A CRITICAL APPRAISAL OF STUDENTS EVALUATING TEACHING EFFECTIVENESS TO MANAGE THE QUALITY OF HIGHER EDUCATION INSTITUTIONS

Ahmed Al Kuwaiti, Imam Abdulrahman Bin Faisal University (IAU)
Arun Vijay Subbarayalu, Imam Abdulrahman Bin Faisal University (IAU)
Ajayan Kamalasanan, Imam Abdulrahman Bin Faisal University (IAU)
Sivasankar Prabaharan, Imam Abdulrahman Bin Faisal University (IAU)
Vinoth Raman, Imam Abdulrahman Bin Faisal University (IAU)

ABSTRACT

The utility of students' evaluations to observe the quality of teaching has demonstrated to be a practical and vital element of the quality management process in higher education institutions (HEIs). Since students are the most affected by teaching, their perception and satisfaction are essential in enhancing the quality of lectures delivered by the teaching faculty. The practice of evaluating lecturing skills of faculty through student feedback is one among the crucial means in the internal quality assurance processes at HEIs. Several factors are influencing the outcome of student evaluations of teaching, and various studies from different perspectives have addressed it. However, to date, no reviews have the findings critically appraised to bring out a list of common factors influencing students evaluating teaching effectiveness (SETE). This study critically evaluated the literature and identified three critical factors, i.e., instructor-centered, student-centered, and course-centered difficulties, that influence students though assessing teaching effectiveness. Through this study, the authors also highlighted the utility of student evaluations to monitor the teaching quality and limitations on using it. The authors concluded that although SETE has several benefits to enhance the quality of higher education, however, it should be used appropriately to overcome the possible barriers, as highlighted in this study.

Keywords: Critical Appraisal (CA), Students Evaluation of Teaching Effectiveness (SETE), Quality, Higher Education Institutions (HEI)

INTRODUCTION

Quality management process in universities is an intricate phenomenon since the course of action involves multifaceted facades, which comprises of all the stakeholders of the academic community like the faculty, students, and different support facilities, namely admission, registration, and other support units like the laboratories, libraries and many more. As fragment of quality management, the Higher Education Institutions (HEIs) are performing continuous student evaluations of the courses and programs, which include assessing teaching skills of teaching faculty, including facilities and services that they offered. To carry out these evaluations, different assessment tools are used by academia, which consists of surveys like course evaluations, mid-course evaluations, and program evaluations (Al-Rubaish, Wosornu & Dwivedi, 2011; Al Rubaish 2011). In addition to the course and program evaluations, the Students Experience Survey (SES) is yet another tool used by the institution to generate student's feedback on the offered quality of education, which is usually considered at the mid-course of a program (Al-Rubaish, 2010; Al Kuwaiti & Subbarayalu, 2015). Further, the Students Evaluating Teaching Effectiveness (SETE) is

another quality assessment tool that is considered as a useful technique for observing the teaching quality and quality of learning in HEIs (Clayson, 2009; Greenwald, 1997; Pounder, 2008). The quality of an education program can be assessed indirectly using students' satisfaction, which is an integral component of academic quality management (Al-Rubaish, 2010).

Even though there are several advantages in utilizing student assessments of teaching in managing the quality of higher education, there are many shortcomings in it, which needs to be explored. This present study is the first of its kind to critically review the historical perspectives, utility, and the factors influencing students' evaluation of teaching in HEIs worldwide. The author addressed the critical issues related to student evaluations of teaching based on the published literature from the year 1927 to 2019.

SETE: A HISTORICAL PERSPECTIVE

Many research publications emerged out from student evaluations of teaching effectiveness, which has an enduring history (Baliyan & Moorad, 2018; Chisholm, 2014; Feistauer & Richter, 2018; Hornstein, 2017; Mart, 2017; Rowan, et al., 2017; Spooren, Brockx & Mortelman, 2013). The initial traces of the literature were found from the contributions of George Brandenburg and Hermann Remmers, who led numerous studies utilizing a self-designed instrument, the "Purdue Rating Scale for Instructors" (Brandenburg & Remmers, 1927; Remmers, 1928, 1930; Remmers & Brandenburg, 1927). Further, Remmers worked together with other researchers on factor analyses of student assessments making use of a type of Purdue scale, which consists of just ten items or "traits" (Smalzried & Remmers, 1943). The ten traits include distinct qualities as "presentation of the subject matter" and "fairness in grading," in addition to certain traits which are not so relevant, such as "personal appearance" and "personal peculiarities." The findings of researchers produced two aspects, which those termed "professional maturity" and "empathy." Later, Edwin Guthrie (1927) led an initial study of SETE and investigated the variations in student feedback at the commencement of the academic year to the scores assumed at the culmination of the year.

During the 1960s, the teachers needed to undergo student evaluations, which emerged out as a theme of significant research in the 1970s. Right from the middle of 1970s through the beginning of the 21st era, Marsh carried out quite a lot of studies on SETE using a reliable tool entitled, the "Students' Evaluations of Educational Quality (SEEQ)" (Marsh 1977, 1980, 1984, 1987; Marsh & Hocevar, 1991; Marsh & Roche, 1997, 2000). When compared to other instruments, the tool is observed to be reliable, relatively valid in measuring teaching effectiveness, and not mostly influenced by the possible prejudices like classifying lenience and class size. Currently, many of the student evaluations aimed on the use of emerging expertise/devices with various alternations and also counting the student and teacher expectations (Slocombe, Miller & Hite, 2011; Umar et al., 2016; Yahya & Rostam, 2015).

The Utility of Students Evaluations to Monitor the Quality of Teaching

The intricacies related with the application of student evaluations endured prevailing in the 21st century, where Douglass (1928) noted the prevalence of contradictory thoughts and ideas related to student rating of teaching assessments. As student ratings are the crucial element to evaluate teaching effectiveness, active and significant input from students are key elements in the triumph of a teaching assessment system. Research proposes that appropriately planned SETE are effective measures of teaching effectiveness (Greenwald, 1997). Also, Moore & Kuol (2005) indicated that students evaluating teaching quality are extensively accepted measures of academic quality management. Seldin (1993) reported that the utility of SETE had been augmented to 86% from the baseline of 29% in the United States of America. The HEIs also depend on students'

feedback on various aspects, which are the basic functions, including lecturing skills, academic programs and courses offered (Aultman, 2006). A study conducted by Vijay (2013) utilized students experience survey to assess the views of the students on learning and teaching facilities and provided an innovative student focused quality-rating practice for the education amenities adopted at HEIs in India

A previous study highlighted four different reasons for collecting SETE (Marsh & Dunkin., 1992): (i) investigative feedback about the instructors' teaching effectiveness; (ii) assessment on teaching effectiveness to be utilized in organizational decision making; (iii) evidence for students for choosing the course units and instructors; (iv) A result or process explanation for integrating research in teaching. Moreover, logical explanations exist for why student assessments should result in enhancement of teaching quality. Those explanations comprise: the drive to attain tenure using worthy outcomes, the supplementary inspiration to pursue the assistance of expert, and the common support feedback on an action can deliver (Murray, 1997). Lately, Al-Abbadi, et al., (2009) indicated that teacher assessments by students meet the two drives that include (i) the students can validate lecturer teaching methods and can alarm the lecturers to create optimistic variations in their teaching, (ii) to reveal how satisfied students are with the organizational process, the course, its content, and the teaching process, as eased by the lecturer.

Factors Perceived to Influence Students' Ratings during the Teaching Evaluation Process

Some researchers found that the student's attitudes over the assessment of teaching effectiveness showed a significant association with the success of teaching assessment system (Marsh, 1984, 1987; Hejase et al., 2013; Umar et al., 2016). Also, students strongly believed that the University administration should rely on students' evaluations for the promotion decisions of the faculty (Sojka, Gupta & Deeter-schmelz, 2002). Here the authors were carrying out a thorough review of literature related to SETE and identified different factors that might influence SETE. Such factors are sorted into three groups that include instructor-centered, student-centered, and; course-centered difficulties. The primary input of this article is to reveal how these several factors affecting students' evaluation of teaching in HEIs.

Students-Centered Factors

Student Gender

While dealing with the gender of the teacher and the student, an intricate interface is experienced between them. This is observed by different studies, which illustrates that female students, when equated to their counterparts, tend to usually provide more ratings indiscriminately (Lavin, Korte & Davies, 2012; Korte, Lavin & Davies, 2013; Young, Rush & Shaw, 2009). This may be because the feminine gender always demonstrated a stereotypic nature of the response, which is usually positive, and it may be influenced by different factors like knowledge, classroom environment, and presentation skills of the instructor (Aronson, Wilson & Akert, 2007). However, different researches specified that the inclination towards same-sex exists, where female students provided higher ratings to faculties of the same gender, while male students favored the instructors from the same sex (Young et al., 2009; Korte et al., 2013; MacNell, Driscoll & Hunt, 2015). In contrary to the above findings, studies also indicate that same-sex preference was restricted to the feminine gender, while male students do not demonstrate any gender preference (Centra & Gaubatz, 2000).

The Cultural Background of the Students

The cultural background of students also influences the outcome of student evaluations of teaching (Davies et al., 2007). The authors observed a significant low rating on the quality of teaching by the female students from the China, Hong Kong, Singapore, and Indian sub-continent, and male students from Singapore and Taiwan. This is significant because they constitute nearly 10 percent of students in all subjects. However, the male and female students from China and UK, studying in the same class had a significantly positive impression on the quality of teaching, but they constitute only two percent of all students in the respective subjects. Whereas a study that investigated the influence of student nationality on SETE ratings reported a different style of feedback by the international students when compared to the local students, and the former gave an average punitive evaluation (Dolnicar & Grun, 2009).

Domain-Specific Vocational Interests

The findings of previous research illustrated that the students' affection towards the course content is associated with a raised SETE score (Dresel & Rindermann, 2011; Marsh & Roche, 1997; Staufenbiel, Seppelfricke & Rickers, 2016). However, the studies do not clarify whether such interest persists before they assume course, or whether such interest was inculcated by a passionate teacher (Marsh & Roche, 1997).

The Personality of the Students

The role of students' personalities on SETE ratings has not explored much. One of the earlier studies indicated that researchers looked for a correlation between the students' mood traits (positive and negative affection) and their mindset at the time of introduction of SETE and the rating obtained for the survey. The results of the study could not establish any correlation between trait variables and SETE ratings. However, it was observed that there exists a constructive association between the students' state of mind at the time of evaluation and the rating of SETE (Munz & Munz, 1997). A study by McCann & Gardner (2014) reported that the personalities of the students were associated with their ratings towards the instructor. It is inferred that the higher agreeableness, lower neuroticism, and higher core self-evaluation were related to higher SETE.

Psychosocial Dynamics

The SETE ratings are also influenced by a range of sub-conscious psychosocial process that interacts between the students and their teachers in the classroom (Bonitz, 2011). Among the ranges of psychological dynamics, the three important phenomena that sought the attention of researchers are the "Halo effect", the significance of initial impressions generated, and the style of delivery. On the perspective of ratings made for SETE, the Halo effect indicates the inclination of students to assess their teachers based on a single feature (for example, whether they like the faculty or not), which then take a broad view to further distinct features (e.g., having familiarity about the course material). The different variables which can persuade the feedback on SETE includes, attractiveness of an instructor (Freng & Webber, 2009; Gurung & Vespia, 2007), his or her inspiring charm (Shevlin et al., 2000), dressing style (Sebastian & Bristow, 2008), and the perceived character of the teacher (Clayson & Sheffet, 2006). Most often, the first impression about a teacher often influences their minds to develop a notion, which leads to a specific judgment, and its influence has been proved recurrently with SETE ratings (Clayson & Sheffet, 2006; Tom, Tong & Hesse, 2010). These studies illustrated that the first impression formed within the first two weeks of teaching hours, or

4

even the impression generated in a few seconds, might retain until the end of a semester can influence the SETE ratings.

Faculty or Teaching Staff Centered Factors

Faculty Gender

Studies demonstrated that, in comparison to the SETE scores of female teachers, the male instructors received higher global SETE scores (MacNell et al., 2015; Smith, 2009), at the same time few studies observed a reverse pattern in scoring (Basow & Montgomery, 2005; Mitchell & Martin, 2018), whereas Smith & Anderson (2005) could not find any logical sex differences. Bastow (2000) studied the characteristics of students to elucidate the relevance of gender perceptions, describing the attributes of their teachers, demonstrated that female teachers were chosen as the best teachers to a better extent by females, but to a lesser extent than expected by the males. Some studies (Korte et al., 2013; Lavin et al., 2012; Young et al., 2009) denote that females are tend to provide positive scores on teacher effectiveness when compared to their counterparts. Also, it is often on different dimensions that capture the established interpersonal relationships between the instructor and the learner: by and large, the women get fostered when they were praised for being considerate, compassionate and approachable, which establishes a mutually supportive and relational classroom environment (Basow, Phelan & Capotosto, 2006; Basow & Montgomery, 2005). The menfolk, contrarily, attract higher SETE ratings on the areas for instance course scheduling, knowledge, proficiency, and management skills (Basow et al., 2006; Smith, 2009). Besides, it is observed that menfolk have been constantly rated on the higher side than the women in physical science education (Basow & Silberg, 1987; Potvin et al., 2009).

Evidence demonstrates that the gender role dynamics between students and teachers play a significant role in SETE ratings. It had been shown that the female teachers who do not conform to the already established conventional womanly gender characteristics, *i.e.,* being caring, complying, pleasant, and relational tend to be judged negatively by the student of both genders (Basow & Montgomery, 2005; Basow et al., 2006). Similarly, the male genders that are not conforming to the traditional masculine features were also perceived negatively by the students (Swaffield, 1996).

Personal Characteristics of the Instructors

It is established that the student assessments inclined towards the professors who demonstrate warmness with a persuading behavior (Best & Addison, 2000). Contrary to this, one study indicated that entertaining lecturers do not unavoidably provide a high level of overall student rating (McKeachie, 1979). However, instructor charisma is one of the factors affecting students' ratings (Shevlin et al., 2000). Instructors who were seeming to be considerate by the students were rated more positively concerning affective class ratings, which depends on how students enjoyed the class.

Further, the student's understanding of cognitive learning is grounded on what the students learned from the lectures, which recommends that the individual features of the instructor may impact SETE (Teven & McCroskey, 1997). A study by Spooren et al., (2013) stated that the instructor-related characteristics would influence SETE in addition to the student-related and course-related characteristics. On the other hand, the personal characteristics of an instructor failed to influence the students' ratings of teaching (Benton & Cashin, 2011).

Another research conducted to study the perception of the students' rating was exclusively based on the student's voice, without the discernible image of both the genders. The result indicated that the younger male professors scored higher rating, which is based on the way they speak and the

tone of voice they use than did young female or old male/female professors (Arbuckle & Williams, 2003).

Mode of Instruction (Online Versus Face-to-Face)

Previous researchers have studied the students' evaluation of professors towards online and face-to-face teaching (Berk, 2013; Carle, 2009; Kelly, Ponton & Rovai, 2007; Marzano & Allen, 2016; Young & Duncan, 2014). Recent studies demonstrated that the students provided lower ratings over online teaching when compared to face-to-face teaching (Marzano & Allen, 2016; Young & Duncan, 2014). Clare (2009) compared the students' evaluation based on the mode of instruction (Online *vs.* face-to-face), and the findings held across several majority/minority-based comparisons (White compared to Black, White compared to Asians, White compared to Hispanic, etc.). It is found that for online lectures, a professor's ethnicity did not affect students' ratings on teaching effectiveness.

Other Faculty Related Variables

Other features of the instructor which influence the respondents are the title, education, and designation of the faculty member (Gokcekus, 2000); however, the duration of teaching experience (Jacobs, 2002) have a meager impact on the overall student rating.

Type, Complexity and Grading of Courses

Course Grades

Previous studies indicated that course grades had been connected with course evaluations (Stumpf & Freedman, 1979). It is also observed that a higher level of course grades resulted in high ratings in teaching evaluations (Greenwald & Gillmore, 1997). Later, McPherson (2006) also indicated that those students who were predicted to have higher grades tend to give significantly better student ratings in both fundamental and higher-level courses. Other studies also supported this notion that there was a significant association between expected course grades and overall ratings of teaching effectiveness (Kozub, 2010; Phipps, Kidd & Latif, 2006). Even though sufficient correlation prevails between high course grades and higher assessments, it is not possible to ascertain that the higher grades are an indication of more learning, preexisting differences, or grading leniency (Marsh & Roche, 2000).

Course Type and Complexity

SETE and its ratings vary based on the type of course, including the academic discipline. Studies indicated that, on average, the teachers who are teaching in the field of arts and social sciences are likely to obtain the highest scores, subsequently the management studies and biological sciences. Whereas, the teachers who teach academic programs like engineering, physical sciences and computer science, are likely to score the lowest ratings (Basow & Montgomery, 2005). In contrary to the usual findings, in some cases, the senior professors who are taking the most demanding courses were able to seek evaluations corresponding to the university means.

In contrast, there are instances where the lecturers teaching fewer challenging courses obtained student ratings that were unusually high (Overbaugh, 1998). However, it is also observed that correlations persist among higher grades with lower workloads and higher teaching evaluations (Greenwald & Gillmore, 1997). A reasonable number of students has an opinion that their teachers

demand less from the students to obtain a satisfactory assessment. Thus, the intricacy of the Course, as perceived by the students, also will affect the students' rating of teaching effectiveness.

Other Factors

Few researchers observed that class size is not a crucial element of student evaluations (Ting, 2000; Williams & Ory, 1992). A study conducted in Australia across four disciplines concluded that the students studying in larger classes provided with lower ratings (Neumann, 2000). Bedard & Kuhn (2005) also reported an exceedingly significant and nonlinear negative effect of class size on SETE ratings.

LIMITATIONS OF STUDENTS EVALUATION SURVEYS

Even though students' surveys have been in practice across the globe, quite a few limitations were also reported (Yorke, 2009). To avoid potential bias, McKeachie (1969) suggested that when the student evaluations are used for faculty appraisals, it is desirable to make use of the student ratings along with other details about the instructor and the course. Emery, Kramer & Tian (2003) also indicated that student ratings are not the only measure of teaching effectiveness, and it does not give evidence in areas linked to teaching effectiveness (for instance, suitability of course content and objectives, and command of subject matter). Though investigators tend to accept that student perception questionnaire should be a factor in teacher evaluations, "there is no agreement on what percentage of an instructor's assessment should come from the results of student survey. There is no sign on the number of times a instructor should be rated by their students every year, and it is a speculation that multiple surveys in one year could give more reliable estimates and also imitate growth during the year.

Furthermore, the instructors are often concerned about the manner at which the SETE data are used by the senior administration, particularly in the context of decision making, as the renewal of contracts and promotions. The studies are abounding with complaints on how the data generated from SETE has been used and the subsequent consequences. The available literature on mismanagement of SETE can be broadly categorized into two: (1) issues related to data interpretation, and (2) the consequences of SETE misuse on the faculty members' job satisfaction and their performance.

Misinterpretation of SETE Data

A common concern raised by the teachers is the lack of knowledge among the administrators to make a psychometrically sound interpretation of SETE data (Algozzine et al., 2004; Wachtel, 1998). Still, many higher education institutions use SETE data alone to assess teaching effectiveness of instructors (Franklin, 2001; Pounder, 2007). Since the ratings of SETE are taken into consideration for high stake decisions, any inappropriate use of SETE data can lead to unnecessary anxiety and tension among the instructors, which ultimately results in low confidence, and lack of job satisfaction (Ory & Ryan, 2001; Wachtel, 1998; Yao & Grady, 2005). To save themselves from building pressure of SETE ratings, some of the instructors also respond to the ponding pressure by generating impressions by compromising the standards of education, unreasonably inflating the results of examinations, or by disregarding SETE feedback altogether (Crumbley & Reichelt, 2009; Wachtel, 1998).

The prevailing competition among higher education institutions leads to a quality race among them, where the student evaluation on teaching effectiveness becomes unavoidable, especially for those universities targeting a sustained quality improvement (Wright, 2008). As a

result, students are more frequently exposed to surveys and evaluations related to courses, program, and other facilities in the College/University that ultimately leads to responder's fatigue in students' surveys.

CONCLUSIONS

Through the critical appraisal of the practice of involving students evaluating teaching effectiveness, this study concludes that HEIs need to be cautious about three essential factors, *i.e.,* Student-centered factors; Faculty centered factors, and Course-centered complexities. Student factors such as gender, cultural background, domain-specific vocational interests, personality-related issues, and psychosocial dynamics influence SETE. Likewise, the faculty related factors include gender, mode of instruction, educational qualification, years of experience, and the position of the faculty member is influencing SETE. Besides, several course-related factors are also influencing students to perceive teaching effectiveness differently, which consists of the grades secured by them, the type & complexity of the course, and class size. The authors concluded that even though SETE has many merits to enhance the quality of higher education, it should be appropriately utilized to extract the full benefits. Further research should focus on studying the influence of these factors on the overall quality of HEIs.

REFERENCES

Al Kuwaiti, A., & Subbarayalu, A.V. (2015). Appraisal of Students Experience Survey (SES) as a measure to manage the quality of higher education in the Kingdom of Saudi Arabia: An institutional study using six sigma model. *Educational Studies*, *41*(4), 430-443

Al. Abbadi, I., Alkhateeb, F., Khanfar, N., Mujtaba, B., & Latif, D. (2009). Pharmacy students perceptions of the teaching evaluation process in Jordan. *Education, Business and Society*, *Contemporary Middle Eastern Issues*, *2*(3), 181-190.

Algozzine, B., Gretes, J., Flowers, C., Howley, L., Beattie, J., Spooner, F., … & Bray, M. (2004). Student evaluation of college teaching: A practice in search of principles. *College Teaching, 52*(4), 134-141.

Al-Rubaish, A. (2010). On the contribution of student experience survey regarding quality management in higher education: An institutional study in Saudi Arabia. *Journal of Service Science and Management*, *3*(4), 464–469.

Al-Rubaish, A., Wosornu, L., & Dwivedi, S.N. (2011). Using deductions from assessment studies towards furtherance of the academic program. *An empirical appraisal of institutional student course evaluation, iBusiness*, *3*(2), 220–228.

Al Rubaish, A. (2011). A comparative appraisal of timings for program evaluation survey and related institutional results in Saudi Arabia: Quality management in higher education. *Journal of Service Science and Management*, *4*, 184–190.

Arbuckle, J., & Williams, B.D. (2003). *Sex Roles*, *49*, 507-516.

Aronson, E., Wilson, T.D., & Akert, R.M. (2007). *Social Psychology, (6th edition)*. Prentice Hall, New Jersey.

Aultman, L.P. (2006). An expected benefit of formative student evaluations. *College Teaching*, *54*(3), 251-285.

Baliyan, S.P., & Moorad, F.R. (2018). Teaching effectiveness in private higher education institutions in Botswana: Analysis of students' perceptions. *International Journal of Higher Education*, *7*(3),143-155.

Basow, S.A., & Silberg, N.T. (1987). Student evaluations of college professors: Are female and male professors rated differently? *Journal of Educational Psychology*, *79*(3), 308- 314.

Basow, S.A. (2000). Gender dynamics in the classroom', in JC Chrisler, C Golden & PD Rozee (edition), *Lectures on the psychology of women (2nd edition)*. McGraw-Hill, New York.

Basow, S.A., & Montgomery, S. (2005). Student ratings and professor self-ratings of college teaching: Effects of gender and divisional affiliation. *Journal of Personnel Evaluation in Education*, *18*(2), 91-106.

Basow, S.A., Phelan, J.E., & Capotosto, L. (2006). Gender patterns in college students' choices of their best and worst professors. *Psychology of Women Quarterly*, *30*(1), 25-35.

Bedard, K., & Kuhn, P. (2005). Where class size really matters: Class size and student ratings of instructor effectiveness. *Economics of Education Review*, *27*(2008), 253–265.

Bennett, S.K., (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. *Journal of Educational Psychology*, *74*(2), 170-179

Benton S.L., & Cashin W.E. (2011). Student ratings of teaching: A summary of the literature'. *IDEA Paper No. 50, The IDEA center, Manhattan, KS.*

Berk, R.A. (2013). Face-to-Face versus Online course evaluations: A "Consumer's Guide" to Seven Strategies. MERLOT *Journal of Online Learning and Teaching*, *9*(1), 140-148.

Best, J.B., & Addison W.E. (2000). A preliminary study of perceived warmth of professor and student evaluations. *Teaching of Psychology*, *27*(1), 60-62.

Bonitz, V.S. (2011). *Student evaluation of teaching: Individual differences and bias effects', Graduate Theses and Dissertations* (12211), Iowa State University, Iowa.

Brandenburg, G.C., & Remmers. H.H. (1927). Rating scale for instructors. *Educational Administration and Supervision*, *13*, 399-406

Carle, A.C. (2009). Evaluating college students' evaluations of a professor's teaching effectiveness across time and instruction mode (online *vs*. face-to-face) using a multilevel growth modeling approach. *Computers & Education*, *53*, 429-435.

Centra, J.A., & Gaubatz, N.B. (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education*, *71*(1), 17-33.

Chisholm, M.E. (2014). Perspectives of faculty on student evaluations of teaching at an Anglophone Caribbean University. *In Adult Education Research Conference 2014 conference proceedings, Harrisburg, PA*.

Clayson, D.E., & Sheffet M.J. (2006). Personality and the student evaluation of teaching. *Journal of Marketing Education*, *28*(2), 149-160.

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

Crumbley, D.L., & Reichelt, K.J. (2009). Teaching effectiveness, impression management, and dysfunctional behavior: Student evaluation of teaching control data. Quality *Assurance in Education. An International Perspective*, *17*(4), 377-392.

Darby, J.A. (2006). Evaluating courses: An examination of the impact of student gender. *Educational Studies*, *32*(2), 187-199.

Davies, M., Hirschberg, J., Lye, J., Johnston, C., & McDonald, I. (2007). Systematic influences on teaching evaluations: The case for caution. *Australian Economic Papers*, *46*(1), 18-38.

Dolnicar, S., & Grun, B. (2009). Response style contamination of student evaluation data. *Journal of Marketing Education*, *31*(2), 160-172.

Douglass, H., (1928). Rating the teaching effectiveness of college instructors. *School and Society, 28*, 192-197.

Dresel, M., & Rindermann, H., (2011). Counseling university instructors based on student evaluations of their teaching effectiveness: A multilevel test of its effectiveness under consideration of bias and unfairness variables. *Research in Higher Education*, *52*, 717–737.

Emery, C.R., Kramer, T.R., & Tian, R.G. (2003). Return to academic standards: A critique of student evaluations of teaching effectiveness. *Quality Assurance in Education*, *11*(1), 37-46.

Feistauer, D., & Richter, T. (2018). Validity of students' evaluations of teaching: Biasing effects of likability and prior subject interest. *Studies in Educational Evaluation*, *59*, 168-178.

Freng, S., & Webber, D. (2009). Turning up the heat on online teaching evaluations: Does "hotness" matter? *Teaching of Psychology*, *36*(3), 189-193.

Gokcekus, O. (2000). How do university students value economics courses? A hedonic approach. *Applied Economics Letters, 7*, 493–496.

Greenwald, A.G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, *52*(11), 1182-1186.

Greenwald, A.G., & Gillmore, G.M. (1997), Grading Leniency is a removable contaminant of student ratings. *American Psychologist*, *52*(11), 1209–1217.

Gurung, R.A.R., & Vespia, K.M. (2007). Looking good, teaching well? Linking liking, looks, and learning. *Teaching of Psychology*, *34*(1), 5-10.

Guthrie, E.R. (1927). Measuring student opinion of teachers. *School and Society*, *25*, 175-176.

Hejase, A.J., Al Kaakour, R.S., Halawi, L.A., & Hejase, H.J. (2013). Students' perceptions of Student Evaluation of Teaching (SET) process. *International Journal of Social Sciences and Education*, *3*(3), 565-575.

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

Jacobs, L.C. (2002). Student ratings of college teaching: What research has to say.

Kelly, H.F., Ponton, M.K., & Rovai, A.P. (2007). A comparison of student evaluations of teaching between online and face-to-face courses. *The Internet and Higher Education*, *10*(2), 89-101.

Korte, L., Lavin, A., & Davies, T. (2013). Does gender impact business students' perceptions of teaching effectiveness? *Journal of College Teaching & Learning*, *10*(3), 167-178.

9

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

Lavin, A., Korte, L., & Davies, T. (2012). Student gender and perceptions of teaching effectiveness. *Research in Higher Education Journal*, *18*, 1-16.

MacNell, L., Driscoll, A., & Hunt, A.N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, *40*(4), 291–303.

Marsh, H.W. (1977). The validity of students' evaluations: Classroom evaluations of instructors independently nominated as best and worst teachers by graduating seniors. *American Educational Research Journal*, *14*(4), 441-447.

Marsh, H.W. (1980). The influence of student, Course, and instructor characteristics in evaluations of university teaching. *American Educational Research Journal*, *17*(2), 219-237.

Marsh, H.W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, *76*(5), 707-754

Marsh, H.W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, *11*(2), 253-388.

Marsh, H.W., & Hocevar, D. (1991). Students' evaluations of teaching effectiveness: The stability of mean ratings of the same teachers over a 13-year period. *Teaching and Teacher Education*, *7*(4), 303-314.

Marsh, H.W., & Dunkin, M.J. (1992). *Students' evaluations of university teaching: A multidimensional perspective', in JC Smart (ed.)*. Higher education: Handbook of theory and research, *8*, Agathon Press, New York

Marsh, H.W., & Roche, L.A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, *52*(11), 1187-1197.

Marsh, H.W., & Roche, L.A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myths, bias, validity, or innocent bystanders?.*Journal of Educational Psychology*, *92*(1), 202-228.

Mart, C.T. (2017). Student evaluations of teaching effectiveness in higher education. *International Journal of Academic Research in Business and Social Sciences*, *7*(10), 57-61.

Marzano, M.P., & Allen, R. (2016). Online *vs*. Face-to-Face course evaluations: Considerations for administrators and faculty. *Online Journal of Distance Learning Administration*, *19*(4).

McCann, S., & Gardner, C. (2014). Student personality differences are related to their responses on instructor evaluation forms. *Assessment & Evaluation in Higher Education, 39*(4), 412-426.

McKeachie, W.J. (1969). Student ratings of faculty. *AAUP Bulletin, 55*, 439-444.

McKeachie, W.J. (1979). Student Ratings of Faculty: A Reprise'. *Academe*, *65*, 384–397.

McPherson, M.A. (2006). Determinants of how students evaluate teachers. *The Journal of Economic Education*, *37(*1), 3–20.

Mitchell, K.M.W., & Martin, J. (2018). Gender bias in student evaluations. *PS: Political Science & Politics, 51*(3), 648-652.

Moore, S., & Kuol, N. (2005). Students evaluating teachers: Exploring the importance of faculty reaction to feedback on teaching. *Teaching in Higher Education, 10*(1), 57–73.

Munz, D.C., & Munz, H.E. (1997). Student mood and teaching evaluations. *Journal of Social Behavior and Personality, 12*(1), 233-242.

Murray, J.H. (1997). *Hamlet on the Holodeck: The future of narrative in cyberspace*. The Free Press, New York.

Neumann, R. (2000). Communicating student evaluation of teaching results: Rating Interpretation Guides (RIGs). *Assessment & Evaluation in Higher Education, 25*(2), 121-134.

Ory, J.C., & Ryan, K. (2001). How do student ratings measure up to a new validity framework? In M Theall, PC Abrami & L Mets (eds.). *The student ratings debate: Are they valid? How can we best use them?* Jossey-Bass, San Francisco.

Overbaugh, R.C. (1998). The effect of course rigor on preservice teachers' course and instructor evaluation. *Computers in the Schools*, *14*, 13-23.

Phipps, S.D., Kidd, R.S., & Latif, D.A. (2006). Relationships among student evaluations, instructor effectiveness, and academic performance. *Pharmacy Education, 6*, 237–243.

Potvin, G., Hazari, Z., Tai, R.H., & Sadler, P.M. (2009). Unraveling bias from student evaluations of their high school science teachers. *Science Education, 93*(5), 827-845.

Pounder, J.S. (2007). Is student evaluation of teaching worthwhile? An analytical framework for answering the question. *Quality Assurance in Education: An International Perspective*, *15*(2), 178-191.

Pounder, J.S. (2008). Transformational classroom leadership: A novel approach to evaluating classroom performance. *Assessment and Evaluation in Higher Education, 33*(3), 233-243.

Remmers, H.H., & Brandenburg, G.C. (1927). Experimental data on the Purdue rating scale for instructors. *Educational Administration and Supervision, 13*, 519-527.

Remmers, H.H. (1928). The relationship between students' marks and students' attitudes toward instructors. *School and Society, 28*, 759-760.

Remmers, H.H. (1930). To what extent do grades influence student ratings of instructors? *Journal of Educational Research, 21*, 314-316.

Rowan, S., Newness, E.J., Tetradis, S., Prasad, J.L., Ko, C.C., & Sanchez, A. (2017). Should student evaluation of teaching play a significant role in the formal assessment of dental faculty? Two viewpoints: Viewpoint 1: Formal faculty assessment should include student evaluation of teaching and Viewpoint 2: student evaluation of teaching should not be part of formal faculty assessment. *Journal of dental education*, *81*(11), 1362–1372.

Sebastian, R.J., & Bristow, D. (2008). Formal or informal? The impact of style of dress and forms of address on business students' perceptions of professors. *Journal of Education for Business, 83*(4), 196-201.

Seldin, P. (1993). How colleges evaluate professors: 1983 Versus 1993. *AAHE Bulletin, 12*, 6-8.

Shevlin, M., Banyard, P., Davies, M., & Griffiths, M. (2000). The validity of student evaluation of teaching in higher education: Love me, love my lectures? *Assessment & Evaluation in Higher Education, 25*(4), 397-405.

Slocombe, T., Miller, D., & Hite, N. (2011). A survey of student perspectives toward faculty evaluations. *American Journal of Business Education, 4*(7), 51–57.

Smalzried, N.T., & Remmers, H.H. (1943). A factor analysis of the Purdue rating scale for instructors. *Journal of Educational Psychology, 34*, 363-367.

Smith, B.P. (2009). Student ratings of teaching effectiveness for faculty groups based on race and gender. *Education, 129*(4), 615-624.

Smith, G., & Anderson K.J. (2005). Students' ratings of professors: The teaching style contingency for Latino/a professors. *Journal of Latinos & Education, 4*(2), 115-136

Sojka, J., Gupta, A.K., & Deeter-schmelz, D.R. (2002). Student and faculty perceptions of student evaluations of teaching: A study of similarities and differences. *College Teaching, 50*(2), 44-49.

Spooren, P., Brockx, B., & Mortelman, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83*(4), 598-642.

Staufenbiel, T., Seppelfricke, T., & Rickers, J. (2016). Pradiktoren student is cher Lehrveranstaltungs evaluation en: Eine Mehrebenen analyse [Predictors of student evaluations of teaching: A multilevel analysis]. *Diagnostica, 62*, 44–59.

Stumpf, S.A., & Freedman, R.D. (1979). Expected grade covariation with student ratings of instruction: Individual versus class effects. *Journal of Educational Psychology*, *71*, 293–302.

Swaffield, B.C. (1996). What happens when male professors enact feminist pedagogies?'

Teven, J.J., & McCroskey, J.C. (1997). The relationship of perceived teacher caring with student learning and teacher evaluation. *Communication Education, 46*, 1-9.

Ting, K. (2000). A multilevel perspective on student ratings of instruction: Lessons from the Chinese experience. *Research in Higher Education*, *41*, 637–661.

Tom, G., Tong, S.T., & Hesse C. (2010). Thick slice and thin slice teaching evaluations. *Social Psychology of Education, 13*(1), 129-136.

Umar, M.A., Ahmad, B.I., Kufena, A.M., Abdulsalami, A.O., Tenuche, S.S., Sahabi, Y.A., & Ahmad, U.M. (2016). Students' perception of online Student Evaluation of Teaching (SET) in Nigeria. Paper presented at the International Conference on Information and Communication Technology and Its Applications (ICTA), 28-30 November, (23-27), Federal University of Technology, Minna, Nigeria.

Vijay, S.A. (2013). Appraisal of student rating as a measure to manage the quality of higher education in India: An institutional study using six sigma model approach. *International Journal of Quality Research, 7*(3), 3-14.

Wachtel, H.K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education, 23*(2), 191-211.

Williams, R., & Ory, J.C. (1992). A further look at class size, discipline differences and student ratings. Urbana-Champaign: Office of Instructional Resources, University of Illinois at Urbana-Champaign.

Wright, R.E. (2008). Student evaluations of faculty: Concerns raised in the literature, and possible solutions. *College Student Journal, 40*(2), 417-422.

Yahya, S.I., & Rostam, D.M. *(2015).* A Sustainable Paperless Online System (SPOS) for engineering quality in teaching: Koya University as a case Study. *The Scientific Journal of Koya University, 3*(1), 1-13

Yao, Y., & Grady, M.L. (2005). How do faculty make formative use of student evaluation feedback? A multiple case study. *Journal of Personnel Evaluation in Education, 18*(2), 107-126.

Yorke, M. (2009). Student experience' surveys: Some methodological considerations and an empirical investigation. *Assessment & Evaluation in Higher Education, 34(*6), 721–739.

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).

Young, S., & Duncan, H.E. (2014). Online and face-to-face teaching: How do student ratings differ? *MERLOT Journal of Online Learning and Teaching, 10*(1), 70-79.