

A literature review on the student evaluation of teaching

An examination of the search, experience, and credence qualities of SET

Gregory Ching

*Graduate Institute of Educational Leadership and Development,
Fu Jen Catholic University, New Taipei City, Taiwan*

Student
evaluation of
teaching

63

Received 29 April 2018
Revised 14 June 2018
10 October 2018
Accepted 19 November 2018

Abstract

Purpose – Competition among higher education institutions has pushed universities to expand their competitive advantages. Based on the assumption that the core functions of universities are academic, understanding the teaching–learning process with the help of student evaluation of teaching (SET) would seem to be a logical solution in increasing competitiveness. The paper aims to discuss these issues.

Design/methodology/approach – The current paper presents a narrative literature review examining how SETs work within the concept of service marketing, focusing specifically on the search, experience, and credence qualities of the provider. A review of the various factors that affect the collection of SETs is also included.

Findings – Relevant findings show the influence of students' prior expectations on SET ratings. Therefore, teachers are advised to establish a psychological contract with the students at the start of the semester. Such an agreement should be negotiated, setting out the potential benefits of undertaking the course and a clear definition of acceptable performance within the class. Moreover, connections should be made between courses and subjects in order to provide an overall view of the entire program together with future career pathways.

Originality/value – Given the complex factors affecting SETs and the antecedents involved, there appears to be no single perfect tool to adequately reflect what is happening in the classroom. As different SETs may be needed for different courses and subjects, options such as faculty self-evaluation and peer-evaluation might be considered to augment current SETs.

Keywords Higher education, Student expectations, Service marketing, Teacher evaluation, Teaching and learning process

Paper type Research paper

1. Introduction

For the past number of years, the increasing number of degree providing institutions has dramatically changed global higher education (Altbach *et al.*, 2009; Usher, 2009). This rising number of higher education institutions has actually led to increased competition among universities (Naidoo, 2016). Furthermore, with cutbacks in government funding for higher education (Mitchell *et al.*, 2016), differentiation is essential for universities to distinguish themselves and compete with other institutions (Staley and Trinkle, 2011). Such differentiation of higher education institutions has become commonplace, forcing universities to become more innovative, cost conscious, and entrepreneurial (Longanecker, 2016; MacGregor, 2015).

These global dilemmas are not new to Taiwan, wherein universities have to outperform each other for financial subsidies, while also competing to recruit new students (Chou and Ching, 2012). The problem of recruitment results from a serious decline of birth rate in Taiwan. The National Statistics Office of Taiwan (2018) reported that birth figures declined



© Gregory Ching. Published in *Higher Education Evaluation and Development*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

Higher Education Evaluation and
Development
Vol. 12 No. 2, 2018
pp. 63-84
Emerald Publishing Limited
2514-5789
DOI 10.1108/HEED-04-2018-0009

from 346,208 in 1985 to 166,886 in 2010, representing a fall of almost 50 percent. Projecting these numbers into university entrants, a drop of around 20,000 incoming students can be noted for the academic year 2016/2017 (Chang, 2014). In fact, only 241,000 freshmen students are noted for the current 2017/2018 academic year and this number is expected to drop to around only 157,000 in 2028 (Wu, 2018). This issue of declining number of students has resulted in financial difficulties for academic institutions (Chen and Chang, 2010). In such difficult times, it is crucial for higher education institutions in Taiwan to differentiate themselves and develop their competitive advantages.

In the age of big data, differentiation can be achieved with the help of large data sets that provide institutions with the capacity to address complex institutional issues (Daniel, 2015; Norris and Baer, 2013). Many researchers have begun to collect and analyze institutional data sets to address various administrative and instructional issues faced by the universities (Picciano, 2012). The results of these studies can provide school administrators and students with useful information (Castleman, 2016). In Taiwan, big data has provided institutions with information on topics such as trends in enrollment rates, students' online learning performances, and research outputs measured by number of academic publications (Tseng, 2016). Another study reported on the advantages of collecting and understanding student learning experiences using big data (Lin and Chen, 2016). Based on the assumption that the core functions of higher education institutions remains to be academic (Altbach, 2011), i.e., teaching and learning, determining and understanding the quality of the teaching learning process with the aid of big data can be extremely useful.

In order to understand the quality of the teaching learning process, higher education institutions in Taiwan and elsewhere have long been using the student evaluation of teaching (SET), which provides feedback on teaching performance and appraises faculty members (Aleamoni and Hexner, 1980; Centra, 1979; Clayson, 2009; Curran and Rosen, 2006; Pozo-Muñoz *et al.*, 2000). Even though the practice of using SETs is well established in higher education institutions (Rice, 1988; Wachtel, 2006) and is considered relatively reliable for evaluating courses and instructors (Aleamoni, 1999; Marsh, 1987, 2007; Nasser and Fresko, 2002), its usefulness and effectiveness has been challenged (Boring *et al.*, 2016).

Over time, numerous issues have arisen in research on SETs. It has been claimed that SETs are used as a tool by students to reward or punish their instructor (Clayson *et al.*, 2006), that SET results differ across areas (course, subject, and discipline) (Chen, 2006) and type (including course design and class size) of study (Feldman, 1978; Marsh, 1980), and that completion rate and background demographics of students significantly affect SETs (Stark and Freishtat, 2014). Moreover, SETs can be biased with respect to the gender of the instructor and that of the students (Boring *et al.*, 2016). Interestingly, recent research has found that effective teachers are receiving low SET ratings (Braga *et al.*, 2014; Kornell and Hausman, 2016). This has caused many institutions, including universities in Taiwan, to analyze and redesign their SETs (Chen, 2006; Zhang, 2003).

In light of these issues, the current paper shall seek to provide a better understanding of the inner workings of SETs. With the better understanding of SETs, more appropriate and effective evaluation tools can be developed. In addition, the categorization of education as a type of services (WTO, 1998) has also opened up new ways of looking into the entire academe. Anchoring on the narrative literature review paradigm, this paper will shape the discussion of SETs within the concept of service marketing. A common framework used to evaluate services, which is to determine the search, experience, and credence qualities of the provider (Fisk *et al.*, 2014, p. 151; Wilson *et al.*, 2012, p. 29). In addition, the paper will review the definitions of SET in the existing literature as well as the dimensions commonly used to measure the quality of teaching. Finally, the various factors that affects the collection of SETs are discussed.

2. Methodology

The current study is anchored on a literature review paradigm. For any study, a literature review is an integral part of the entire process (Fink, 2005; Hart, 1998; Machi and McEvoy, 2016). In general, literature reviews involve database retrievals and searches defined by a specific topic (Rother, 2007). To perform a comprehensive literature review, researchers adopt various approaches for organizing and synthesizing information, adopting either a qualitative or quantitative perspective for data interpretation (Baumeister and Leary, 1997; Cronin *et al.*, 2008; Fink, 2005; Hart, 1998; Lipsey and Wilson, 2001; Petticrew and Roberts, 2005; Rocco and Plakhotnik, 2009; Torraco, 2005).

For the current study, the researcher adopts a narrative literature review approach. Narrative review or more commonly refer to as traditional literature review is a comprehensive, critical, and objective analysis of the current knowledge on a topic (Charles Stuart University Library, 2018). The review should be objective insofar as it should have a specific focus but should provide critiques of important issues (Dudovskiy, 2018). More importantly, the results of a narrative review are qualitative in nature (Rother, 2007).

The study follows the suggestion of Green *et al.* (2006) with regard to synthesizing search results retrieved from computer databases. For the current study, the researcher used Google Scholar as a starting point, followed by searches within ProQuest and PsycINFO. Keywords used for searches were “student evaluation of teaching” and related terminologies (see next section for more information on SET synonymous terms). Selections of relevant articles are explicit and potentially biased insofar as the researcher focuses on the search, experience, and credence qualities of providers within SET studies. Data analysis methods consist of a procedure for organizing information into specific themes developed by Miles and Huberman (1994) and Glaser’s (1965, 1978) technique for continuous comparison of previously gathered data.

3. Defining student evaluation of teaching

In relation to students’ college experience, determining whether a course or a teacher is good or bad can be equated to measuring service quality (Curran and Rosen, 2006). This is especially the case with regard to SETs. The concepts behind SETs have been discussed since the early 1920s (Otani *et al.*, 2012), and literally thousands of studies have been carried out on these various interrelated concepts (Marsh, 2007). Furthermore, within the vast spectrum of literature on the topic, a variety of related terms are used interchangeably. Hence, a thorough, comprehensive literature review is impossible.

SET is a relatively recent term that is used synonymously with several previous terminologies such as Student Evaluation Of Educational Quality (SEEQ) (Coffey and Gibbs, 2001; Grammatikopoulos *et al.*, 2015; Lidice and Saglam, 2013), SET effectiveness (Marsh, 1987, 2007), student evaluation of teacher performance (Chuah and Hill, 2004; Coburn, 1984; Flood, 1970; Poonyakanok *et al.*, 1986), student evaluation of instruction (Aleamoni, 1974; Aleamoni and Hexner, 1980; Clayson *et al.*, 2006; Powell, 1977), student course satisfaction (Betoret, 2007; Bolliger, 2004; Rivera and Rice, 2002), or just simply student course evaluation (Anderson *et al.*, 2005; Babcock, 2010; Bembenuddy, 2009; Chen, 2016; Duggan and Carlson-Bancroft, 2016; Huynh, 2015; Pravikoff and Nadasen, 2015; Stark and Freishtat, 2014). Despite the difference in terms, the core objectives of all of the above are similar.

According to the classification and definition of basic higher education terms set by the United Nations Educational, Scientific and Cultural Organization, SET is described as:

[...] the process of using student inputs concerning the general activity and attitude of teachers. These observations allow the overall assessors to determine the degree of conformability between student expectations and the actual teaching approaches of teachers. Student evaluations are

expected to offer insights regarding the attitude in class of a teacher and/or the abilities of a teacher [...] (Vlăsceanu *et al.*, 2004, pp. 59-60).

This definition implies three main aspects, namely, the evaluation of the teacher (the teacher itself), the teaching process (general activity and teaching approaches), and the learning outcomes as perceived by the students (student expectations) This is similar to the framework for evaluating service marketing, whereby the teacher corresponds to the “search” qualities, the teaching process to the “experience” qualities, and the learning outcomes to the “credence” qualities (see Figure 1).

3.1 Search qualities in SET

As previously mentioned, one of the first aspects of SET that focuses on the teacher is the student evaluation of the teacher, or rather the student’s perception of the teacher’s characteristics (Fox *et al.*, 1983). As Tagiuri (1969) notes, in an early study, a person’s (in this case a teacher’s) personality, characteristics, qualities, and inner states (p. 395) matter significantly. Early research findings suggest that students sometimes interpret a teacher’s creativeness as a positive characteristic (Costin *et al.*, 1971), while others note that a teacher’s personality traits affect their SET ratings (Clayson and Sheffet, 2006; Mogan and Knox, 1987; Murray *et al.*, 1990). For instance, the interpersonal characteristics of teachers influence interactions between the students (Mogan and Knox, 1987), which ultimately leads to better engagement and learning (Hu and Ching, 2012; Hu *et al.*, 2015; Skinner and Belmont, 1993).

This focus on the teacher also leads to various biases in SET. For example, teachers’ physical appearance can have an effect on their SET ratings (Bonds-Raacke and Raacke, 2007; Hultman and Oghazi, 2008). Felton *et al.* (2004) in their study of the university teachers rating website (www.ratemyprofessors.com/) conclude after analyzing 3,190 faculty members across 65,678 posts that physically attractive teachers get higher ratings. In addition, a study by Buck and Tiene (1989) finds that attractive female teachers, even if they are considered authoritarian, tend to receive higher SET ratings compared to their less attractive female counterparts. Besides the teacher’s physical appearance, gender, and age are also important (Buck and Tiene, 1989; Sohr-Preston *et al.*, 2016). Younger male faculty members were found to receive higher ratings (Boring *et al.*, 2016), while more senior teachers received lower SET ratings (Clayson, 1999). Similarly, a teacher’s ethnicity is also a factor (Dee, 2005; Ehrenberg *et al.*, 1995). For instance, students may consider female African American teachers more sympathetic (Patton, 1999), which can affect their SET ratings. These biases in SETs are unfair since an individual’s demographics and personality traits are fixed and cannot be changed.

Drawing on the concept of service marketing, the aforementioned teacher factors can be considered the search qualities that students look for before enrolling in a particular course. Students sometimes look for easy teachers just to pass a subject (Felton *et al.*, 2004).

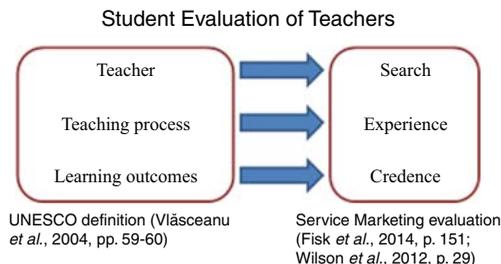


Figure 1.
SET concepts

However, research shows that most students tend to search for competent teachers (Feldman, 1984) and credible faculty members (Patton, 1999; Pogue and Ahyun, 2006). This disproves the fallacy that easy teachers receive high SET ratings (Beatty and Zahn, 1990; Marsh and Roche, 2000).

By definition, search qualities are the easily observable and most common physical attributes a product (or in this case a teacher or course) may possess (Curran and Rosen, 2006, p. 136). Moreover, these search qualities are apparent and can be judged relative to similar options (Lubienski, 2007). What is most important is that students are able to judge these search qualities beforehand. This means that students have certain initial preferences with regard to aspects such as the type of teacher, the physical characteristics of the classroom, or even the schedule of the course. Students tend to compare various options before signing up for a class. In addition, social psychology literature has long demonstrated the influence of beauty on individual judgments (Adams, 1977; Berscheid and Walster, 1974). Individuals tend to relate beauty to being good (Eagly *et al.*, 1991). This halo effect explains why teachers' attractiveness tends to influence their SET ratings (Felton *et al.*, 2004). Furthermore, students may also have a preference with regard to the physical situation of the classroom (Douglas and Gifford, 2001; Hill and Epps, 2010), which influences their overall level of satisfaction.

In summary, more emphasis should be placed on the perceived expectations of students, which can be discerned from their search qualities. As studies by Buck and Tiene (1989) and Patton (1999) find, students tend to associate physical attributes with certain types of behavior, such as expecting attractive female teachers to be more feminine and female African American teachers to be more sympathetic. Another important issue here is that students are expecting something, regardless of their reasons for having these expectations of their teachers and courses. These expectations, whether arising from stereotyping of attributes or hearsay from schoolmates, must be met to satisfy the students. However, this should not be the case, and teachers should focus on building their professionalism and credibility (Carter, 2016). In-class behaviors such as self-disclosure, humor, warmth, clarity, enthusiasm, verbal and nonverbal messages, teacher immediacy; nonverbal interactions that enhance closeness (Mehrabian, 1968, p. 203), and affinity seeking; the creation of positive feelings toward oneself (Bell and Daly, 1984) are just a few examples of effective strategies that can positively affect how students view teachers (Patton, 1999). These behaviors make for effective teaching and can also prevent students from stereotyping teachers because of their appearance or on account of demographic features.

3.2 *Experience qualities in SET*

Besides the teacher, the second aspect of SET identified is the teaching process. In reality, this is what the majority of SETs currently being used measures (Algozzine *et al.*, 2004; Wachtel, 2006). The goal of SETs is to determine the teachers' teaching effectiveness (Marsh, 2007). Such instruments have been used throughout academia for a long time, but their validity, reliability, and usefulness are still being challenged (Aleamoni, 1974, 1999; Aleamoni and Hexner, 1980; Arreola, 2007; Costin *et al.*, 1971; Feldman, 1978, 1984; Marsh, 1984, 2007; Marsh and Roche, 1997; Rodin and Rodin, 1972; Wright *et al.*, 1984). This makes sense since teaching is a complex activity (Shulman, 1987), so the factors used to measure a teacher's effectiveness are multidimensional (Marsh, 1991, 2007; Marsh and Bailey, 1993) and difficult to comprehend. Nonetheless, it has been proven that SETs contribute to faculty development by enhancing the teaching and learning experience (Centra, 1979; Marsh, 1980, 1984, 1987, 1991, 2007; Marsh and Bailey, 1993; Marsh and Roche, 1997). This function is formative (Shulman, 1987), meaning that SETs can provide evidence to support improvements that shape the overall quality of teaching (Berk, 2005, p. 48).

Evaluating the teaching process is a complex and complicated undertaking, requiring a full understanding of how students came to conclusions with regard to their teachers and courses (Curran and Rosen, 2006). Typically, taking a university course would require the student to attend class every week, which corresponds to repeated service encounters that are critical to later evaluation (Solomon *et al.*, 1985). Within the concept of service marketing, these repeated service encounters (which in this case are the repeated classroom encounters) correspond to the experience qualities that students perceive when taking the course. These experience qualities are not readily apparent and can only be judged when the entire service experience is over (generally at the end of the course) (Curran and Rosen, 2006; Lubienski, 2007). However, because such experiences are repeated, it can be difficult to know whether the resulting SET ratings are based on an overall experience of the course or just on one single event that made a lasting impression (Curran and Rosen, 2006). Furthermore, students attend class with their classmates, so there are other individuals partaking of the service delivery at the same time. Therefore, the interactions of these students within the class might either enhance or reduce the service quality, which might, in turn, affect an individual student's experience (Grove and Fisk, 1997).

Based on the above points, evidence shows that students can compare their teachers with other teachers teaching the same course before actually signing up for the class. However, it is most likely that student would ask around, seeking other students who had taken the course already and asking for their comments. This is because students generally do not have access to SET results (Marlin, 1987). Marsh (2007) notes that although a few universities do publish their SET summaries, this is solely for the purpose of course or subject selection. The publication of SET results is controversial (Babad *et al.*, 1999; Perry *et al.*, 1979) and is generally regarded negatively by faculty members (Howell and Symbaluk, 2001).

It is important to note that, based on asking around prior to taking a course, students might expect to receive a certain grade or a certain amount of classwork, or even have expectations with regard to how the lectures are conducted (Nowell, 2007; Remedios and Lieberman, 2008; Sander *et al.*, 2000; Voss *et al.*, 2007). If teachers then behave contrary to the students expectations, students may be disappointed and SET ratings may be affected (Bennett, 1982). Such student expectations can also contribute to the development of a psychological contract between the teacher and the students. These prior expectations, whether arising from the students' desire to benefit from the course (Voss *et al.*, 2007) or from hearsay, are found to contribute to such a psychological contract (Roehling, 1997).

A psychological contract can be defined as any individual beliefs, shaped by the institution, regarding the terms of an exchange agreement, between students and their teachers (Kolb, Rubin, and McIntyre, 1984; Rousseau, 1995, 2001). Recent research finds that when the psychological contract between the teacher and the students is positive, learning motivation is enhanced (Liao, 2013). Furthermore, these agreements might be either implicitly or explicitly made between the teachers and students. To make them more effective, the agreements should be negotiated at the start of the term, and should constitute a shared psychological contract between the teacher and the students (Pietersen, 2014). More importantly, Cohen (1980) notes that if SETs are accomplished during the middle of the semester, teachers are still able to improve their teaching pedagogy re-aligning the previously agreed upon psychological contract. Hence, faculty members who received mid-semester feedback ended up with significantly higher SET ratings than their counterparts who did not have a mid-semester evaluation (Cohen, 1980). Ultimately, mid-semester feedbacks provide ample opportunity for instructional improvement (Overall and Marsh, 1979).

In summary, it has been noted in the literature that evaluating the teaching process, or rather the effectiveness of teaching, is a complex task. It is multidimensional and mostly concerns the experience qualities of the students who have taken the course. More important, in relation to the numerous biases associated with SETs discussed in the

introduction of this paper, perceptions of course delivery and service quality are affected by a variety of issues, including peers, class size, and type of course. Adding to the fact that students have their own personal expectations of what the course should offer, it is difficult to satisfy every student. Pietersen (2014) suggests the making of a psychological contract between the teacher and the students, to provide clear study goals and remove false expectations. In addition, an evaluation can be conducted in the middle of a semester, giving the teacher ample opportunity to address students' doubts and to re-adjust the shared contract based on students' abilities. Furthermore, as the goal of SETs is to provide formative suggestions for teachers to improve teaching, it is also prudent to include statements on the provision of formative lessons and on how course designs contribute to student learning (Brownell and Swaner, 2009; Kuh, 2008; Kuh *et al.*, 2013).

3.3 *Credence qualities and SETs*

The last component of SETs identified is the evaluation of learning outcomes, more specifically, the accomplishment of goals. It has long been accepted that goals are important predictors of educationally relevant learning outcomes (Ames and Archer, 1988), while also focusing on the motivational aspects driven by mastery and performance-approach goals (Harackiewicz *et al.*, 2002). In simple terms, if students clearly understand the skills necessary for future employment, while also understanding taking a certain course will enable them to master those skills, they should be motivated to do well in that course. Research shows that students are more engaged with their academic classwork when future career consequences are clearly understood (Greene *et al.*, 2004; Miller *et al.*, 1996). However, in reality many students are uncertain of their study goals and are at risk of dropping out (Mäkinen *et al.*, 2004).

A university education is characterized by high credence properties (Curran and Rosen, 2006). Credence qualities are those properties that are not apparent, can be never be fully known or appreciated by students (Lubienski, 2007), and might, therefore, be impossible to evaluate (Curran and Rosen, 2006). Credence properties are generally found in goods and services that are characterized by high levels of complexity (Darby and Karni, 1973), such as the teaching and learning process. More importantly, even after the service has been used (in this case, when a student graduate from the university), the consumer (student) may still find it difficult to evaluate the service (Zeithaml, 1981). A famous example of credence qualities in a product can be found in the taking of vitamin pills, where there is low verification of the alleged effectiveness and quality of the product, even after it has been tried by the consumer (Galetzka *et al.*, 2006). In higher education, the true value of a course may be realized by a student only after the skills and knowledge learned are used in employment, which might be several months or even years after the service has ceased (Curran and Rosen, 2006).

The credence qualities of higher education are related to the concept of academic delay of gratification (Bembenutty, 2009). Academic delay of gratification is a term used to describe the "postponement of immediately available opportunities to satisfy impulses in favor of pursuing chosen important academic rewards or goals that are temporally remote but ostensibly more valuable" (Bembenutty and Karabenick, 1998, p. 329). Similar to what is described by the achievement goals theory, students are motivated when they clearly perceive benefits that lead to future success (Bembenutty, 1999). In addition, students who adhere to the academic delay of gratitude principle tend to become autonomous learners (Bembenutty and Karabenick, 2004). If students know the usefulness of the course subject, they are more willing to delay gratification, participate in class, and complete academic tasks, and are ultimately more satisfied and hence give high SET ratings (Bembenutty, 2009).

In summary, the literature shows that besides formative evaluations, SETs also include summative evaluations (Kuzmanovic *et al.*, 2012; Mortelmans and Spooren, 2009; Otani *et al.*, 2012; Spooren and Van Loon, 2012), which involve summing up the overall performance of teachers (Berk, 2005). Summative SETs generally contribute to teacher audits and evaluations that may lead to the hiring, tenure, and even promotion of faculty members (Arthur, 2009; Berk, 2005; Marks, 2000; Stark and Freishtat, 2014). Literature suggests that school administrators should be careful in using SET results containing many summative evaluations (Spooren *et al.*, 2013), because, with respect to the credence properties of education, students might be unable to grasp the entire and actual benefits of certain courses. In order for effective learning to occur, the potential benefits of the course and an outline of acceptable performance should be defined in advance (Otter, 1995). Moreover, connections should be made between previous, current, and future courses, thus providing an overview of the entire program together with a clear outline of future career pathways.

4. Dimensions of SET

As has been noted, SETs are complex and involves multiple interrelated dimensions. In his early meta-analysis, Feldman (1978) shows that although most studies focus on the overall rating of the instructor. However, SETs that focus only on summative evaluations and that use global measures (few summary items) are highly discouraged (Cashin and Downey, 1992; Marks, 2000; Sproule, 2000). The majority of SETs aim at a more comprehensive rating of teachers and, as Marsh (2007) notes, are mostly constructed around the concept of effective teaching. The usefulness and effectiveness of an SET depends on how well it can capture the concepts it measures. Hence, careful design is essential (Aleamoni, 1974, 1999; Aleamoni and Hexner, 1980; Arreola, 2007).

One of the early syntheses of SETs is conducted by analyzing students' views of the characteristics of a superior teacher (Feldman, 1976). For the study, three categories are identified: presentation, which includes teachers' enthusiasm for teaching and for the subject matter, their ability to motivate students', their knowledge of the subject matter, clarity of presentation, and organization of the course; facilitation, which denotes teachers' availability for consultation (helpfulness), their ability to show concern and respect for students (friendliness), and their capacity to encourage learners through class interactions and discussions (openness); and regulation, which includes the teachers' ability to set clear objectives and requirements, appropriateness of course materials (including supplementary learning resources) and coursework (with regard to difficulty and workload), fairness in evaluating students and providing feedback, and classroom management skills (Feldman, 1976).

Another early analysis of SETs conducted by Hildebrand (1973) and Hildebrand *et al.* (1971) and his associates identifies five constructs for measuring the effectiveness of teaching: analytic/synthetic skills, which includes the depth of a teacher's scholarship and his or her analytic ability and conceptual understanding of the course content; organization/clarity, denoting the teacher's presentation skills in the course subject area; instructor group interaction, which describes the teacher's ability to actively interact with the class, his or overall rapport with the class, sensitivity to class responses, and ability to maintain active class participation; instructor-individual student interaction, which includes the teacher's ability to establish mutual respect and rapport with individual students; and dynamism/enthusiasm, which relates to the teacher's enthusiasm for teaching and includes confidence, excitement about the subject, and pleasure in teaching (Hildebrand *et al.*, 1971, p. 18).

More recently, the SEEQ is frequently used by many higher education institutions. The SEEQ measures nine factors that constitute quality instruction (Marsh, 1982, 1987; Marsh and Dunkin, 1997; Richardson, 2005). These are assignments and readings, breadth of coverage, examinations and grading, group interaction, individual rapport, instructor enthusiasm, learning and academic value, organization and clarity, and workload and

difficulty (Marsh, 2007, p. 323). Some SEEQ studies include an overall summative evaluation of the course subject as an additional factor (Schellhase, 2010). The similarities with Hildebrand (1973) and Hildebrand *et al.* (1971) and Feldman's (1976) criteria of effective teaching are apparent.

In a series of studies conducted at the University of Hawaii, SET is first analyzed with respect to the perspectives of faculty members, which identifies important factors such as evaluation information based from students, information from peers (colleagues), student performance and grades, and external performance evaluations of teachers (Meredith, 1977). A study that included apprentice teachers (practice teachers) found that students preferred instructors who exhibited classroom organizational skills, who focused on students' learning outcomes, and who interacted well with students (Meredith and Bub, 1977). A set of evaluation criteria was developed based on a study of both faculty members and students in the School of Law at the University of Hawaii, which included dimensions such as knowledge of subject matter, ability to stimulate interest and motivate students, organization of the course, preparation for the course, concern for students, quality of course materials, and an overall summative evaluation of the teacher (Meredith, 1978). Other studies measured teaching excellence by subject mastery, teaching skills, and personal qualities of the teacher (Meredith, 1985b), while an overall analysis of student satisfaction used the criteria social interaction, teaching quality, campus environment, employment opportunities, and classroom facilities (Meredith, 1985a), all of which contribute to SET ratings.

In summary, it is noted that SETs can vary depending on whether the evaluations are from the perspective of faculty members (how teachers teach) or from the students (how students learn). However, although several variations of SETs exist, comparisons suggest that as long as the overall objective is to evaluate effective teaching, dimensions within these SETs are interrelated and may overlap (Marsh, 1984, 2007; Marsh and Bailey, 1993; Marsh and Dunkin, 1997). A study conducted by the American Association of University Professors involving 9,000 faculty members found that SETs are generally established with controversial biases and issues (Flaherty, 2015). The more important issue is the establishment of the objectives for SET implementation within the university and careful decision making about who should participate in the development of such an evaluation instrument.

5. Antecedents of SET

Within the vast literature on SETs, analysis of their validity and reliability has identified various antecedents affecting effective evaluation. SET ratings are dependent on several issues, including the various biases already discussed. The first obvious antecedent is the instructor, as can be discerned from the previous discussions. Besides personality issues, gender plays an important role. Boring *et al.* (2016) find that SETs are statistically biased against female faculty, and that such biases can cause effective teachers to get lower SET ratings than less effective ones. MacNeill *et al.* (2015) conducted an experiment in which students were blind to the gender of their online course instructors. For the experiment, two online course instructors were selected, one male and one female, and each was given two classes to teach. Later in the course, each instructor presented as one gender to one class and the opposite gender to the other class. The SET results gathered at the end of the semester are interesting. Regardless of the instructor's real gender, students gave the teacher they thought was male and the actual male teacher higher SET ratings than the teacher they perceived as female. This experiment clearly shows that the rating difference results from gender bias (Marcotte, 2014).

Previous studies also show that the time of SET evaluation matters. As discussed, when SET evaluations are administered during the middle of the semester, results can assist teachers in re-evaluating their course design to better fit with the students' needs and capabilities. However, this phenomenon is limited. SETs are mostly given before the end of

the term or during final examinations, and studies have shown that ratings taken at this time tend to be lower compared to evaluations conducted a few weeks before final exams (Braskamp *et al.*, 1984). Interestingly, no significant differences were found when comparing SET ratings gathered before the end of the semester with those taken in the first week of the succeeding term (Frey, 1976). This debunks the fallacy that students tend to seek revenge on teachers because of issues with the grades received (Clayson *et al.*, 2006; Skinner and Belmont, 1993). In fact, studies have proven that students who received poor grades were less likely to care enough to complete the SET (Liegle and McDonald, 2005).

In terms of the students themselves, as previously mentioned the background demographics of students do significantly affect SETs (Stark and Freishtat, 2014). Although some biases are found between gender and SET ratings (Boring *et al.*, 2016; Feldman, 1977), still there are no consistent evidence of such difference exists (Wachtel, 2006). For instance, different studies have shown that male and female students give higher ratings as compared to their peers of opposite genders (Tatro, 1995). In some instances, students evaluate their same gender teachers higher than their opposite gender instructors (Centra, 1993a, b). With regards to ethnicity, Marsh *et al.* (1997) translated the SEEQ instrument into Chinese and found that there are no significant differences with the results reported as compared with the studies done in the USA. In other Chinese studies, besides the significant differences in SET ratings between students of various discipline and nature (Chen and Watkins, 2010; Liu *et al.*, 2016), it is well noted that linguistics or foreign language teachers tend to received higher evaluations than the faculty of other discipline (Chen and Watkins, 2010).

Administration conditions or the way SETs are administered also matters. Currently, SETs are mostly collected using online course evaluations (Spooren and Van Loon, 2012). However, literature shows that online SETs results in lower participation (Anderson *et al.*, 2005; Avery *et al.*, 2006), although reminders do increase the response rate (Norris and Conn, 2005). With paper-and-pen SETs, the person administering the evaluation also contributes to any inconsistencies in the ratings. This holds true even if the teacher leaves the room during the SET administration and the forms are anonymous, as students may still be reluctant to provide an objective evaluation (Pulich, 1984). Many researchers have agreed that SETs should be entrusted to a third-party individual for effective collection (Braskamp *et al.*, 1984; Centra, 1979).

The characteristics of the course subject also matters. Wachtel (2006) notes that the nature of the course subject, such as whether it is a required course or an elective, affects how students rate its importance. Sometimes students give higher ratings for elective course subjects due to their having a prior interest in the subject (Feldman, 1978). Class schedule can sometimes affect ratings, and odd schedules such as early morning classes or late afternoon classes have been found to receive the lowest SET ratings (Koushki and Kuhn, 1982). However, inconsistencies were found in several other studies (Aleamoni, 1999; Centra, 1979; Feldman, 1978; Wachtel, 2006), but it has been suggested that the level of the course is a relevant factor. The year or level of the course is closely related to the students' age; as students continue with their studies, they becomes more mature and become aware that their opinions are taken seriously by the school administration (Spooren and Van Loon, 2012). Class size has also been found to have an impact (Feldman, 1978; Marsh, 2007) since bigger classes tend to present less opportunities for interaction between the teacher and the individual students, which can affect ratings (Meredith and Ogasawara, 1982). Finally, the subject area and the discipline also greatly influence SET ratings. Since the discipline affects how classes are held (e.g. laboratory classes compared to lecture intensive courses), comparisons between colleges are not advisable (Wachtel, 2006). For instance, task-oriented subjects such as mathematics and science offer less interaction than the social sciences (Centra, 1993a, b).

In summary, apart from the issues relating to students that affect SETs discussed in the “Experience Qualities” section of this paper, including their gender, learning motivations, and grade expectations (Boring *et al.*, 2016), many more have been added to the discussion. Having examined the various antecedents of SETs, it is apparent that one model is not suitable for all instances. More specifically, one single type of SET cannot and should not be used to collect students’ perception across all courses and subjects. This is actually the main reason why some higher education institutions choose to use global measures to collect the summative evaluations of the class. In practice, separate SETs should be used for different course types. Since this can place a significant burden on institutions, careful analysis and research is necessary.

6. Conclusion

To sum up, literature has shown that the use of SETs to collect information regarding the teaching–learning process is commonplace. However, given the complex nature of academic processes, the data resulting from SETs are questionable and limited. The current paper presents a review of the literature on SETs, focusing on the concept of service marketing evaluation. The framework’s three criteria are used to examine SETs, whereby the teacher represents the “search” qualities, the teaching process the “experience” qualities, and the learning outcomes the “credence” qualities.

The search qualities with regard to SETs are the easily observable attributes of teachers. These may include the appearance, gender, age, ethnicity, and personalities traits of faculty members. In practice, course subject selections are made prior to enrollment in a course; students can compare faculty members when deciding which one to enroll with. Hence, the expectations of students are important. It has been noted that stereotyping faculty members according to certain demographic factors such as gender and age is unfair since these features are fixed and impossible to change. Students should look beyond these obvious factors and focus more on the teachers’ credibility and competencies.

Beyond initial search preferences, students place much importance on evaluating their learning experiences. As the literature suggests, for the sake of simplicity, many SETs include only global summative evaluations of the teaching–learning process. However, given that the nature of the learning experience is complex and multidimensional, evidence to support student development should be in the form of formative judgments. Furthermore, the actual teaching–learning process is composed of repeated service encounters (a semester in Taiwan typically lasts around 18 weeks). It is, therefore, difficult to determine whether a single class experience or the collective sum of the semester’s learning encounters contribute to the SET ratings. Considering the influence of prior expectations on SET ratings, teachers are advised to establish a psychological contract with the students. To make these agreements effective, they should be negotiated at the start of the term, so that they are shared contracts between the teacher and the students.

Finally, accepting that university education is characterized by high credence qualities, students must be aware of the concept of academic delay of gratification, so that they understand and accept that the benefits of undertaking a course are not immediate. Combining this with the importance of students’ expectations and the usefulness of creating a psychological contract, clear definitions of the potential benefits and acceptable performance should be provided during the first class. Moreover, connections should be made between previous, current, and future courses, thus providing an overview of the entire program together with career pathways.

In summary, since SETs are frequently used to collect information on effective teaching, it is important for higher education institutions to establish what kinds of SETs are effective. Given the complex factors involved and the various antecedents of SETs, it appears that no one perfect tool exists to accurately measure what happens in the classroom. As different

SETs may be necessary for different courses and subjects, options such as faculty members' self-evaluation and/or faculty members' peer-evaluation might be considered to provide what is lacking in SETs. It is hoped that as technology advances, an innovative way of collecting SETs might be found to make the process more productive.

6.1 Recommendations for further research

Having analyzed the above issues, several recommendations for further research are proposed:

Develop and validate an SET. The development of the SET is important in the ongoing dialogue within the literature. As the literature shows, SETs are only useful if they can appropriately capture what they are being used to measure. Hence, in order to develop a relevant and constructive SET, the participation of significant stakeholders, such as school administrators, faculty, and students, is essential. A constructive SET would be capable of providing formative recommendations to improve the performance of both faculty members and students. More important, an effective SET should consider the service attributes (the search, experience, and credence qualities) that students want.

Develop an SET software program. In the current age of technological advances and big data, students are adept at using mobile devices (Gikas and Grant, 2013). Therefore, an app designed to collect SET ratings – either directly after each class, twice a semester (after midterm exams and before the end of the semester), or once before the end of the semester – could be made available to students for easy and convenient collection of data. This could initiate a new strand of SET literature. Combining technology with pedagogy can provide a more accurate evaluation of teaching, by facilitating the collection of real-time SET results.

References

- Adams, G.R. (1977), "Physical attractiveness research: toward a developmental social psychology of beauty", *Human Development*, Vol. 20 No. 4, pp. 217-239, available at: <https://doi.org/10.1159/000271558>
- Aleamoni, L.M. (1974), "Typical faculty concerns about student evaluation of instruction", *NACTA*, Vol. 20 No. 1, pp. 16-21.
- Aleamoni, L.M. (1999), "Student rating myths versus research facts from 1924 to 1998", *Journal of Personnel Evaluation in Education*, Vol. 13 No. 2, pp. 153-166, available at: <https://doi.org/10.1023/A:1008168421283>
- Aleamoni, L.M. and Hexner, P.Z. (1980), "A review of the research on student evaluation and a report on the effect of different sets of instructions on student course and instructor evaluation", *Instructional Science*, Vol. 9 No. 1, pp. 67-84.
- Algozzine, B., Gretes, J., Flowers, C., Howley, L., Beattie, J., Spooner, F., Mohanty, G. and Bray, M. (2004), "Student evaluation of college teaching: a practice in search of principles", *College Teaching*, Vol. 52 No. 4, pp. 134-141, available at: <https://doi.org/10.3200/CTCH.52.4.134-141>
- Altbach, P.G. (2011), "Introduction", in Altbach, P.G. (Ed.), *Leadership for World-Class Universities: Challenges for Developing Countries*, Routledge, New York, NY, pp. 1-7.
- Altbach, P.G., Reisberg, L. and Rumbley, L.E. (2009), *Trends in Global Higher Education: Tracking an Academic Revolution*, UNESCO, Paris.
- Ames, C. and Archer, J. (1988), "Achievement goals in the classroom: students' learning strategies and motivation processes", *Journal of Educational Psychology*, Vol. 80 No. 3, pp. 260-267.
- Anderson, H.M., Cain, J. and Bird, E. (2005), "Online student course evaluations: review of literature and a pilot study", *American Journal of Pharmaceutical Education*, Vol. 69 No. 1, pp. 34-43.
- Arreola, R.A. (2007), *Developing a Comprehensive Faculty Evaluation System: A Guide to Designing, Building, and Operating Large-Scale Faculty Evaluation Systems*, Jossey-Bass, San Francisco, CA.

- Arthur, L. (2009), "From performativity to professionalism: lectures' responses to student feedback", *Teaching in Higher Education*, Vol. 14 No. 4, pp. 441-454, available at: <https://doi.org/10.1080/13562510903050228>
- Avery, R.J., Bryant, W.K., Mathios, A., Kang, H. and Bell, D. (2006), "Electronic course evaluations: does an online delivery system influence student evaluations?", *The Journal of Economic Education*, Vol. 36 No. 1, pp. 21-37, available at: <https://doi.org/10.3200/JECE.37.1.21-37>
- Babad, E., Darley, J.M. and Kaplowitz, H. (1999), "Developmental aspects in students' course selection", *Journal of Educational Psychology*, Vol. 91 No. 1, pp. 157-168.
- Babcock, P. (2010), "Real costs of nominal grade inflation? New evidence from student course evaluations", *Economic Inquiry*, Vol. 48 No. 4, pp. 983-996, available at: <https://doi.org/10.1111/j.1465-7295.2009.00245.x>
- Baumeister, R.F. and Leary, M.R. (1997), "Writing narrative literature reviews", *Review of General Psychology*, Vol. 1 No. 3, pp. 311-320.
- Beatty, M.J. and Zahn, C.J. (1990), "Are student ratings of communication instructors due to 'easy' grading practices? An analysis of teacher credibility and student-reported performance levels", *Communication Education*, Vol. 39 No. 4, pp. 275-282, available at: <https://doi.org/10.1080/03634529009378809>
- Bell, R.A. and Daly, J.A. (1984), "The affinity-seeking function of communication", *Communication Monographs*, Vol. 51 No. 2, pp. 91-115, available at: <https://doi.org/10.1080/03637758409390188>
- Bembenutty, H. (1999), "Sustaining motivation and academic goals: the role of academic delay of gratification", *Learning and Individual Differences*, Vol. 11 No. 3, pp. 233-257, available at: [https://doi.org/10.1016/S1041-6080\(99\)80002-8](https://doi.org/10.1016/S1041-6080(99)80002-8)
- Bembenutty, H. (2009), "Teaching effectiveness, course evaluation, and academic performance: the role of academic delay of gratification", *Journal of Advanced Academics*, Vol. 20 No. 2, pp. 326-355.
- Bembenutty, H. and Karabenick, S.A. (1998), "Academic delay of gratification", *Learning and Individual Differences*, Vol. 10 No. 4, pp. 329-346, available at: [https://doi.org/10.1016/S1041-6080\(99\)80126-5](https://doi.org/10.1016/S1041-6080(99)80126-5)
- Bembenutty, H. and Karabenick, S.A. (2004), "Inherent association between academic delay of gratification, future time perspective, and self-regulated learning", *Educational Psychology Review*, Vol. 16 No. 1, pp. 35-57, available at: <https://doi.org/10.1023/B:EDPR.0000012344.34008.5c>
- 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*, Vol. 74 No. 2, pp. 170-179, available at: <https://doi.org/10.1037/0022-0663.74.2.170>
- Berk, R.A. (2005), "Survey of 12 strategies to measure teaching effectiveness", *International Journal of Teaching and Learning in Higher Education*, Vol. 17 No. 1, pp. 48-62.
- Berscheid, E. and Walster, E. (1974), "Physical attractiveness", *Advances in Experimental Social Psychology*, Vol. 7 No. 1, pp. 157-215, available at: [https://doi.org/10.1016/S0065-2601\(08\)60037-4](https://doi.org/10.1016/S0065-2601(08)60037-4)
- Betoret, F.D. (2007), "The influence of students' and teachers' thinking styles on student course satisfaction and on their learning process", *Educational Psychology: An International Journal of Experimental Educational Psychology*, Vol. 27 No. 2, pp. 219-234, available at: <https://doi.org/10.1080/01443410601066701>
- Bolliger, D.U. (2004), "Key factors for determining student satisfaction in online courses", *International Journal on E-Learning*, Vol. 3 No. 1, pp. 61-67.
- Bonds-Raacke, J. and Raacke, J.D. (2007), "The relationship between physical attractiveness of professors and students' ratings of professor quality", *Journal of Psychiatry, Psychology and Mental Health*, Vol. 1 No. 2, pp. 1-7.
- Boring, A., Ottoboni, K. and Stark, P.B. (2016), "Student evaluations of teaching (mostly) do not measure teaching effectiveness", *ScienceOpen Research*, available at: <https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1> (accessed March 30, 2018).

- Braga, M., Paccagnella, M. and Pellizzari, M. (2014), "Evaluating students' evaluations of professors", *Economics of Education Review*, Vol. 41 No. 1, pp. 71-88, available at: <https://doi.org/10.1016/j.econedurev.2014.04.002>
- Braskamp, L.A., Brandenburg, D.C. and Ory, J.C. (1984), *Evaluating Teaching Effectiveness*, Sage, Newbury Park, CA.
- Brownell, J.E. and Swaner, L.E. (2009), "High-impact practices: applying the learning outcomes literature to the development of successful campus programs", *Peer Review*, Vol. 11 No. 2, pp. 26-30.
- Buck, S. and Tiene, D. (1989), "The impact of physical attractiveness, gender, and teaching philosophy on teacher evaluations", *The Journal of Educational Research*, Vol. 82 No. 3, pp. 172-177, available at: <https://doi.org/10.1080/00220671.1989.10885887>
- Carter, R.E. (2016), "Faculty scholarship has a profound positive association with student evaluations of teaching: except when it doesn't", *Journal of Marketing Education*, Vol. 38 No. 1, pp. 18-36.
- Cashin, W.E. and Downey, R.G. (1992), "Using global student rating items for summative evaluation", *Journal of Educational Psychology*, Vol. 84 No. 4, pp. 563-572.
- Castleman, B. (2016), "Data-driven behavioral nudges: a low-cost strategy to improve postsecondary education", paper presented at the Annual Conference of the Association for Institutional Research, New Orleans, LA.
- Centra, J.A. (1979), *Determining Faculty Effectiveness: Assessing Teaching, Research, and Service for Personnel Decisions and Improvement*, Jossey-Bass, San Francisco, CA.
- Centra, J.A. (1993a), *Determining Faculty Effectiveness*, Jossey-Bass, San Francisco, CA.
- Centra, J.A. (1993b), *Reflective Faculty Evaluation*, Jossey-Bass, San Francisco, CA.
- Chang, J. (2014), "Number of universities should be reduced: education minister", available at: www.chinapost.nownews.com/20140925-50233 (accessed March 30, 2018).
- Charles Stuart University Library (2018), "Literature review: traditional or narrative literature reviews", available at: <http://libguides.csu.edu.au/c.php?g=476545&p=3997199> (accessed March 30, 2018).
- Chen, C.Y. (2006), "A study on teaching evaluation in public universities in Taiwan (Wguó gōnglìdàxué jiàoshī jiàoxué píngjiàn zhī yánjiū)", unpublished doctoral dissertation, National ChengChi University, Taipei.
- Chen, D.S. and Chang, M.-K. (2010), "Higher education in Taiwan: the crisis of rapid expansion", available at: www.isa-sociology.org/universities-in-crisis/?p=417 (accessed March 30, 2018).
- Chen, G.-H. and Watkins, D. (2010), "Stability and correlates of student evaluations of teaching at a Chinese university", *Assessment and Evaluation in Higher Education*, Vol. 35 No. 6, pp. 675-685, available at: <https://doi.org/10.1080/02602930902977715>
- Chen, L. (2016), "Do student characteristics affect course evaluation completion?", paper presented at the 2016 Annual Conference of the Association for Institutional Research, New Orleans, LA.
- Chou, C.P. and Ching, G.S. (2012), *Taiwan Education at the Crossroad: When Globalization Meets Localization*, Palgrave Macmillan, New York, NY.
- Chuah, K.L. and Hill, C. (2004), "Student evaluation of teacher performance: random pre-destination", *Journal of College Teaching & Learning*, Vol. 1 No. 6, pp. 109-114.
- Clayson, D.E. (1999), "Students' evaluation of teaching effectiveness: some implications of stability", *Journal of Marketing Education*, Vol. 21 No. 1, pp. 68-75.
- 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*, Vol. 31 No. 1, pp. 16-30.
- Clayson, D.E. and Sheffet, M.J. (2006), "Personality and the student evaluation of teaching", *Journal of Marketing Education*, Vol. 28 No. 2, pp. 149-160.
- Clayson, D.E., Frost, T.F. and Sheffet, M.J. (2006), "Grades and the student evaluation of instruction: a test of the reciprocity effect", *Academy of Management: Learning and Education*, Vol. 5 No. 1, pp. 52-65, available at: <https://doi.org/10.5465/AMLE.2006.20388384>

-
- Coburn, L. (1984), "Student evaluation of teacher performance", ERIC Document Reproduction Service No. ED289887, National Institute of Education, Washington, DC.
- Coffey, M. and Gibbs, G. (2001), "The evaluation of the Student Evaluation of Educational Quality Questionnaire (SEEQ) in UK higher education", *Assessment & Evaluation in Higher Education*, Vol. 26 No. 1, pp. 89-93, available at: <https://doi.org/10.1080/02602930020022318>
- Cohen, P.A. (1980), "Effectiveness of student-rating feedback for improving college instruction: a meta-analysis of findings", *Research in Higher Education*, Vol. 13 No. 4, pp. 321-341, doi: 10.1007/bf00976252.
- Costin, F., Greenough, W.T. and Menges, R.J. (1971), "Student ratings of college teaching: reliability, validity, and usefulness", *Review of Educational Research*, Vol. 41 No. 5, pp. 511-535.
- Cronin, P., Ryan, F. and Coughlan, M. (2008), "Undertaking a literature review: a step-by-step approach", *British Journal of Nursing*, Vol. 17 No. 1, pp. 38-43.
- Curran, J.M. and Rosen, D.E. (2006), "Student attitudes toward college courses: an examination of influences and intentions", *Journal of Marketing Education*, Vol. 28 No. 2, pp. 135-148, available at: <https://doi.org/10.1177/0273475306288401>
- Daniel, B. (2015), "Big data and analytics in higher education: opportunities and challenges", *British Journal of Educational Technology*, Vol. 46 No. 5, pp. 904-920, available at: <https://doi.org/10.1111/bjet.12230>
- Darby, M.R. and Karni, E. (1973), "Free competition and the optimal amount of fraud", *The Journal of Law & Economics*, Vol. 16 No. 1, pp. 67-88.
- Dee, T.S. (2005), "A teacher like me: does race, ethnicity, or gender matter?", *The American Economic Review*, Vol. 95 No. 2, pp. 158-165.
- Douglas, D. and Gifford, R. (2001), "Evaluation of the physical classroom by students and professors: a lens model approach", *Educational Research*, Vol. 43 No. 3, pp. 295-309, available at: <https://doi.org/10.1080/00131880110081053>
- Dudovskiy, J. (2018), "The ultimate guide to writing a dissertation in business studies: a step-by-step assistance", *Research Methodology*.
- Duggan, M. and Carlson-Bancroft, A. (2016), "How Emerson College increased participation rates in course evaluations and NSSE", paper presented at the Annual Conference of the Association for Institutional Research, New Orleans, LA.
- Eagly, A.H., Ashmore, R.D., Makhijani, M.G. and Longo, L.C. (1991), "What is beautiful is good, but ...: a meta-analytic review of research on the physical attractiveness stereotype", *Psychological Bulletin*, Vol. 110 No. 1, pp. 109-128, available at: <https://doi.org/10.1037/0033-2909.110.1.109>
- Ehrenberg, R.G., Goldhaber, D.D. and Brewer, D.J. (1995), "Do teachers' race, gender, and ethnicity matter? Evidence from the national education longitudinal study of 1988", *Industrial and Labor Relations Review*, Vol. 48 No. 3, pp. 547-561.
- Feldman, K.A. (1976), "The superior college teacher from the students' view", *Research in Higher Education*, Vol. 5 No. 3, pp. 243-288, available at: <https://doi.org/10.1007/BF00991967>
- Feldman, K.A. (1977), "Consistency and variability among college students in rating their teachers and courses: a review and analysis", *Research in Higher Education*, Vol. 6 No. 3, pp. 223-274.
- Feldman, K.A. (1978), "Course characteristics and college students' ratings of their teachers: what we know and what we don't", *Research in Higher Education*, Vol. 9 No. 3, pp. 199-242, available at: <https://doi.org/10.1007/BF00976997>
- Feldman, K.A. (1984), "Class size and college students' evaluations of teachers and courses: a closer look", *Research in Higher Education*, Vol. 21 No. 1, pp. 45-116, available at: <https://doi.org/10.1007/BF00975035>
- Felton, J., Mitchell, J. and Stinson, M. (2004), "Web-based student evaluations of professors: the relations between perceived quality, easiness, and sexiness", *Assessment & Evaluation in Higher Education*, Vol. 29 No. 1, pp. 91-108, available at: <https://doi.org/10.1080/0260293032000158180>

- Fink, A. (2005), *Conducting Research Literature Reviews: From the Internet to Paper*, 2nd ed., Sage, Thousand Oaks, CA.
- Fisk, R.P., Grove, S.J. and John, J. (2014), *Services Marketing: An Interactive Approach*, 4th ed., Cengage Learning, Mason, OH.
- Flaherty, C. (2015), "Flawed evaluations", available at: www.insidehighered.com/news/2015/06/10/aaup-committee-survey-data-raise-questions-effectiveness-student-teaching (accessed March 30, 2018).
- Flood, B. (1970), "Student evaluation of teacher performance", *Journal of Education for Librarianship*, Vol. 10 No. 4, pp. 283-285, available at: <https://doi.org/10.2307/40322085>
- Fox, R., Peck, R.F., Blattstein, A. and Blattstein, D. (1983), "Student evaluation of teacher as a measure of teacher behavior and teacher impact on students", *The Journal of Educational Research*, Vol. 77 No. 1, pp. 16-21.
- Frey, P.W. (1976), "Validity of student instructional ratings: does timing matter?", *The Journal of Higher Education*, Vol. 47 No. 3, pp. 327-336.
- Galetzka, M., Verhoeven, J.W.M. and Pruyn, A.T.H. (2006), "Service validity and service reliability of search, experience and credence services: a scenario study", *International Journal of Service Industry Management*, Vol. 17 No. 3, pp. 271-283, available at: <https://doi.org/10.1108/09564230610667113>
- Gikas, J. and Grant, M.M. (2013), "Mobile computing devices in higher education: student perspectives on learning with cellphones, smartphones & social media", *The Internet and Higher Education*, Vol. 19 No. 1, pp. 18-26.
- Glaser, B.G. (1965), "The constant comparative method of qualitative analysis", *Social Problems*, Vol. 12 No. 4, pp. 436-445.
- Glaser, B.G. (1978), *Theoretical Sensitivity: Advances in the Methodology of Grounded Theory*, Sociology Press, Mill Valley, CA.
- Grammatikopoulos, V., Linardakis, M., Gregoriadis, A. and Oikonomidis, V. (2015), "Assessing the Students' Evaluations of Educational Quality (SEEQ) questionnaire in Greek higher education", *Higher Education*, Vol. 70 No. 3, pp. 395-408, available at: <https://doi.org/10.1007/s10734-014-9837-7>
- Green, B.N., Johnson, C.D. and Adams, A. (2006), "Writing narrative literature review for peer-reviewed journals: secrets of the trade", *Journal of Chiropractic Medicine*, Vol. 5 No. 3, pp. 101-117.
- Greene, B.A., Miller, R.B., Crowson, H.M., Duke, B.L. and Akey, K.L. (2004), "Predicting high school students' cognitive engagement and achievement: contributions of classroom perceptions and motivation", *Contemporary Educational Psychology*, Vol. 29 No. 4, pp. 462-482, available at: <https://doi.org/10.1016/j.cedpsych.2004.01.006>
- Grove, S.J. and Fisk, R.P. (1997), "The impact of other customers on service experiences: a critical incident examination of 'getting alone'", *Journal of Retailing*, Vol. 73 No. 1, pp. 63-85.
- Harackiewicz, J.M., Barron, K.E., Pintrich, P.R., Elliot, A.J. and Thrash, T.M. (2002), "Revision of achievement goal theory: necessary and illuminating", *Journal of Educational Psychology*, Vol. 94 No. 3, pp. 638-645, available at: <https://doi.org/10.1037/0022-0663.94.3.638>
- Hart, C. (1998), *Doing a Literature Review: Releasing the Social Science Research Imagination*, Sage, Thousand Oaks, CA.
- Hildebrand, M. (1973), "The character and skills of the effective professor", *The Journal of Higher Education*, Vol. 44 No. 1, pp. 41-50.
- Hildebrand, M., Wilson, R.C. and Dienst, E.R. (1971), *Evaluating University Teaching*, Center for Research and Development in Higher Education, Berkeley, CA.
- Hill, M.C. and Epps, K.K. (2010), "The impact of physical classroom environment on student satisfaction and student evaluation of teaching in the university environment", *Academy of Educational Leadership Journal*, Vol. 14 No. 4, pp. 65-79.

- Howell, A.J. and Symbaluk, D.G. (2001), "Published student ratings of instruction: revealing and reconciling the views of students and faculty", *Journal of Educational Psychology*, Vol. 93 No. 4, pp. 790-796, available at: <https://doi.org/10.1037/0022-0663.93.4.790>
- Hu, Y.-L. and Ching, G.S. (2012), "Factors affecting student engagement: an analysis on how and why students learn", *Conference on Creative Education, Scientific Research Publishing, Irvine, CA*, pp. 989-992.
- Hu, Y.-L., Hung, C.-H. and Ching, G.S. (2015), "Student-faculty interaction: mediating between student engagement factors and educational outcome gains", *International Journal of Research Studies in Education*, Vol. 4 No. 1, pp. 43-53, available at: <https://doi.org/10.5861/ijrse.2014.800>
- Hultman, M. and Oghazi, P. (2008), "Good looks - good courses: the link between physical attractiveness and perceived performance in higher educational services", in Thyne, M., Deans, K.R. and Gnoth, J. (Eds), *Australian and New Zealand Marketing Academy Conference*, University of Otago, Dunedin, pp. 2588-2597.
- Huynh, P. (2015), "Overcoming low response rates for online course evaluations", paper presented at the Annual Conference of the Association for Institutional Research, Denver, CO.
- Kolb, D.A., Rubin, I.M. and McIntyre, J.M. (1984), *Organizational Psychology: An Experimental Approach to Organizational Behavior*, Prentice-Hall, Englewood Cliffs, NJ.
- Kornell, N. and Hausman, H. (2016), "Do the best teachers get the best rating?", *Frontiers in Psychology*, Vol. 7, No. 570, pp. 1-8, available at: <https://doi.org/10.3389/fpsyg.2016.00570>
- Koushki, P.A. and Kuhn, H.A.J. (1982), "How reliable are student evaluations of teachers?", *Engineering Education*, Vol. 72 No. 3, pp. 362-367.
- Kuh, G.D. (2008), *High-Impact Educational Practices: What they are, who has Access to them, and why they Matter*, AACU, Washington, DC.
- Kuh, G.D., O'Donnell, K. and Reed, S. (2013), *Ensuring Quality and Taking High-Impact Practices to Scale*, AACU, Washington, DC.
- Kuzmanovic, M., Savic, G., Popovic, M. and Martic, M. (2012), "A new approach to evaluation of university teaching considering heterogeneity of students' preferences", *Procedia - Social and Behavioral Sciences*, Vol. 64 No. 1, pp. 402-411, available at: <https://doi.org/10.1016/j.sbspro.2012.11.047>
- Liao, S. (2013), "Psychological contract between teacher and student improves teaching process in the network courses of college: a study based on the network course of psychology in Shaoguan University", in Luo, X. (Ed.), *International Conference on Education Technology and Management Science*, Atlantis Press, Amsterdam, pp. 885-887.
- Lidice, A. and Saglam, G. (2013), "Using students' evaluations to measure educational quality", *Procedia - Social and Behavioral Sciences*, Vol. 70 No. 25, pp. 1009-1015, available at: <https://doi.org/10.1016/j.sbspro.2013.01.152>
- Liegle, J.O. and McDonald, D.S. (2005), "Lessons learned from online vs paper-based computer information students evaluation system", *Information Systems Education Journal*, Vol. 3 No. 37, pp. 1-14, available at: <http://isedj.org/3/37/>
- Lin, J.-H. and Chen, J.-H. (2016), "SELF-COLA: assessing students learning experiences and first-year outcomes", paper presented at the International Conference: Higher Education Institutional Research, New Orleans, LA.
- Lipsey, M.W. and Wilson, D.B. (2001), *Practical Meta-Analysis*, Vol. 49, Sage, Thousand Oaks, CA.
- Liu, S., Keeley, J. and Buskist, W. (2016), "Chinese college students' perceptions of excellent teachers across three disciplines", *Psychology, Chemical Engineering, and Education*, Vol. 43 No. 1, pp. 70-74, available at: <https://doi.org/10.1177/0098628315620888>
- Longanecker, D. (2016), "Higher education in the new normal of the 21st century: an era of evidence based change", paper presented at the Annual Conference of the Association for Institutional Research, New Orleans, LA.
- Lubienski, C. (2007), "Marketing schools: consumer goods and competitive incentives for consumer information", *Education and Urban Society*, Vol. 40 No. 1, pp. 118-141.

- MacGregor, K. (2015), "Six key elements of an entrepreneurial university", University World News, available at: www.universityworldnews.com/article.php?story=20151106141848199 (accessed March 30, 2018).
- Machi, L.A. and McEvoy, B.T. (2016), *The Literature Review: Six Steps to Success*, 3rd ed., Sage, Thousand Oaks, CA.
- MacNell, L., Driscoll, A. and Hunt, A.N. (2015), "What's in a name: exposing gender bias in student ratings of teaching", *Innovative Higher Education*, Vol. 40 No. 4, pp. 291-303, available at: <https://doi.org/10.1007/s10755-014-9313-4>
- Mäkinen, J., Olkinuora, E. and Lonka, K. (2004), "Students at risk: students' general study orientations and abandoning/prolonging the course of studies", *Higher Education*, Vol. 48 No. 2, pp. 173-188, available at: <https://doi.org/10.1023/B:HIGH.0000034312.79289.ab>
- Marcotte, A. (2014), "Best way for professors to get good student evaluations? Be male", available at: www.slate.com/blogs/xx_factor/2014/12/09/gender_bias_in_student_evaluations_professors_of_online_courses_who_present.html (accessed March 30, 2018).
- Marks, R.B. (2000), "Determinants of student evaluations of global measures of instructor and course value", *Journal of Marketing Education*, Vol. 22 No. 2, pp. 108-119, available at: <https://doi.org/10.1177/0273475300222005>
- Marlin, J.W. Jr (1987), "Student perceptions of end-of-course evaluations", *The Journal of Higher Education*, Vol. 58 No. 6, pp. 704-716.
- Marsh, H.W. (1980), "The influence of student, course, and instructor characteristics in evaluations of university teaching", *American Educational Research Journal*, Vol. 17 No. 2, pp. 219-237.
- Marsh, H.W. (1982), "SEEQ: a reliable, valid, and useful instrument for collecting students' evaluations of university teaching", *British Journal of Educational Psychology*, Vol. 52 No. 1, pp. 77-95, available at: <https://doi.org/10.1111/j.2044-8279.1982.tb02505.x>
- Marsh, H.W. (1984), "Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases, and utility", *Journal of Educational Psychology*, Vol. 76 No. 5, pp. 707-754, available at: <https://doi.org/10.1037/0022-0663.76.5.707>
- Marsh, H.W. (1987), "Students' evaluations of university teaching: research findings, methodological issues, and directions for future research", *International Journal of Educational Research*, Vol. 11 No. 3, pp. 253-388, available at: [https://doi.org/10.1016/0883-0355\(87\)90001-2](https://doi.org/10.1016/0883-0355(87)90001-2)
- Marsh, H.W. (1991), "Multidimensional students' evaluations of teaching effectiveness: a test of alternative higher-order structures", *Journal of Educational Psychology*, Vol. 83 No. 2, pp. 285-296, available at: <https://doi.org/10.1037/0022-0663.83.2.285>
- Marsh, H.W. (2007), "Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases and usefulness", in Perry, R.P. and Smart, J.C. (Eds), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, Springer, Dordrecht, pp. 319-383.
- Marsh, H.W. and Bailey, M. (1993), "Multidimensional students' evaluations of teaching effectiveness: a profile analysis", *The Journal of Higher Education*, Vol. 64 No. 1, pp. 1-18, available at: <https://doi.org/10.2307/2959975>
- Marsh, H.W. and Dunkin, M.J. (1997), "Students' evaluations of university teaching: a multidimensional perspective", in Perry, R.P. and Smart, J.C. (Eds), *Effective Teaching in Higher Education: Research and Practice*, Agathon, New York, NY, pp. 241-320.
- Marsh, H.W. and Roche, L.A. (1997), "Making students' evaluations of teaching effectiveness effective: the critical issues of validity, bias, and utility", *American Psychologist*, Vol. 52 No. 11, pp. 1187-1197, available at: <https://doi.org/10.1037/0003-066X.52.11.1187>
- Marsh, H.W. and Roche, L.A. (2000), "Effects of grading leniency and low workload on students' evaluations of teaching: popular myth, bias, validity, or innocent bystanders?", *Journal of Educational Psychology*, Vol. 92 No. 1, pp. 202-228, available at: <https://doi.org/10.1037/0022-0663.92.1.202>

-
- Marsh, H.W., Hau, K.-T., Chung, C.-M. and Siu, T.L.P. (1997), "Students' evaluations of university teaching: Chinese version of the students' evaluations of educational quality instrument", *Journal of Educational Psychology*, Vol. 89 No. 3, pp. 568-572, available at: <https://doi.org/10.1037/0022-0663.89.3.568>
- Mehrabian, A. (1968), "Some referents and measures of nonverbal behavior", *Behavior Research Methods & Instrumentation*, Vol. 1 No. 6, pp. 203-207, available at: <https://doi.org/10.3758/BF03208096>
- Meredith, G.M. (1977), "Faculty-based indicators of teaching effectiveness in higher education", *Psychological Reports*, Vol. 41 No. 2, pp. 675-676, available at: <https://doi.org/10.2466/pr0.1977.41.2.675>
- Meredith, G.M. (1978), "Student-based ratings of teaching effectiveness in legal education", *Psychological Reports*, Vol. 43 No. 3, pp. 953-954, available at: <https://doi.org/10.2466/pr0.1978.43.3.953>
- Meredith, G.M. (1985a), "Student-based indicators of campus satisfaction as an outcome of higher education", *Psychological Reports*, Vol. 56 No. 2, pp. 597-598, available at: <https://doi.org/10.2466/pr0.1985.56.2.597>
- Meredith, G.M. (1985b), "Two rating indicators of excellence in teaching in lecture format courses", *Psychological Reports*, Vol. 56 No. 1, pp. 52-54, available at: <https://doi.org/10.2466/pr0.1985.56.1.52>
- Meredith, G.M. and Bub, D.N. (1977), "Evaluation of apprenticeship teaching in higher education", *Psychological Reports*, Vol. 40 No. 3, pp. 1123-1126, available at: <https://doi.org/10.2466/pr0.1977.40.3c.1123>
- Meredith, G.M. and Ogasawara, T.H. (1982), "Preference for class size in lecture-format courses among college students", *Psychological Reports*, Vol. 51 No. 3, pp. 961-962, available at: <https://doi.org/10.2466/pr0.1982.51.3.961>
- Miles, M. and Huberman, M. (1994), *Qualitative Data Analysis*, 2nd ed., Sage, Beverly Hills, CA.
- Miller, R.B., Greene, B.A., Montalvo, G.P., Ravindran, B. and Nichols, J.D. (1996), "Engagement in academic work: the role of learning goals, future consequences, pleasing others, and perceived ability", *Contemporary Educational Psychology*, Vol. 21 No. 4, pp. 388-422, available at: <https://doi.org/10.1006/ceps.1996.0028>
- Mitchell, M., Leachman, M. and Masterson, K. (2016), "Funding down, tuition up", available at: www.cbpp.org/research/state-budget-and-tax/funding-down-tuition-up (accessed March 30, 2018).
- Mogan, J. and Knox, J.E. (1987), "Characteristics of 'best' and 'worst' clinical teachers as perceived by university nursing faculty and students", *Journal of Advanced Nursing*, Vol. 12 No. 3, pp. 331-337, available at: <https://doi.org/10.1111/j.1365-2648.1987.tb01339.x>
- Mortelmans, D. and Spooren, P. (2009), "A revalidation of the SET37 questionnaire for student evaluations of teaching", *Educational Studies*, Vol. 35 No. 5, pp. 547-552, available at: <https://doi.org/10.1080/03055690902880299>
- Murray, H.G., Rushton, J.P. and Paunonen, S.V. (1990), "Teacher personality traits and student instructional ratings in six types of university courses", *Journal of Educational Psychology*, Vol. 82 No. 2, pp. 250-261, available at: <https://doi.org/10.1037/0022-0663.82.2.250>
- Naidoo, R. (2016), "Higher education is trapped in a competition fetish", University World News, available at: www.universityworldnews.com/article.php?story=20160413131355443 (accessed March 30, 2018).
- Nasser, F. and Fresko, B. (2002), "Faculty views of student evaluation of college teaching", *Assessment & Evaluation in Higher Education*, Vol. 27 No. 2, pp. 187-198, available at: <https://doi.org/10.1080/02602930220128751>
- National Statistics office of Taiwan (2018), "Statistical tables", available at: <https://eng.stat.gov.tw/lp.asp?ctNode=1629&CtUnit=779&BaseDSD=7&mp=5> (accessed January 1, 2018).
- Norris, D. and Baer, L. (2013), *Building Organizational Capacity for Analytics*, Educause, Louisville, CO.
- Norris, J. and Conn, C. (2005), "Investigating strategies for increasing student response rates to online delivered course evaluations", *Quarterly Review of Distance Education*, Vol. 6 No. 1, pp. 13-29.

- Nowell, C. (2007), "The impact of relative grade expectations on student evaluation of teaching", *International Review of Economics Education*, Vol. 6 No. 2, pp. 42-56, available at: [https://doi.org/10.1016/S1477-3880\(15\)30104-3](https://doi.org/10.1016/S1477-3880(15)30104-3)
- Otani, K., Kim, B.J. and Cho, J.-I. (2012), "Student evaluation of teaching (SET) in higher education: how to use SET more effectively and efficiently in public affairs education", *Journal of Public Affairs Education*, Vol. 18 No. 3, pp. 531-544.
- Otter, S. (1995), "Learning outcomes in higher education", in Burke, J. (Ed.), *Outcomes, Learning and the Curriculum: Implications for NVQ's, GNVQ's and Other Qualifications*, Falmer Press, Bristol, PA, pp. 273-284.
- Overall, J.U. and Marsh, H.W. (1979), "Midterm feedback from students: Its relationship to instructional improvement and students' cognitive and affective outcomes", *Journal of Educational Psychology*, Vol. 71 No. 6, pp. 856-865.
- Patton, T.O. (1999), "Ethnicity and gender: an examination of its impact on instructor credibility in the university classroom", *The Howard Journal of Communications*, Vol. 10 No. 2, pp. 123-144, available at: <https://doi.org/10.1080/106461799246852>
- Perry, R.P., Abrami, P.C., Leventhal, L. and Check, J. (1979), "Instructor reputation: an expectancy relationship involving student ratings and achievement", *Journal of Educational Psychology*, Vol. 71 No. 6, pp. 776-787, available at: <https://doi.org/10.1037/0022-0663.71.6.776>
- Petticrew, M. and Roberts, H. (2005), *Systematic Reviews in the Social Sciences: A Practical Guide*, Blackwell Publishers, Malden, MA.
- Picciano, A.G. (2012), "The evolution of big data and learning analytics in American higher education", *Journal of Asynchronous Learning Networks*, Vol. 16 No. 3, pp. 9-20.
- Pietersen, C. (2014), "Negotiating a shared psychological contract with students", *Mediterranean Journal of Social Sciences*, Vol. 5 No. 7, pp. 25-33, available at: <https://doi.org/10.5901/mjss.2014.v5n7p25>
- Pogue, L.L. and Ahyun, K. (2006), "The effect of teacher nonverbal immediacy and credibility on student motivation and affective learning", *Communication Education*, Vol. 55 No. 3, pp. 331-344, available at: <https://doi.org/10.1080/03634520600748623>
- Poonyakanok, P., Thisayakorn, N. and Digby, P.W. (1986), "Student evaluation of teacher performance: some initial research findings from Thailand", *Teaching and Teacher Education*, Vol. 2 No. 2, pp. 145-154, available at: [https://doi.org/10.1016/0742-051X\(86\)90013-2](https://doi.org/10.1016/0742-051X(86)90013-2)
- Powell, R.W. (1977), "Grades, learning, and student evaluation of instruction", *Research in Higher Education*, Vol. 7 No. 3, pp. 193-205, available at: <https://doi.org/10.1007/BF00991986>
- Pozo-Muñoz, C., Reboloso-Pacheco, E. and Fernández-Ramírez, B. (2000), "The 'ideal teacher': implications for student evaluation of teacher effectiveness", *Assessment & Evaluation in Higher Education*, Vol. 25 No. 3, pp. 253-263, available at: <https://doi.org/10.1080/02602930050135121>
- Pravikoff, P. and Nadasen, D. (2015), "Course evaluations simplified: the largest US public university did it and you can too", paper presented at the Annual Conference of the Association for Institutional Research, Denver, CO.
- Pulich, M.A. (1984), "Better use of student evaluations for teaching effectiveness", *Improving College and University Teaching*, Vol. 32 No. 2, pp. 91-94.
- Remedios, R. and Lieberman, D.A. (2008), "I liked your course because you taught me well: the influence of grades, workload, expectations and goals on students' evaluations of teaching", *British Educational Research Journal*, Vol. 34 No. 1, pp. 91-115, available at: <https://doi.org/10.1080/01411920701492043>
- Rice, L.C. (1988), "Student evaluation of teaching: problems and prospects", *Teaching Philosophy*, Vol. 11 No. 4, pp. 329-344, available at: <https://doi.org/10.5840/teachphil198811484>
- Richardson, J.T.E. (2005), "Instruments for obtaining student feedback: a review of the literature", *Assessment & Evaluation in Higher Education*, Vol. 30 No. 4, pp. 387-415, available at: <https://doi.org/10.1080/02602930500099193>

- Rivera, J.C. and Rice, M.L. (2002), "A comparison of student outcomes and satisfaction between traditional and web based course offerings", *Online Journal of Distance Learning Administration*, Vol. 5 No. 3, pp. 1-11, available at: www.westga.edu/~distance/ojdla/fall53/rivera53.html (accessed March 30, 2018).
- Rocco, T.S. and Plakhotnik, M.S. (2009), "Literature reviews, conceptual frameworks, and theoretical frameworks: terms, functions, and distinctions", *Human Resource Development Review*, Vol. 8 No. 1, pp. 120-130, available at: <https://doi.org/10.1177/1534484309332617>
- Rodin, M. and Rodin, B. (1972), "Student evaluations of teachers", *Science*, Vol. 177 No. 4055, pp. 1164-1166, available at: <https://doi.org/10.1126/science.177.4055.1164>
- Roehling, M.V. (1997), "The origins and early development of the psychological contract construct", *Journal of Management History*, Vol. 3 No. 2, pp. 204-217.
- Rother, E.T. (2007), "Systematic literature review × narrative review", *Acta Paulista de Enfermagem*, Vol. 20 No. 2, pp. vii-viii.
- Rousseau, D.M. (1995), *Psychological Contracts in Organizations*, Sage, Thousand Oaks, CA.
- Rousseau, D.M. (2001), "Schema, promise and mutuality: the building blocks of the psychological contract", *Journal of Occupational and Organizational Psychology*, Vol. 74 No. 4, pp. 511-541, available at: <https://doi.org/10.1348/096317901167505>
- Sander, P., Stevenson, K., King, M. and Coates, D. (2000), "University students' expectations of teaching", *Studies in Higher Education*, Vol. 25 No. 3, pp. 309-323, available at: <https://doi.org/10.1080/03075070050193433>
- Schellhase, K.C. (2010), "The relationship between student evaluation of instruction scores and faculty formal educational coursework", *Athletic Training Education Journal*, Vol. 5 No. 4, pp. 156-164.
- Shulman, L. (1987), "Knowledge and teaching: Foundations of the new reform", *Harvard Educational Review*, Vol. 57 No. 1, pp. 1-23, available at: <https://doi.org/10.17763/haer.57.1.j463w79r56455411>
- Skinner, E.A. and Belmont, M.J. (1993), "Motivation in the classroom: reciprocal effects of teacher behavior and student engagement across the school year", *Journal of Educational Psychology*, Vol. 85 No. 4, pp. 571-581, available at: <https://doi.org/10.1037/0022-0663.85.4.571>
- Sohr-Preston, S.L., Boswell, S.S., McCaleb, K. and Robertson, D. (2016), "Professor gender, age, and 'hotness' in influencing college students' generation and interpretation of professor ratings", *Higher Learning Research Communications*, Vol. 6 No. 3, pp. 1-23, available at: <https://doi.org/10.18870/hlrc.v6i3.328>
- Solomon, M.R., Surprenant, C.F., Czepiel, J.A. and Gutman, E.G. (1985), "A role theory perspective on dyadic interactions: the service encounter", *Journal of Marketing*, Vol. 49 No. 1, pp. 99-111.
- Spooren, P. and Van Loon, F. (2012), "Who participates (not)? A non-response analysis on students' evaluations of teaching", *Procedia – Social and Behavioral Sciences*, Vol. 69 No. 1, pp. 990-996.
- Spooren, P., Brockx, B. and Mortelmans, D. (2013), "On the validity of student evaluation of teaching: the state of the art", *Review of Educational Research*, Vol. 83 No. 4, pp. 598-642, available at: <https://doi.org/10.3102/0034654313496870>
- Sproule, R. (2000), "Student evaluation of teaching: methodological critique", *Education Policy Analysis Archives*, Vol. 8 No. 50, pp. 1-23, available at: <https://doi.org/10.14507/epaa.v8n50.2000>
- Staley, D.J. and Trinkle, D.A. (2011), "The changing landscape of higher education", *Educause*, pp. 16-32, available at: <http://er.educause.edu/articles/2011/2/the-changing-landscape-of-higher-education> (accessed March 30, 2018).
- Stark, P. and Freishtat, R. (2014), "An evaluation of course evaluations", *ScienceOpen Research*, available at: <https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1> (accessed March 30, 2018).
- Tagiuri, R. (1969), "Person perception", in Lindzey, G. and Aronson, E. (Eds), *The Handbook of Social Psychology: The Individual in a Social Context*, 2nd ed., Vol. 3, Addison-Wesley, Reading, MA, pp. 395-449.
- Tatro, C.N. (1995), "Gender effects on student evaluations of faculty", *Journal of Research and Development in Education*, Vol. 28 No. 3, pp. 169-173.

- Torraco, R.J. (2005), "Writing integrative literature reviews: guidelines and examples", *Human Resource Development Review*, Vol. 4 No. 3, pp. 356-367.
- Tseng, Y.-H. (2016), "Development and application of databases for institutional research and analysis (Xiàowù yánjiū zīliàokù de jiàngòu yfēnxī yīngyòng)", *Contemporary Educational Research Quarterly*, Vol. 24 No. 1, pp. 107-134, available at: <https://doi.org/10.6151/CERQ.2016.2401.04>
- Usher, A. (2009), "Ten years back and ten years forward: developments and trends in higher education in Europe region", paper presented at the UNESCO Forum on Higher Education in the Europe Region, Bucharest.
- Vlăsceanu, L., Grünberg, L. and Pârlea, D. (2004), *Quality Assurance and Accreditation: A Glossary of Basic Terms and Definitions*, United Nations Educational, Scientific and Cultural Organization, Bucharest.
- Voss, R., Gruber, T. and Szmigin, I. (2007), "Service quality in higher education: the role of student expectations", *Journal of Business Research*, Vol. 60 No. 9, pp. 949-959, available at: <https://doi.org/10.1016/j.jbusres.2007.01.020>
- Wachtel, H.K. (2006), "Student evaluation of college teaching effectiveness: a brief review", *Assessment & Evaluation in Higher Education*, Vol. 23 No. 2, pp. 191-212, available at: <https://doi.org/10.1080/0260293980230207>
- Wilson, A., Zeithaml, V.A., Bitner, M.J. and Gremler, D.D. (2012), *Services Marketing: Integrating Customer Focus Across the Firm*, 2nd European ed., McGraw-Hill Education, Berkshire.
- WTO (1998), "Education services", Document No. S/C/W/49 98-3691, World Trade Organization, Geneva, available at: www.wto.org/english/tratop_e/serv_e/w49.doc (accessed March 30, 2018).
- Wright, P., Whittington, R. and Whittenburg, G.E. (1984), "Student ratings of teaching effectiveness: what the research reveals", *Journal of Accounting Education*, Vol. 2 No. 2, pp. 5-30, available at: [https://doi.org/10.1016/0748-5751\(84\)90002-2](https://doi.org/10.1016/0748-5751(84)90002-2)
- Wu, P.-M. (2018), "The declining birthrate threatens. The number of college freshmen is reduced to only 100,000 after 10 years (shǎo zhūà fā wēi dà zhuān xiào yuàn xīn shēng 10 nián hòu jiǎn jìn 10 wàn rén)", available at: <https://udn.com/news/story/7266/3156797> (accessed March 30, 2018).
- Zeithaml, V.A. (1981), "How consumer evaluation processes differ between goods and services", in Donnelly, J.H. and George, W.R. (Eds), *Marketing of Services*, American Marketing Association, Chicago, IL, pp. 186-190.
- Zhang, Y.-W. (2003), "Development of student instructional rating scale (Dàxuéshēng jiàoxuépíngjiǎnliángbiǎo zhī fāzhǎnyánjiū)", *Journal of Education and Psychology*, Vol. 26 No. 2, pp. 227-239.

Corresponding author

Gregory Ching can be contacted at: gregory_ching@yahoo.com