (Young et al., 2009).

In their study on business students' perceptions of best university professors Das and Das (2001) examined the effect of gender in this matter. The results showed first, male faculty members were more likely to be chosen as best instructors than females. Both male and female students were more likely to choose a male as their best instructor. However, female students were far more likely than male students to choose a female as their best instructor. Second, masculine students were more likely to choose instructors they perceived to be masculine, and feminine

students preferred instructors they perceived to be feminine. Third, female students chose instructors perceived to be low in femininity (i.e., masculine/undifferentiated) as their best instructors more often than those perceived to be high in femininity (i.e., androgynous or feminine). Fourth, there was no significant difference in the perceived gender roles of the male and female best instructors chosen by female students (Das & Das, 2001).

In their research on "Students' perceptions of expressiveness: age and gender effects on teacher evaluations", Arbuckle and Williams' (2003) purpose of the study was to determine whether the perceptions of college professors' expressiveness in the classroom are strongly associated with students' implicit attitudes toward age and gender. Their first hypothesis predicted that the professor gender would influence student ratings of the professors on the teacher-expressive items (enthusiasm, conscientiousness, acceptance, desirable voice tone, interest, and confidence). The results revealed that students rated male professors higher than female professors. Their second hypothesis predicted that students would rate male and female professors under the age of 35 higher than male and female professors over the age of 55. The results showed that students rated a young professor higher on these items only when the professor was a man (Arbuckle et al., 2003).

To study the Influence of student sex and instructor sex on student ratings of instructors, Smith et al. (2007) posed research questions as to whether male and female students would rate male or female instructors more highly on five dimensions of student rating forms (Instructor Involvement, Student Interest, Student-Instructor Interaction, Course Demands, and Course Organization). Results indicated that male and female students rated female instructors more highly on all five dimensions (Smith et al., 2007).

Relationship between Student Ratings of Instructors and Class (Type, Size and Level)

Class type is usually defined by such characteristics as lecture, combined lecture discussion, and laboratory. Wigington, Tollefson and Rodriguez (1989), as reported by Simmons (1996), showed that instructors with discussion classes had higher overall ratings than other class types and smaller classes showed a higher rating than larger within a class type. They found no differences among class types in comparisons of class rank (lower, upper or graduate). But they did show that in some types the more experienced instructors got the lower ratings and the least experienced got the lowest in others – instructor rank is apparently interactive with class type. The Wigington et al. study (1989) on the interaction of class type and instructor gender showed that women got better ratings in lecture, discussions and laboratory classes but received lower scores in lecture/discussion classes.

Using a demarcation of small as 25 or less, midsized as 26-49, and large as any class over 50, as reported by Simmons (1996), Wigington et al. (1989) found that the interaction of instructor gender and class size produces better ratings for women in small classes and men in large classes. The interaction of class level (lower, upper or graduate division) and size showed higher ratings for instructors in upper division, large classes in comparison with mid-sized classes.

As stated by Simmons (1996), Romeo and Weber (1985) and others found that there were no difference for ratings of instructors of higher or lower classes. Cranton and Smith (1986) and others found that instructors of upper level classes received higher ratings. Wigington et al. (1989) found that while instructors of lower level classes received lower ratings than those of higher level classes, there was no significant difference between men and women in lower levels. However, in upper division or graduate courses men got higher ratings than women (Simmons, 1996).

Relationship between Student Ratings of Instructors and Course Content Characteristics

On their study examining the relationship between student ratings and instructors' predictions of these ratings, taking into account other instructor, student and course characteristics, Nasser et al. (2006) found a systematic positive relationship between instructors' predictions and actual student ratings with respect to overall ratings and the ratings of three dimensions of teaching. Results also demonstrated that low-rated instructors tend to overestimate their student ratings, high rated instructors underestimated student ratings and moderately rated instructors gave accurate predictions.

The characteristics of the three dimensions were as follows:

a) Instructor – rating predictions, status, experience and academic degree

b) Student – interest in subject and expected grade

c) Course – course discipline and course length

The findings showed that the instructor predictions of student ratings were consistent with the actual student ratings regardless of overall course rating or specific dimensions of rating (open–mindedness, organization and stimulation). Prior interest in the course was prominent with respect to overall student ratings, ratings in the area of open-mindedness and ratings in the area of stimulation whereas instructors' academic degree was prominent with the area of stimulation.

With respect to student variables, findings indicated that instructors were rated highly when students expected high grades and when they were interested in course content characteristics. Course variables (discipline and length) and other instructor variables (status and experience) did not have any significant weights (Nasser et al., 2006).

Relationship between Student Ratings of Instructors and Instructor Rank and Reputation

Instructor rank can be differentiated as graduate teaching assistant (GTA) and various levels of professor (graduate, doctorate). Research by Wigington et al. (1989), as stated by Simmons (1996), into the interaction of class size and instructor rank showed that some ranks were greatest at the small class size, decreased in mid class size and somewhat greater in large class size – a 'U' shaped profile. In other cases, the class size showed an inverse relationship with the instructors rank – bigger classes, lower ratings. When they studied the interaction of instructor rank and instructor gender Wigington et al. (1989) found that at both ends of the ranking system, graduate assistants and full-professors, women got higher than men of the same rank.

The reputation of the instructor among the students may also be a factor. Wigington et al. (1989) found that students often take a class on the basis of the instructors' reputation. When this happened, the ratings were higher for lecture and laboratory class types. The interaction for reputation and rank showed the highest variation for professors. They received the highest scores from students who took classes for the instructors' reputation but they were also the lowest among the ranks when the students did not take the class because of the instructor's reputation. GTAs showed the least variation (Simmons, 1996).

Relationship between Student Ratings of Instructors and Distance from Teacher

In their study "Does the Distance from the Teacher Influence Student Evaluation", Safer, Farmer, Segala, and Elhoubi (2005) recorded student ratings of the overall effectiveness of their instructor in 75 freshman classes in California State University. The ratings were appraised in relation to seven independent variables; including number of students per class, number of rows

per class, mean student grade, instructor, time of the class, frequency of instruction per week, and whether we-instruction was offered. The findings show that student assessment of their instructors is influenced by the dimensions of the classroom (1) the greater the number of rows in the classroom, the lower the average student evaluations (2) higher scores in student evaluations were associated with higher student grades (3) individual instructors differed significantly from one another in student evaluations and (4) distance from the teacher is a significant variable in student evaluations (Safer et al., 2005).

Relationship between Student Ratings of Instructors and Attendance

Several student and course characteristics were examined in relation to student ratings of instruction. Beran and Violato (2005) asked students at a major Canadian University to complete the Universal Student Rating of Instruction (USRI) instrument at the end of every course over a three-year period. The analysis of between group differences indicated that students who attend class often and expect high grades provide high ratings of their instructors. In addition, Beran et al (2005) found that lab-type courses received higher ratings than lectures or tutorials and courses in the social sciences received higher ratings than courses in the natural sciences. They also concluded that student ratings were more related to teaching instruction and the behavior of the instructor than to student and course characteristics (Beran & Violato, 2005).

In the current study, the main goal was to examine the relationships among student ratings of university instructors based on interpersonal characteristics, pedagogical characteristics and course content characteristics. Based on the work of Young et al. (2009), we hypothesized that gender bias is expected in student ratings of their instructors. Specifically, we expected that male students would rate male instructors higher than female students. We also expected that female students would rate female instructors higher than male instructors. To find if student evaluations of instructors would be affected by student level, we expected that graduate students would rate instructors higher than undergraduate students. We also hypothesized that student ratings of their instructors would be affected by class size. Specifically, we anticipated that students in smaller classes would rate teachers higher than students in larger classes.

Chapter 3

Methodology

Participants

The sample included 161 graduate and undergraduate students in five universities in Lebanon participated in this study. The sample consisted of 81 students from Haigazian University (HU), 16 students from Notre Dame University (NDU), 20 students from Universite Saint Joseph de Beirut (USJ), 19 students from Balamand University (BU) and 22 students from the American University of Sciences and Technology – Zahle (AUST). Two surveys were deducted due to missing data.

The final sample included 75 male teachers and 84 female teachers. The sample also consisted of 81 male students and 78 female students. One hundred and forty seven of the responding students were working toward their undergraduate degrees and 12 for their graduate degrees. The sample was divided into four categories that represent the students' university majors. Sixty-six students were from the business major, 16 students from the engineering major, 43 students from the education major and 34 from the 'other' major (e.g. Biology, English, Armenian Studies…). Eighty-five percent of the respondents were under 22 years of age, 15% in the 22- 40 age group. Over 90% of the students chose to evaluate a course that was required for them. Students also reported their class sizes. There were 67 students in a class that included students less than 20, 81 in the 20-39 class size and 11 were in the 40-59 class size

Using a random sampling technique, students across all levels, majors and through different years of university experience were chosen for the study. In the case of the university with a graduate program, it was found that students from Education major only participated in the sample.

Instrument

Teacher Evaluation Scale (TES) was used to examine the students' evaluation of a memorable university instructor of their choice and not particularly the professor of the class they were attending. Students were asked to rate the instructor within the context of a particular course and to regard him/her in relation to other university professors they have had.

The scale included 25 items that were research based and showed a relationship with teacher effectiveness. Items included information about the instructor subject matter knowledge, communication skills, concern for students learning, sense of humor, preparation for class, and others. Since all items were literature based, content validity was strong (Young et al., 2009). The items were rated on a scale from one to nine, in which one meant 'not at all descriptive' and nine meant 'very descriptive'. In case no information was provided or the student felt the item did not apply, students were asked to answer 'Not applicable'. Items that included this notion were rated as zero.

To simplify the interpretation of items, three factors were identified. The first factor - interpersonal characteristics - included items (3, 5, 6, 7, 8, 9, 10, 11, 13, 14 and 18) that revealed how the instructors developed interpersonal relationships with students. The second factor – pedagogical characteristics – included items (1, 2, 4, 12, 15, 16, 17 and 22) that had a relationship

with the instructors' teaching approaches. The third factor – course content characteristics – included items (21, 23, 24, and 25) that reflected the importance of the course.

Items 19 and 20 did not load under any of three factors. To confirm the validity of the scale, a pilot study two months prior to the present study was designed and administrated. Then, the reliability coefficient for each factor was found. Crombach's Alpha for each factor was found; interpersonal characteristics, pedagogical characteristics, and course content characteristics were found to be .86, .92, and .94 respectively. Considering .7 and above as an acceptable reliability coefficient in Social Sciences, these high reliabilities indicated the consistency in students' evaluation of their instructors and further validated the current scale.

Procedure

A self-administered questionnaire survey design was used to collect data. The Teacher Evaluation Scale used in the research "Evaluating Gender Bias in Ratings of University Instructors' Teaching Effectiveness" by Drs. from the University of Wyoming is a recent scale. Prior to using it, Dr. Suzanne Young and Dr. Leslie Rush, were contacted to check about its validity and the frequency of its use. The reply was that the scale had high reliability coefficients across all items and it was only used for the purpose of that particular research. Therefore, it was decided to replicate the previous research on a Lebanese sample and to further validate its use.

Data were collected during and outside regular class meetings. Surveys were administered within a two week period during the spring semester. The participants were asked to rate the university professor they had in the recent past. Besides gender and gender role (of students and their best professors), information on a variety of demographic and background variables that have

been shown to effect teacher evaluations was also collected. The study involved a quantitative survey questionnaire to inspect students' evaluations of their instructor's quality characteristics.

Data Analysis

Data were analyzed using SPSS. Frequencies and descriptive information were found per item. Means for interpersonal characteristics, pedagogical characteristics, and course content characteristics were found and studied from different perspectives. Students' evaluations of instructors were examined in terms of student gender, instructor gender, student level, student major, course required and class size. To compare and investigate the means, Analysis of Variance (ANOVA) and t-test were used.

Chapter 4

Results

Sample

The sample included 159 students who rated 159 teachers from five universities of Lebanon (See Table 1).

Table 1

Description of the Sample

	n	Percent
Student Gender		
Male	81	50.9
Female	78	49.1
Instructor Gender		
Male	75	47.2
Female	84	52.8
Student Age		
18 – 22	135	85
23 – 40	24	15
Student Level		
Undergraduate	147	92.4
Graduate	12	7.6
Student Major		
Business	66	41.5
Engineering	16	10.1
Education	43	27
Other	34	21.4
Required Course		
Yes	144	90.6
No	15	9.4
Class Size		
Less than 20	67	42.1
20 to 39	81	50.9
40 to 59	11	6.9

Student Ratings in All Items

In Table 2, overall ratings refer to mean ratings of instructors per item. Results from student evaluations revealed that faculty received a mean score above 4.0 on the 25 items. The mean scores ranged from a low of 4.29 to a high of 6.78. Results revealed that item 3 was the lowest item rated (The instructor was enthusiastic about online teaching) and item 1 was the highest item rated (The instructor was knowledgeable about subject matter).

Students Ratings on Interpersonal Items

Table 2 also shows means of items that represent the interpersonal characteristics (3, 5, 6, 7, 8, 9, 10, 13, 14 and 18). Besides that item 3 received the lowest mean score, item 18 (The instructor was accessible outside of class) in this section had the lowest mean score of 5.78 and item 13 (The instructor genuinely enjoyed teaching) received the highest mean score of 6.45.

Students Ratings on Pedagogical Items

Table 2 also refers to means of items that represent the Pedagogical characteristics (1, 2, 4, 12, 15, 16, 17 and 22). All the items for this factor had mean scores that ranged from a low of 6.11on item 22 (The course was well organized) to a high of 6.78 on item 1 (The instructor was knowledgeable about subject matter).

Students Ratings on Course Content Items

The mean scores of items that correspond to the course content characteristics (21, 23, 24, and 25) are also found in Table 2. Item 21 (The course increased my interest in the subject matter)

received the lowest rating of mean score of 5.46 and item 24 (The course improved my understanding of concepts in the field) received the highest mean score of 6.13.

Table 2

Mean Scores per Item in Teacher Evaluation Scale

Item	Mean
1. The instructor was knowledgeable about subject matter.	6.78
2. The instructor communicated effectively.	6.17
3. The instructor was enthusiastic about online teaching.	4.29
4. The instructor was well prepared for each class.	6.35
5. The instructor created a comfortable learning atmosphere.	6.05
6. The instructor adapted to student needs.	5.95
7. The instructor was tolerant of others' ideas and views.	6.18
8. The instructor was genuinely respectful of students.	6.42
9. The instructor was warm and friendly.	6.19
10. The instructor had a good sense of humor.	6.10
11. The instructor motivated students to do their best.	6.04
12. The instructor was self-confident.	6.72
13. The instructor genuinely enjoyed teaching.	6.45
14. The instructor was concerned about student learning.	6.29
15. The instructor was able to explain material clearly.	6.45
16. The instructor identified important ideas.	6.30
17. The instructor used good examples to explain concepts.	6.44
18. The instructor was accessible outside of class.	5.78
19. The assignments were appropriate in amount and level.	5.43
20. The evaluation methods were appropriate.	5.72
21. The course increased my interest in the subject matter.	5.46
22. The course was well organized.	6.11
23. The course materials (text, readings, etc.) were worthwhile.	5.91
24. The course improved my understanding of concepts in the field.	6.13
25. The course was valuable to me.	6.01

Overall Student Ratings

Overall ratings refer to undergraduate and graduate mean ratings of the instructors in the three factors and across all variables. The means of each of the three factors by the independent variables were supplied. In the ANOVA findings of this study, the three factors; interpersonal characteristics, pedagogical characteristics and course content characteristics were carried out as

dependent variables. Student gender, instructor gender, student level, required course and class size were conducted as independent variables.

To determine the differences between mean scores on the dependent and independent variables, the significance level was set at .05. It would be worthy to note that differences in mean scores tend to be a result of chance rather than significance. Four of the ANOVAs, analyzing differences in course content characteristics and total score, yielded significant interaction effects. For interpersonal characteristics and pedagogical characteristics, groups did not differ significantly among any of the three independent variables or in their interactions.

This study used the Teacher Evaluation Scale (TES). The twenty-five items from the instrument were grouped into three factors: interpersonal characteristics, pedagogical characteristics and course content characteristics. As a preliminary contribution to check the reliability of the scale, the reliability coefficient for each factor was found by calculating Crombach's alpha for each factor. Crombach's alpha coefficient for each factor; interpersonal characteristics, pedagogical characteristics and course content characteristics were found to be .86, .92 and .85 respectively (See Table 3).

Table 3

Previous and Current Crombach's alpha for the three factors

	Previous Crombach	Current Crombach	Number of Items
Interpersonal Characteristics	**.94**	**.86**	**11**
Pedagogical Characteristics	**.93**	**.92**	**8**
Course Content Characteristics	**.91**	**.85**	**4**
Total Score		**.94**	**25**

Hypothesis 1: Male students rate male instructors higher than they rate their female instructors.

When we examined the differences among groups on course content characteristics, results showed a significant interaction between student gender and instructor gender ($F(1, 79) = 5.021$, $p = .028 < .05$) (See Table 4).

When answering the question of how does instructors' gender affect male students' evaluation on all three factors, the following results from the mean scores were found. It was found that male students rated male instructors ($M = 27.54$) significantly higher than they rated female instructors ($M = 23.82$) (See Table 4).

When we examined the differences among groups on total score, results showed a significant interaction between student gender and instructor gender ($F(1, 79) = 4.401$, $p = .039 < .05$) (See Table 4). It was found that male students rated male instructors ($M = 159.18$) significantly higher than they rated female instructors ($M = 145.71$) (See Table 4). Therefore, hypothesis 1 was confirmed.

Table 4

Males vs. Instructor Gender

	Male		Female			
	Mean	**SD**	**Mean**	**SD**	**F**	**P**
Interpersonal	73.92	13.64	69.32	13.46	2.105	.151
Pedagogical	57.71	11.09	52.57	13.39	3.405	.069
Content	27.54	7.15	23.82	7.03	5.021	**.028**
Total Score	159.18	6.49	145.71	29.32	4.401	**.039**

Note: Significant effects are in bold type

Hypothesis 2: Female students rate female instructors higher than they rate their male instructors.

When studying the effect of female students' ratings on female and male instructors, no significant differences were found on the student ratings in all factors and total score for female students towards instructor gender. Hence, hypothesis 2 was not confirmed (See Table 5).

Table 5

Females vs. Instructor Gender

	Male		Female			
	Mean	**SD**	**Mean**	**SD**	**F**	**P**
Interpersonal	70.22	11.76	72.66	18.17	.337	.563
Pedagogical	58.54	8.84	56.89	12.93	.302	.584
Content	27.59	5.62	25.30	8.47	1.362	.247
Total Score	156.36	22.74	154.85	3.02	.032	.859

Hypothesis 3: Graduate students rate instructors higher than undergraduate students.

When studying the effect of level on the three factors, hypothesis 3 was not confirmed because there was no significant difference in undergraduate and graduate students' ratings of instructors in all factors and total score (See Table 6).

Table 6

All Scales vs. Level

	Male		Female			
	Mean	**SD**	**Mean**	**SD**	**F**	**P**
Interpersonal	72.11	15.48	72.66	10.00	.015	.904
Pedagogical	56.49	12.12	58.33	10.46	.259	.611
Content	25.95	7.63	28.00	5.95	.820	.367
Total Score	154.56	31.28	159.00	23.90	.230	.632

Hypothesis 4: Students in smaller classes rate instructors higher than students in larger classes.

When examining the effect of the three factors on class size, it was found that students' ratings of instructors in classes 40 to 59 were highest in all factors (M = 72.63, M = 62.18 and M = 29.63) and in total score (M = 164.45) and lowest in classes that have students less than 20 in interpersonal (M = 71.56), in course content (M = 25.80) and in total score (M = 153.83). Hypothesis 4 was not confirmed since no significant differences were found in terms of class size in all three factors and in total scores (See Table 7).

Table 7

All Scales vs. Class Size

	Less than 20		20 – 39		40 - 59			
	Mean	SD	Mean	SD	Mean	SD	F	P
Interpersonal	71.56	19.03	72.58	11.71	72.63	10.87	.087	.917
Pedagogical	56.46	14.49	56.02	9.93	62.18	7.61	1.294	.277
Content	25.80	8.26	25.87	7.15	29.63	4.50	1.306	.274
Total Score	153.83	38.07	154.48	24.81	164.45	18.23	.576	.564

Further analysis on the student ratings of their instructors in terms of instructor gender, course type, class size and major revealed the following results.

Effect of Instructor Gender on Student Ratings of Instructors

When we examined the differences among groups in course content characteristics, results revealed a significant interaction between all students, irrespective of gender, and instructor gender ($F(1,157) = 5.438$, $p = .021 < .05$). All students rated male instructors ($M = 27.56$) significantly higher than female instructors ($M = 24.8$) (See Table 8).

Table 8

All Students vs. Instructor Gender

	Male		Female			
	Mean	**SD**	**Mean**	**SD**	**F**	**P**
Interpersonal	72.84	13.14	71.54	16.74	.288	.592
Pedagogical	57.96	10.43	55.45	13.17	1.741	.189
Content	27.56	6.70	24.80	8.00	5.438	**.021**
Total Score	158.36	25.33	151.80	34.74	1.808	.181

Note: Significant effects are in bold type

Effect of Course Type on Student Ratings of Instructors

When we examined the differences among groups on course content characteristics, results showed a significant interaction between all students and course type ($F = 8.402$, $p = .004 < .05$). Instructors were rated higher on required courses ($M = 72.20$, $M = 57.00$, $M = 26.65$ and $M = 155.86$) than elective courses ($M = 71.66$, $M = 53.13$, $M = 20.86$ and $M = 145.66$) in all factors and total score respectively (See Table 9).

Table 9

All Scales vs. Course Required

	Yes		No			
	Mean	**SD**	**Mean**	**SD**	**F**	**P**
Interpersonal	72.20	15.48	71.66	11.45	.017	.895
Pedagogical	57.00	11.88	53.13	12.79	1.417	.236
Content	26.65	7.32	20.86	7.67	8.402	.004
Total Score	155.86	31.01	145.66	27.27	1.498	.223

No significant results were found in terms of major in all three factors and total score.

Table 10

All scales vs. Major

	Business		Engineer		Education		Other			
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**F**	**P**
Interpersonal	71.24	11.75	70.68	11.94	76.44	15.52	69.20	20.38	1.737	.162
Pedagogical	55.87	11.94	53.31	11.16	60.72	11.5	54.50	13.02	2.598	.054
Content	26.51	15.52	24.37	8.33	26.93	7.90	25.08	9.19	.722	.540
Total Score	153.63	20.38	148.37	27.10	164.09	32.11	148.79	39.85	2.046	.110

In summary, male and female students rated their male and female instructors on three factors that associated with effective teaching. The three factors that were studied were interpersonal characteristics, pedagogical characteristics, and course content characteristics. Male students rated male instructors, in course content characteristics and total score, significantly higher than they rated their female instructors. All students rated male instructors, in course content characteristics, significantly higher than they rated female instructors. When the course was required, all students rated their instructors, in course content characteristics, significantly higher than when the course was not required. Female instructors were not rated by female students higher than male instructors neither in the three factors, nor in the total score. Undergraduate and graduate students did not rate their instructors differently. Students in smaller classes also did not rate their instructors higher than in larger classes.

Chapter 5

Discussion

The main goal of this research was to examine whether or not gender differences in three factors of instructor ratings (interpersonal characteristics, pedagogical characteristics, and course content characteristics) were found in student ratings and, if so, whether they were due to gender differences in student raters. This study was also intended to investigate if variables such as student level or class size had an effect on students' evaluations of their instructors' teaching effectiveness.

In this study, the researcher analyzed data of the Teacher Evaluation Scale that related to students' ratings of instructors teaching effectiveness. Four major findings emerged. Results showed that the Teacher Evaluation Scale was proved to be a valid scale. The most consistent findings in this study related to the course content characteristics factor. Results revealed that male students rated male instructors, in course content characteristics and in total score, significantly higher than female instructors. All students rated male instructors, in course content characteristics, significantly higher than female instructors. When the course was required, all students rated, in the course content factor, significantly higher than when the course was not required.

Effect of Student Ratings on Instructors in Terms of Gender

A - Male Students vs. Male Instructors

The present study showed that student gender and instructor gender played an essential role in how students perceived effective teaching. Bias refers to an attitude either for or against a

particular gender which involuntarily influences student ratings of teachers. Gender bias is often displayed in the form of opposite –sex or same-sex bias (Lahiri, 2010). In the interaction between student gender and instructor gender, Crombie et al., (2003) found that female students perceived less support and encouragement from their male instructors. This finding is also consistent with the student gender/teacher gender interaction found by Feldman (1993). Feldman reported that "students tend to rate same-gendered teachers a little higher than opposite-gendered teachers" (p. 151), and indeed the results were replicated in our study (Crombie et al., (2003). Previous research also revealed that male students rated male instructors, in pedagogical characteristics and course content characteristics, significantly higher than they rated their female instructors (Young et al., 2009). The results in the current study support this finding in the course content characteristics. Basow and Silberg (1987) reported that male students gave female professors significantly lower ratings than they gave male professors on six teaching evaluation measures (Basow & Silberg, 1987).

Corresponding findings have been reported by a number of other researchers. Basow (1998), Helgeson (1994), and Mulac and Lundell (1982) have all proposed that dynamism (enthusiasm) is more consistent with the masculine gender role stereotype; thus, students may not have perceived dynamism in the female professor. Therefore, male students rated male instructors higher than female instructors (Arbuckle et al., 2003).

Generalizing from this earlier work, we might expect that male students would perceive their male instructors more positively than their female instructors; this finding is in accord with the result in the present study. When rated by male students, male instructors received, in the course content characteristics and the total score significantly higher ratings than did female

instructors. Hence, gender bias might be found since student gender interacted with instructor gender for one of the three factors of teaching effectiveness. Male students tended to pick professors with perceived masculine behavior as their 'best' instructors. This might be related to the idea that men tend to be more traditional than women in terms of their attitudes toward women (Lahiri, 2010). Moreover, male students might refer to their male instructors as their role models and hence, they rate them significantly higher than they rate female instructors.

B – All Students (male and female) vs. Male Instructors

Less favorable ratings of women tend to mostly occur when women are seen as not fitting gender stereotypes, for example by participating in a sex-atypical profession (e.g., Etaugh & Riley, 1983). Even if instructors treated male and female students similarly, research by Tannen (1990), suggested that males and females might be using different rules to evaluate the communication patterns (and other behaviors) of instructors. This might result in female professors being judged more negatively by male students and vice versa (Das & Das, 2001). Basow and Silberg (1987) found that female students evaluated female professors less favorably than male professors on three measures of teaching effectiveness (Basow & Silberg, 1987). A study by Das and Das (2001) examined students who had been in the workplace and returned to study. The findings revealed that male and females students favored their male instructors. The researchers hypothesized that this happened because most businesses in today's world encourage masculine values and the fact that males occupy most positions of power (Das & Das, 2001). In addition to the aforementioned and to the truth that college teaching might be considered a male occupation, all the above might have resulted in that all students (male and female) rated male instructors, in course content characteristics, significantly higher than female instructors.

Effect of Course Content Characteristics on Student Ratings of Instructors

Although it might be assumed that students evaluate more positively courses in their major and smaller classes, existing research indicates that the relationships among course characteristics, student characteristics, and faculty evaluations are mixed at best. For example, Centra (1981) found that the relationships among course and student characteristics and student evaluations are very small and generally insignificant. Marsh (1987) and Simmons (1996) reported that smaller classes tend to be evaluated more highly and less reliably, although Feldman (1978) found no significant difference by class size (Obenchain et al., 2001).

Addison et al. (2006) hypothesized that students' evaluations of instruction would be associated with their perceptions of the difficulty of the course. For instance, it was assumed that students who found the class more difficult than they thought it was going to be tend to assign low evaluations, while those who found the class easier than they thought it was going to be assign somewhat higher evaluations. This hypothesis was confirmed in the sense that undergraduate students evaluated different professors more or less favorably, based in part on their perceptions of the course as more or less difficult as they had anticipated. As reported by Smith et al. (2007), Centra (1993) noted that the structure and nature of some courses are more appropriately measured on certain dimensions. For instance, student-instructor interaction is essential in courses with small enrollments or in seminars and might need to be weighted more heavily at times. Instructor involvement or organization might be most relevant in large lecture contexts where developing and preserving student interest is essential to student learning. Therefore, applying specific rating criteria for certain types of courses will increase the effect of theoretical and practical issues, especially focusing on the rationale for and validation of criteria selection (Smith et al., 2007).

Implications for the Previous Findings

What implications could be drawn from this study? Although it might be assumed differently, Lebanese students in the current study revealed promising results. Unlike college American students, Lebanese students seemed to have overcome differences that might exist between male and female instructors on interpersonal and pedagogical characteristics. This is promising in the sense that students tended to judge their instructors teaching behaviors and techniques in an objective way irrespective of gender, character or behavior differences. The students' ratings showed no preference based on character or behavior, they based their ratings on the course content characteristics. This might be considered relevant since Lebanese students did not portray differences in instructors' interpersonal characteristics and pedagogical characteristics based on gender. In this sense, they tended to reflect a view of equality between male and female instructors and this was reflected in their ratings.

Coming from a patriarchal society, male students tended to rate male instructors higher than female instructors. Even female students seemed to have kept the traditional view of having male as the dominant figure in the Lebanese society and hence, they tended to replicate this cultural fact. Hence, our assumptions that male and female students would rate male instructors higher than female instructors were confirmed.

Since gender bias existed, in the current study, only from the male perspective and not from the female perspective, it is worthy to note the following. If the differences in ratings were to be treated as mere data collection, then no harm would seem to be done. If, however, the differences in ratings were to be treated for promotions and personnel decisions in a way that whichever sex was favored by the ratings would enjoy benefits, and then gender bias might exist. Feldman

(1993) noted, as reported by Smith et al. (2007), that "differences in ratings are not in and of themselves reflective of [gender] bias. The differences might reflect actual differences in instructor involvement, student interest, student-instructor interaction, course demands, and course organization that are reflective of the instructor skill of one sex over the other" (p. 7). Hence, the verification of gender bias entails other more objective measures of teaching skills to compare with the student rating scores (Smith et al., 2007).

In the current study, students perceived a significant difference in instructors, in course content characteristics, such as valuable course, improved understanding, and worthwhile materials. However, this difference was not relevant in interpersonal characteristics or pedagogical characteristics. That gender bias did not play a role in the student evaluations of instructors' interpersonal characteristics suggested that all students did not perceive a difference in their instructors in terms of personality characteristics, such as friendliness, humor, motivation and enthusiasm. All students also did not perceive a difference in their instructors' pedagogical characteristics, in things that were related to clear explanations, subject matter knowledge, communication, self-confidence, organization and preparation for the course. This is promising since it tended to diminish the potential of gender bias in students' evaluations of their instructors. It could be concluded that gender bias did not have an effect on the way students rated their instructors on interpersonal characteristics and pedagogical characteristics but had an effect on the course content characteristics. This might be related to the fact that the interpersonal characteristics and pedagogical characteristics did not affect how students perceived their instructors and their classroom strategies. The students' ratings tended to be affected by the nature of the course itself.

Effect of Required Course on Student Ratings of Instructors

Several studies have shown that student ratings differ in academic disciplines such that courses in sciences and engineering tend to be rated lower than courses in humanities and social sciences (Centra, 1993; Kwan, 1999; Ory, 2001; Watchel, 1998). This phenomenon is usually clarified by the general difficulty students have learning course material in scientific and technical fields (Nasser & Fresco, 2006).Upon checking the reliability of students' ratings of faculty teaching effectiveness, Obenchain et al. (2001) found a significant difference between teacher education students' evaluations of teacher education courses and non-teacher education courses. They found that teacher education students rated teacher education faculty higher than non-teacher education faculty (Obenchain et al., 2001).

Previous research by Crombie et al., (2003) showed significant discipline effects that were observed in general favored the arts/social science courses. For example in the case of active students, their total participation scores were significantly higher and the length of their interactions longer in arts/social science classes than in science classes. Professors in arts/social science courses were perceived to investigate or seek elaboration more frequently than were science professors.

Beran and Violato (2005) reported that courses in social sciences received significantly higher ratings than courses in natural sciences. Required courses entail support and often individual or small group interaction between students and instructors, which might leave students with more positive impressions about their instructors than if the instructor had less contact with students through lectures (Beran and Violato, 2005).

Therefore, the finding that students' ratings of their instructors are higher, when the course is required, is consistent with what previous researchers tried to find. When the course was required, all students rated their instructors, in the course content characteristics, significantly higher than when the course was not required. Students seemed to agree that the course improved their own understanding of concepts in the field. This was relevant in the high mean score that item 24 received (6.13). It could be common sense for students to rate a required course higher than a non required course. Required courses tend to be the courses that students have in their major in which they have chosen and consequently would like or be interested in. However, when it comes to non required courses, a student might feel obliged to attend it just because it is a university requirement whether they like it or not. In required courses, students tend to get higher grades than non – required courses. Grades have been associated with higher ratings in instructors' evaluations. Grades might be considered clearly predictive of student evaluations in the stereotypically predicted direction, with higher grades associated with more favorable student evaluations. Addison et al. (2006) found that when the high grades were earned in a class that is perceived as easy, then high grades might not have had a particularly strong effect on student evaluations. They also found that when the high grades were earned in a course perceived as difficult, then earned high grades played a strong role in helping to produce high student evaluations (Addison et al., 2006).

Effect of Student Level on Student Ratings of Instructors

Although the current results did not yield differences in terms of level, however, Basow and Silberg (1987) reported that there was a positive correlation between student level and teacher

ratings for undergraduate students participating in their study. They examined five factors: scholarship, organization and clarity, interaction with the group, interaction with individual students, and enthusiasm. They reported that the higher the student level, the higher was the teacher rating (Basow & Silberg, 1987). Donaldson, Flannery, and Ross-Gordon (1993) reported comparative findings from three studies of adult students. They found differences in age group expectations: younger students were most interested in traits that might improve their own tasks (i.e. being successful in school) while older students were more attentive to relationship issues such as teachers who are devoted and who inspire students to do their best (Young et al., 2009).

Although, in the current study, it was expected that graduate students would give higher ratings than undergraduate students. However, graduate and undergraduate students did not rate their instructors differently. This might be related more to the small sample of graduate students (n = 12) represented in the current study rather than that the interaction of student gender and instructor gender which generalized across student levels.

Effect of Class Size on Student Ratings of Instructors

Contrary to results in other studies, the current study did not find that student ratings of their own instructors were influenced by class size. Research by Crombie at al. (2003) on the effect of class size on student participation in class showed that students in smaller classes perceived their professors more favorably, on all five measures examined, than did students in larger classes. We expected students in smaller classes to evaluate teachers more positively, and this is indeed the finding reported in the literature (Marsh & Roche, 1997).

Effect of Major on Student Ratings of Instructors

When studying evaluation of teaching in terms of discipline or major, Cashin (1990) and Feldman (1978) reported a fairly consistent pattern of student-completed evaluations that favored soft and applied disciplines over math, engineering, and many sciences. Smart and Elton (1982) provided a justification for the difference; they found that faculty in soft and applied disciplines place a larger emphasis on the process of teaching (Obenchain et al., 2001). In the current study, major, as a contextual factor, did not yield any significant effect.

Limitations of the Study

While the primary goal of this study was to examine students' ratings of their instructors through Teacher Evaluation Scale, there were certain limitations that should be emphasized. First, it is possible that the findings were subject to two biases. One is confirmation bias - students noticing feminine behaviors in female instructors and masculine behaviors in male instructors and second role model bias - students choosing the instructor they liked most as being similar to themselves. Second, although the findings of this study are significant, it is possible that students' evaluations of their instructors were affected by other variables. Third, this study was limited to 5 universities in Lebanon and it could not be generalized to all Lebanese society. Fourth, the current sample tended to be small, selective and underrepresented in the case of graduate students; hence a replication in other settings might be needed. Fifth, the fact that students were asked to choose one instructor for their ratings might limit the possibility of rating other effective instructors of same or different gender.

Conclusion and Further Research

In sum, this study explored student ratings of instructors' teaching effectiveness. The results of this study provide only a partial support for the research done in the field of students' evaluation of teaching. Male students rated male instructors, in course content characteristics and in total score, significantly higher than female students. This same finding is paralleled in the research performed by Young et al. (2009). All students rated male instructors, in course content characteristics, significantly higher than female instructors. When the course was required, all students rated, in the course content factor, significantly higher than when the course was not required. Hence, these findings have implications in terms of students' evaluation of instructors.

Therefore, from a global perspective, it is important that future research in this context examine teaching effectiveness through assessment that includes multiple means along with student evaluations. For instance, to ensure effective evaluation of instructors, clear communication between students and faculty regarding the purpose of evaluation is essential to make the process valid and useful. According to Rolheiser and Ross (2010), 'self-evaluation is defined as students or teachers' judging the quality of their work, based on evidence and explicit criteria, for the purpose of doing better work in the future' (p:1) (Rolheiser & Ross, 2010). An example of self-assessment is the development of a Teaching Portfolio. The Teaching Portfolio might include a record of achievements, success stories, and items that document personal and professional growth over the course of a faculty member's tenure. Young et al. (2009) suggested that it is important that awareness be stressed of the potential for gender bias in ratings of course content characteristics for instructors and that instructors themselves, should work to make students aware of this potential for gender bias in evaluations and to help them become aware of it

and reduce it (Young et al., 2009). DeFina (1996), as reported by Smith (2008), suggested that a combination of student, peer, and self-evaluation would supply a more reliable assessment of teaching effectiveness than any form of evaluation alone (Smith, 2008). Narasimhan (2001) suggested that teachers do not have to change their delivery just to suit the students; they might have to attempt to change student's expectations, if they were found to be unrealistic. This could be applied by generating discussions, which could be an educational experience for the students. For this exercise to be beneficial, a climate of trust between the teacher and the students is essential. McKeachie & Kaplan (1996) suggested that this approach could motivate students to complete the student evaluation questionnaires more seriously. Furthermore, Deming (1986) considered that this open approach might assist in `driving out fear' from students and teachers and might help create an atmosphere of trust, which is necessary for innovation (Narasimhan, 2001).

From a relevant perspective based on the results of the current research, it would be interesting to find interactions between students' ratings of instructors and the physical environment that surrounds them such as the distance of the student from the teacher. It would also be remarkable to replicate this research on the Lebanese University or in other private Universities that follow an Arabic curriculum rather than an American one and check whether cultural differences persist.

Since the current study only revealed significant results on the course content characteristics, it is evident that policy decisions dealing with faculty should not be based on such findings. On the practical level, any rewards for instructional excellence should not be based on the sex of the instructor, however, on the student ratings of instructors determined on a course-by-course basis. Hence, applying specific rating criteria for certain types of courses might be useful.

References

Addison, W.E., Best, J. & Warrington, J. D. (2006). Students' Perceptions of Course Difficulty and their Ratings of the Instructor. *College Student Journal*, 40 (2), 409 – 416.

Ahmadi, R.T. & Cotton, S.E. (1998). Assessing Students' Rtaings of Faculty. *Assessment Update*, 10 (5), 5 – 7.

Arbuckle, J. & Williams, B. D. (2003). Students' perceptions of expressiveness: age and gender effects on teacher evaluations. *Sex Roles: A Journal of Research*, 49 (9 - 10), 507 – 519.

Bedard, K. & Kuhn, P. (2005). Where Class Size Really Matters: Class Size and Student Ratings of Instructors Effectiveness. University of California, Santa Barbara, 1 – 26.

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

Boysen, G. A. (2008). Revenge and Student Evaluations of Teaching. T*eaching of Psychology*, 35, 218 – 222.

Brown, M. J. (2008). Student Perceptions of Teaching Evaluations. *Journal of Instructional Psychology*, 35 (20), 177 – 183.

Bruce, L. B. (2001). Student Self – Assessment: Making Standards Come Alive. *Classroom Leadership*, 1 – 9.

Chermesh, R. & Tzelgov, J. (1979). The College Instructor as a Leader: Some Theoretical Derivations from a Generalization of a Causal Model of Students' Evaluation of Their Instructors. *Journal of Educational Research*, 73 (2), 101 – 115.

Combie, g., Pyke, S. W., Silverthorn, N., Jones, A. & Piccinin, S. (2003). Students' perceptions of their classroom participation and instructor as a function of gender and context. *Journal of Higher Education*, 74 (1), 51 – 69.

Das, M. & Das, H. (2001). Business students' perceptions of best university professors: does gender role matter. *Sex roles: A Journal of Research*, 45 (9 -10), 665 – 674.

Davidovitch, N. & Soen, D. (2006). Class Attendance and Students' Evaluation of Their College Instructors. *The College Student Journal*, 40 (3), 691 – 703.

Diseth, A. (2007). Students' Evaluation of Teaching, Approaches to Learning, and Academic Achievement. *Scandinavian Journal of Educational Research*, 51 (2), 185 – 204.

Dolmans, D., Van Lujik, S. J., Wolfhager, I. & Scherpbier, A. (2006). The relationship between professional behaviour grades and tutor performance ratings in problem – based learning. *Medical Education*, 40, 180 – 186.

Doyle, T. (2001). Evaluating Teacher Effectiveness – Research Summary. *Center of Teaching and Learning.*

El Hassan, K. (2009). Investigating substantive and consequential validity of student ratings of instruction. *Higher Education Research & Development*, 28 (3), 319 – 333.

Fernald, P.S. (1990). Students' Ratings of Instruction: standardized and Customized. *Teaching of Psychology*, 17 (2), 101 – 109.

Gallagher, H. A. (2004). Vaugh Elementary's Innovative Teacher Evaluation System: Are Teacher Evaluation Scores Related to Growth in Student Achievement?. PEABODY *Journal of Education*, 79 (4), 79 – 107.

Guevara, A. P. & Adriano, C. T. (2002). Student Evaluation of Teachers. *Office of the Director of Instruction.*

Harris, E. (1982). Student Ratings of Faculty Performance: Should Departmental Committees Construct the Instruments?. *Journal of Educational Research*, 76 (2), 100 – 107.

Heckert, T. M., Latier, A., Ringwald, A. & Silvey, B. (2006). Relation of Course, Instructor, and Student Characteristics to Dimensions of Student Ratings of Teaching Effectiveness. *College Student Journal*, 40 (1), 195 – 203.

Higher Education in Lebanon (2009).

Hoxby, C. M. (2000). The effect of Class Size on Student Achievement: New evidence from population variation. *Quarterly Journal of Economics*, 1239 – 1285.

Intime (2001). Knowledge of Students' Characteristics: Definition and checklist. Integrating new technologies into the methods of education.

Kember, D., Jenkins, W. & Chi Ng, K. (2004). Adult students' perceptions of good teaching as a function of their conceptions of learning – Part 2. Implications for the evaluation of teaching. *Studies in Continuing Education,* 26 (1), 81 - 97.

Koon, J. & Murray, H. G. (1995). Using multiple outcomes to validate student ratings of overall teacher effectiveness. *Journal of Higher Education*, 66 (1), 61 – 73.

Kyriakides, L., Denetriou, D. & Charalambous, C. (2006). Generating Criteria for evaluating teachers through teacher effectiveness research. *Educational Research*, 48 (1), 1 – 20.

Lahiri, S. (2010). Status of Female Teachers When Students Evaluating Teachers. Faculty of Education, Banaras Hindu University.

Lomas, L. (2001). Does the development of mass education necessarily mean the end of quality? The sixth QHE Seminar.

Lynch, d. J. (2006). Motivational Factors, Learning Strategies and Resource Management as Predictors of Course Grades. *College Student Journal*, 40 (2), 423 – 428.

Marsh, H. W. & Bailey, M. (1993). Multidimensional students' evaluations of teaching effectiveness: a profile analysis. *Journal of Higher Education*, 64 (1), 1 – 8.

Miley, W. M. & Gonslaves, S. (2009). What you don't know can hurt you: students' perceptions of professors' annoying teaching habits. *College Student Journal*, 37 (3), 447 – 453.

Millea, M. & Grimes, P. (2002). Grade expectations and student evaluation of teaching. *College Student Journal*, 36 (4), 582 – 589.

Mintzes, J. J. (1979). Overt Teaching Behaviors and Student Ratings of Instructors. *Journal of Experimental Education,* 48 (2), 139 – 153.

Munz, D. C. (2009). Student Mood and Teaching Evaluations. *Journal of Social & Behavior Personality.*

Narasimhan, K. (2001). Improving the Climate of Teaching Sessions: the use of evaluations by students and instructors. *Quality in Higher Education*, 7 (3), 179 – 192.

Nasser, F. & Fresko, B. (2006). Predicting student ratings: the relationship between actual student ratings and instructors' predictions. *Assessment & Evaluation in Higher Education*, 31 (1), 1 – 18.

Nizam, J. (2009). Traits of Faculty Members. Higher Education in Lebanon. (Chapter 10).

Obenchain, K. M., Abernathy, T. V. & Wiest, L. R. (2001). The Reliability of Students' Ratings of Faculty Teaching Effectiveness. *College Teaching*, 49 (3), 100 – 107.

Office of Institutional Research (2010). Student ratings of Instruction – Report and Recommendations.

Okpala, C. O. & Ellis, R. (2010). The perceptions of College Students on Teacher Quality: A Focus on Teacher Qualifications. *Education*, 126 (2), 374 – 382.

Ozogul, G. & Sullivan, H. (2009). Student performance and attitudes under formative evaluation by teacher, self and peer evaluators. *Education Tech Research Dev*, 57, 393 – 410.

Perlman, B., McCann, L.I. & Prust, A. (2007). Students' Grades and Ratings of Percieved Effectiveness of Behaviors Influencing Academic Performance. *Psychology Press*, 34 (4), 236 – 242.

Rolheiser, C. & Ross, J. A. (2010). Student Self –Evaluation: What Research Says and What Practice Shows. *Center of Development and Learning.*

Rosenblit, S.G., Sebkova, H. & Teichler, U. (2007). Massification and Diversity of Higher Education Systems: Interplay of Complex Dimensions. *Higher Education Policy*, 20, 373 – 389.

Safer, A. M., Farmer, L. S. J., Segalla, A. & Elhoubi, A. F. (2005). Does the Distance from the Teacher Influence Student Evaluations?. *Educational Research Quarterly*, 28 (3), 28 – 35.

Scriven, M. (1995). Student ratings offer useful input to teacher evaluations. *Practical Assessment, Research & Evaluation*, 4 (7), 1 – 5.

Simmons, T.L. (1996). Student evaluation of teachers: professional practice or punitive policy?. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 1 (1), 12 – 19.

Smith, B. P. (2008). Student Ratings of Teaching Effectiveness: An Analysis of End – of Course Faculty Evaluations. *College Student Journal*, 41 (4), 788 – 900.

Smith, S. W., Yoo, J. H., Farr A. C., Salmon, C. T., & Miller, V. D. (2007). The Influence of Student Sex and Instructor Sex on Student Ratings of Instructors: Results from a college of Communication. *Women's Studies in Communication*, 30 (1), 64 – 74.

Sojka, J., Grupta, 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), 44 – 53.

Spooren, P. & Mortelmans, D. (2006). Teacher professionalism and student evaluation of teaching: will better teachers receive higher ratings and will better students be given higher ratings?. *Educational Studies*, 32 (2), 201 – 214.

Strage, A. (2008). Traditional an Non – Traditional College Students' Descriptions of the " Ideal" Professor and the " Ideal" Course and Perceived Strengths and Limitations. *College Student Journal*, 42 (1), 225 – 231.

Tata, J. (1999). Grade Distributions, Grading Procedures, and Students' Evaluations of Instructors: A Justice Perspective. *Journal of Psychology*, 133 (3), 262 – 271.

The Higher Education System in Lebanon, A Brief Overview (2009).

Weber, L. J. & Frary, R. B. (1982). Profile Uniqueness in Student Ratings of Instruction. *Journal of Experimental Education*, 51 (1), 38 – 45.

Wikipedia, (2010) Student Evaluation of Instructor.

Wilson, J. H. (2006). Predicting Student Attitudes and Grades from Perceptions of Instructors' Attitudes. *Teaching of Psychology,* 33 (2), 91 – 95.

Wright, R. E. & Palmer, J. C. (2010). A Comparative Analysis of Different Models Explaining the Relationship between Instructor Ratings and Expected Student Grades. *Educational Reasearch Quarterly*, 30 (2), 3 – 20.

Yao, Y., Weissinger, E. & Grady, M. (2003). Practical Assessment, Research & *Evaluation. Research & Evaluation*, 8 (21), 1 – 12.

Young, S., Rush, L. & Shaw, D. (2009). Evaluating Gender Bias in Ratings of University Instructors' Teaching Effectiveness. *International Journal for the Scholarship of Teaching and Learning*, 3 (2), 1 – 14.

Appendix A

To whom it may concern,

My name is Samar Haddad. I am a graduate student in Education with emphasis in "Administration and Supervision" at Haigazian University in Beirut.
I kindly ask you to allow me to do my research study at your university. I would highly appreciate your support in supplying 5 minutes of your students' and instructors' time to conduct this study.
The purpose of my research is to study "Student Ratings of their College Instructors". This research is tailored to measure the relation between class attendance, course characteristics and instructor gender and the evaluation of teachers.
Your participation in this study would provide valuable data and it is essential to note that the research will by no means reveal personal information about your institution.

Thank you for your cooperation, your help is highly appreciated.

Regards,

Samar Haddad

Appendix B

Teacher Evaluation Scale

Thank you for taking the time to answer the following questions about an instructor you have had in the recent past. As you rate your instructor within the context of a particular course, consider him/her relative to other university instructors you have had. Please rate each item indicating the degree to which you feel the item is descriptive of the instructor or course; where 1=not at all descriptive and 9=very descriptive. If you have no information or you feel the item does not apply, circle NA (Not applicable).

1. The instructor was knowledgeable about subject matter.
2. The instructor communicated effectively.
3. The instructor was enthusiastic about online teaching.
4. The Instructor was well prepared for each class.
5. The instructor created a comfortable learning atmosphere.
6. The instructor adapted to student needs.
7. The instructor was tolerant of others' ideas and views.
8. The instructor was genuinely respectful of students.
9. The instructor was warm and friendly.
10. The instructor had a good sense of humor.
11. The instructor motivated students to do their best.
12. The instructor was self-confident.
13. The instructor genuinely enjoyed teaching.
14. The instructor was concerned about student learning.
15. The instructor was able to explain material clearly.
16. The instructor identified important ideas.
17. The instructor used good examples to explain concepts.
18. The instructor was accessible outside of class.
19. The assignments were appropriate in amount and level.
20. The evaluation methods were appropriate.
21. The course increased my interest in the subject matter.
22. The course was well organized.
23. The course materials (text, readings, etc.) were worthwhile.
24. The course improved my understanding of concepts in the field.
25. The course was valuable to me.

Please tell us a little about yourself and about the course.

You are: ____ Male ____Female

Your age: _____

Your student level: ____ Undergraduate ____Graduate

Your major: _____

Your instructor was: ____ Male ____ Female

Was the course required? ____ Yes ____ No

Approximate class size?

_____less than 20 _____ 20-39 _____ 40-59 _____Greater than 59