Journal of World Englishes and Educational Practices
ISSN: 2707-7586
DOI: 10.32996/jweep
Journal Homepage: www.al-kindipublisher.com/index.php/jweep

JWEEP
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Students' Evaluation of Teaching (SET) for Improving Learning and Teaching Quality in HE: Students' and Teachers' Perspectives

Ibtisam Al Kharusi
Lecturer, English Language Center, University of Technology and Applied Science, Muscat-Oman
Corresponding Author: Ibtisam Al Kharusi, **E-mail**: ibtisamh85@gmail.com

| ABSTRACT
Various academic institutions in higher education (HE) allocate their resources differently to improve the quality of teaching and learning. To measure and identify the teaching and learning quality in any educational context, institution performance in this field must be evaluated using different techniques and instruments, one of which is Student Evaluation of Teaching (SET). The study aims to determine the use of SET to improve teaching quality based on the opinions of teachers (n = 6) and students (n = 413). This is achieved by investigating 1) the degree of alignment or divergence between teachers' and students' ratings of a particular course, 2) how teachers value students' feedback on teaching to improve their teaching practises, 3) students' perspectives on using SET in improving teaching quality and 4) the implications for implementing SET effectively in HE context. These were addressed using an adapted version of the Students' Course Evaluation Questionnaire (SCEQ), and the Mann-Whitney test was used to analyse the data. The findings reveal a statistically significant correlation coefficient (r = 0.739) between teachers' and students' evaluations. The associated p-value (0.058) is slightly greater than the conventional threshold of $p \leq 0.05$. The results of the Mann-Whitney U test indicated an overall concurrence between teachers' and students' evaluations, although certain inconsistencies were observed. The findings indicate a consensus among teachers and students regarding SET's efficacy in enhancing instructional quality.

| KEYWORDS

Student Evaluation of Teaching (SET), evaluation, quality assurance, teaching practises, higher education, performance indicators

| ARTICLE INFORMATION

ACCEPTED: 24 November 2023 **PUBLISHED:** 04 December 2023 **DOI:** 10.32996/jweep.2023.5.3.8

1. Introduction

Almost all HE systems and different academic institutions worldwide have undergone substantial reforms aimed at enhancing their quality improvement systems (Chalmers, 2008). The quality system implemented within a specific academic institution must prioritise delivering high-quality performance, which can be achieved by cultivating a skilful workforce to enhance labour markets and economic growth. Regarding the national level, a particular institution's performance should be evaluated by a national quality system that provides a standardised audit and review of the organisation's practices, processes, and outcomes. Since the intensifying impact of the global environment cannot be disregarded, there has been a growing concern about HE global comparison, benchmarking, and ranking of institutions and systems (Marginson & van Der Wende, 2007). Thus, this study presents international comparisons of performance indicators of HE systems in different colleges and universities worldwide, some of which have been referenced by UNESCO (Chalmers, 2008).

Although many countries have successfully implemented SET to evaluate the teaching quality in HE, the use of SET has yet to yield the desired outcomes in some countries. In Oman, the significance of SET has been limited due to the prevailing belief that students are unable to offer assessments of teaching quality that are both reliable and valid (Al-Hinai, 2011). Thus, it is imperative to investigate the utilisation of SET in Oman since all Higher Education Institutions (HEIs) are undergoing an institutional accreditation

process. According to the standardised assessment criteria, HEIs are expected to demonstrate satisfactory levels of teaching quality (Oman Academic Accreditation Authority, 2016), which can be achieved through meaningful input collected from SET. Institutions, even those that have obtained accreditation, are required to undergo the Standards Assessment stage every five years, indicating a continuous reflection on teaching quality is necessary (ibid, 2014). Hence, the aforementioned justifications underscore the imperative nature of exploring the utilisation of SET as a means to improve teaching quality. This study has undertaken an examination of the various concepts associated with SET. In doing so, it has considered the perspectives of individuals directly involved in the educational process, namely students and teachers. By exploring these perspectives, this research aims to establish a foundation for administrators to implement practical and constructive measures to enhance teaching practises within their respective educational institutions.

2. Literature Review

As with any other organisation, HEIs are displaying a growing inclination towards enhancing their quality system to monitor performance in different units and areas, mainly the teaching-learning process. Academic units in HEIs must prioritise their time and effort to monitor and reflect on teaching quality (Harris et al., 1997) using multidimensional metrics and performance indicators. Unfortunately, research claims that active implementation of quality assurance (QA) in higher academic institutions has faced many challenges (Mårtensson et al., 2014; Chalmers, 2008). One of these challenges is legitimising QA effects, which have been shown to be more related to governance than teaching and learning quality (Stensaker, 2008). Furthermore, Newton (2000) claims that it is challenging for academic staff to understand how QA perspectives relate to academic life. They find it difficult to establish a connection between QA principles and teaching and learning quality. As a result, there is growing interest in implementing QA standards to address and monitor the effectiveness of HE by evaluating learning outcomes and proposing a qualification framework (Prøitz, 2010). Thus, QA stakeholders and policymakers should consider these challenges when policies related to teaching and learning quality are anticipated. In addition, these policies' rationale and procedure must be unambiguous and concise for academics to stimulate the 'quality practice'.

2.1. Types of Key Performance Indicators

Generally, there are four core performance indicators to evaluate the quality of practises: input, process, output, and outcome (Harris et al., 1997). According to Chalmers (2008), these can also be classified as qualitative and quantitative indicators.

2.1.1. Quantitative indicators

Quantitative indicators are those used to measure input and output performance based on quantity or numerical values. First, input performance indicators include the organisation's human, financial, and physical resources. Chalmers (2008) pointed out the challenges in determining the quality of teaching and learning as it requires a clear interpretation; for instance, enrollment data can be understood by student ratio as an input indicator. The second is output performance indicators, which represent final measurable outcomes (Chalmers, 2008) and any organisational activities or practises that directly contribute to the achieved outcomes (Burke, 1998). The number of graduates and degrees awarded are examples of output performance indicators (Burke & Minassians, 2002).

However, there is an argument that quantitative performance indicators do not reveal much about teaching and learning quality despite being less controversial (ibid, 2002). Stakeholders and developers may find limited opportunities for observation and comprehension due to the inherent limitations of quantitative data. Unlike qualitative data, which offers detailed insights into instructional practises, collaboration, and learning processes within an educational programme, quantitative data does not facilitate a thorough examination of these aspects.

2.1.2. Qualitative indicators

Qualitative indicators refer to observations and descriptions of quality or non-numerical results, such as policies or processes for assessing teaching and learning quality. Qualitative indicators comprise outcome and process indicators. Outcome indicators evaluate the quality of educational programmes and services for students, parents, employers, and society (Burke, 1998; Yorke, 2016). These indicators do not deal with numerical data, as with output performance indicators; instead, they highlight the impact and quality of final results. Both output and outcome performance indicators measure the final effects and achievements of HE programmes; however, while the former indicators evaluate them quantitatively, the latter evaluate them qualitatively.

On the other hand, process indicators relate to the ways and means of delivering activities and services provide insight into system operation in an academic context (Burke, 1998; Yorke, 2016). These indicators deal with all practises, procedures, and policies linked to teaching and learning quality, including curriculum development and student assessment. In addition, these indicators provide a comprehensive view by highlighting strengths and weaknesses for enhancing the quality of different practises.

2.2. Using SET in HEIs

The effectiveness of teaching practises can be evaluated by teachers (self-evaluation), administrators (class visit), colleagues (peer evaluation), external or trained observers, or students themselves. In reality, the implementation of SET is a widely adopted practise among HEIs worldwide. Since the data obtained from SET can efficiently improve teaching at the programme, department, and college levels in general, it has been used as the main tool to assess teaching quality in many universities and colleges (Otani et al., 2012). Boysen (2016) added that academics support using SET because of its positive effects on pedagogical enhancement.

2.2.1. Multidimensionality of SET

Educators and researchers generally concur that teaching is a complex practise involving different dimensions, such as student-teacher interaction, organisation, class management, teaching and learning styles and course outline (Cashin and Downey, 1992; Duy et al., 2008; Friedrich, 1998; Marsh et al., 1997; Marsh & Roche, 1993, 1997; Zerihun et al., 2011). Marsh (1983:150) clarified that "common senses and a considerable body of empirical research indicate that students' evaluations are multidimensional". Therefore, the multidimensionality of teaching has to be reflected when it is evaluated. In addition, detailed and specific feedback is required for teachers to develop and improve their pedagogical skills.

Standardised SET instruments have been developed and are widely used in HEIs. For example, Students' Evaluation of Educational Quality (SEEQ) is based on nine factors: learning/value, instructor enthusiasm, organisation and clarity, group interaction, individual rapport, breadth of coverage, examinations and grading, assignments and readings, and workload and difficulty (Marsh, 1983; Marsh et al., 1997; Marsh & Roche, 1997). Another standardised SET survey is the SCEQ, which comprises five categories: good teaching, clear goals and standards, appropriate assessment, appropriate workload, and generic skills (Boysen, 2016). In addition to these three, the Instrumental Development and Effectiveness Assessment (IDEA), which consists of four main categories (instructor methods, student progress rating on course objectives, course description, and students' self-rating), has been acknowledged as a standardised SET instrument (Boysen, 2016; Cashin & Downey, 1992). Thus, the multidimensionality of teaching practises is readily apparent in these internationally recognised surveys widely used to evaluate teaching effectiveness in educational institutions.

However, there has been an argument about how to weigh each SET component and decide if specific categories are more important than others. In this regard, Marsh and Roche (1997) suggested deciding on each component's weight according to its importance in a specific teaching context.

2.3. Validity and reliability of SET

The validity and reliability of students' evaluations have become a subject of growing concern and debate in light of their significant impact on academic careers (Al-Hinai, 2011; Boysen, 2016; Cashin & Downey, 1992; Friedrich, 1998; Kwan, 2000; Marsh & Roche, 1997). Based on the numerous studies conducted, the prevailing consensus suggests that SET is reasonably valid and reliable (Feldman, 1989, Kwan, 2000).

Notably, there is a positive correlation between SET results and the various instruments used to evaluate factors related to teaching effectiveness (Kwan, 2000; Marsh, 1983; Marsh and Roche, 1997). Some empirical research proved that SET agreed with teachers' self-evaluation with a correlation of 0.29 overall rating (Feldman, 1989) and external observation (Murray, 1983). In contrast, based on class visits, colleagues' and administrators' evaluations did not correlate with SET (Koon and Murray, 1995) or teachers' self-evaluation (Feldman, 1989).

Another way to validate the use of SET is by looking at its stability. For instance, in one of the very earliest studies of SET effectiveness, "correlations of 0.87 and 0.89 between students' rankings of the quality of their teachers from one year to the next" were found, according to Guthrie (1954:512, as cited by Costin et al., 1971). In addition, Krantz-Girod et al. (2004) investigated the stability of students' evaluations of 37 academics over two successive academic years, and the findings showed that the students' evaluations were stable, reliable, and valid throughout the study. Similarly, Marsh and Hocevar (1991) examined the ratings of 195 teachers from 31 different academic departments evaluated over 13 consecutive academic years. Amazingly, there was no significant change in students' ratings over time, neither in course evaluation nor in teachers' ratings. These findings suggest that SET can be a valid tool for improving curriculum and teaching quality.

3. Methodology

3.1. The rationale and significance of the study

The growing concern about using SET to evaluate teaching quality in HE has piqued the interest of educators and researchers. As a result, many studies have found that implementing SET to evaluate and improve teaching practises and quality in HEIs is effective (Alhija and Fresko, 2009; Al-Hinai, 2011; Duy et al., 2008; Ferguson, 2012; Hay and Van der Merwe, 2010; Johnson, 2000; Marsh et al., 1997; Marsh and Roche, 1993, 1997).

In Oman, it is commonly hypothesised that SET is not a valid indicator because it is associated with "distrust and suspicion" (Al-Hinai, 2011:ii). One possible explanation is the belief that students lack the necessary skills and experience to evaluate teaching quality fairly. Furthermore, Omani HEIs are undergoing accreditation, with one of the requirements being an ongoing evaluation of teaching where the effective implementation of SET is required (OAAA, 2014). The recent accreditation experience of HE in Oman makes it an ideal context to investigate the use of SET, not only for accreditation attainment but also to update and maintain national standards that are internationally benchmarked. An accredited institution has to undergo the Standards Assessment stage every five years to maintain its accreditation status (ibid, 2014). Furthermore, the study is generally helpful for institutions new to the QA system and seeking accreditation, as well as those who want to reflect, update, and improve their existing quality system. Thus, examining this topic will help conceptualise the use of SET, especially with the limited literature on this topic.

The research investigated students' and teachers' perspectives on using SET to improve teaching quality in the Omani HE context. The research seeks to answer the following research questions:

Research Question 1: To what extent do teachers' and students' evaluations of teaching practises in HE in Oman match or mismatch?
Research Question 2: How do teachers value students' feedback on teaching to improve their teaching practises?
Research Question 3: What are students' perspectives about using SET to improve the college's teaching quality?
Research Question 4: What are the implications for HEIs to implement SET effectively in Omani HE?

3.2. Data Collection and Analysis

3.2.1. Instrument

This study used an adapted version of the SCEQ. The questionnaire items were adapted to ensure their suitability for the assessment of a writing course. The questionnaire comprises two parts: scaled items to collect quantitative data and open-ended questions, which are considered qualitative data (Boysen, 2016). This was considered because when multidimensional and complex human behaviours (e.g., teaching and learning) are investigated, the mixed method is necessary to provide a clear insight into participants' perspectives (Drew, 2008; Mackey and Gass, 2015; Teddlie, 2009; Walliman, 2006).

The questionnaire's first part included 22 Likert-type items presented on a 5-point scale from 1 (disagree) to 5 (strongly agree). These items were categorised into five main factors to evaluate teaching effectiveness: good teaching scale (6 items), clear goals and standards scale (3 items), appropriate assessment scale (3 items), appropriate workload scale (4 items), and generic skills scale (4 items). Based on Cashin's and Downey's (1992) recommendation to include an overall rating in SET, two items were included in the survey, asking about overall satisfaction with teachers' performance and course materials. In the second part of the questionnaire, some open-ended questions allowed participants to express their thoughts about the topic since these questions "can catch the authenticity, richness, depth of responses, honesty and candour" (Cohen et al., 2017:475). Furthermore, participants' responses will not be only the researcher's choices; rather, they will be able to respond spontaneously (Hong, 1984), which might highlight some unexpected dimensions and concerns leading to further investigation. Thus, the researcher ensured a balance between objective ratings and subjective comments.

There were two versions of the questionnaire: the students' version and the teachers' version. In the students' version, the components were translated into their mother tongue, Arabic, to ensure students understood each item and encouraged them to answer open-ended questions freely, using English or Arabic. In the teachers' version, teachers were asked to evaluate teaching quality (self-evaluation) on the same questionnaire items as their students did. Students' evaluations were compared to teachers' self-evaluations since self-reflection is preferred in many institutions as it is easy, applicable to all education settings, and persuasive to teachers (Marsh & Roche, 1997). In addition, some research found that SET correlated positively with teachers' self-evaluation (Feldman, 1989).

Paper-based questionnaires were used, and students were administered in groups, while teachers were administered one-on-one. A printed questionnaire was chosen for several reasons. First, the chance of getting a reasonable response rate is significantly higher than for other types (Dörnyei & Taguchi, 2009). Second, the researcher's presence during questionnaire administration is an advantage as any clarification or concern can be answered, and participants would be more motivated to respond honestly, according to Strange et al. (2003, as cited by Dörnyei & Taguchi, 2009).

3.2.2 Participants

The population was stratified based on specific characteristics using proportional stratification sampling (Creswell, 2013; Gay, 1987). Participants included students and their writing teachers in a standard writing course. In order to obtain a confidence level of 95% with a margin error of 5%, the selected sample from a total population of 957 students should be ≥ 275. Four hundred and twenty students were chosen. For comparison, all full-time writing teachers teaching the course for over three years were targeted to

maintain a confidence level of 95%. Six teachers were chosen because they all taught students included in the sample. Since six teachers participated in the study, the students' sample was divided into six sub-groups. In order to sustain the validity of the collected data from these six sub-groups, a minimum of 10% of the descriptive data was collected from each cluster (Gay and Diehl, 1992).

Oversampling was considered to avoid incomplete, spoiled, or lost questionnaires, as it is recommended by Cohen et al. (2017). One of the techniques to deal with missing data is list-wise deletion (ibid, 2017), which was implemented in this study. Thus, the questionnaires with missing responses in the closed-ended questions were discarded. Only eight cases had missing data, which did not affect the investigated cases.

3.2.3. Data Analysis

As the questionnaire was meant to collect quantitative and qualitative data, the data were computed and analysed using appropriate techniques. Quantitative data were analysed using Excel 2016 with add-ins from Analysis ToolPak to obtain descriptive statistics like ANOVA, means, standard Deviation (SD), multiple r, p-value, and frequencies because of its easy and effective built-in tools (Dunlop, 2015). In addition, the statistical software SPSS was used to calculate the Mann-Whitney test to obtain the difference between students' and teachers' perceptions, as the two groups were not customarily distributed (Cohen et al., 2017). Descriptive data is crucial because it provides both the range and meaning of variables and percentages, which are easier to understand than raw frequencies (Dörnyei & Taguchi, 2009).

For qualitative data, the obtained data was re-written by selecting repeated and essential statements, words, phrases and patterns, and then categorised and sorted into themes to develop core categories through examining, conceptualising and categorising the data and responses (Wilson, 2013; Creswell, 2013). A cross-sectional approach to analysis was used to analyse, interpret, and correlate the collected data. The difference between a teacher's and students' evaluations was investigated, and their responses were analysed and compared.

3.3. Ethical Considerations

The ethical issues related to social research were taken into consideration. The ultimate goal of research ethics is to balance research demands to purify the truth and respondents' rights, which could be threatened by the research procedures (Cohen et al., 2017). Considering these ethical issues, ethical approval was obtained from the university. Furthermore, the Participant Information Sheet was provided, and participants were asked to sign a Constant Form to acknowledge their approval to participate in the study. Confidentiality was ensured, and anonymity was secured through a coding system followed in data analysis.

4. Results and Discussion

The interpretation and discussion of the results are presented in line with the research questions to create a clear connection between the research objectives and the obtained data (Cohen et al., 2017).

4.1. Academics' and students' evaluations of teaching practises in HE

Statistical descriptive data and the Mann-Whitney U test were calculated, discussed, and interpreted. The differences between the evaluations of six different groups of students and the differences between teachers and their students were also considered. The following table (Table 1) presents the descriptive statistics of students' and teachers' ratings on five scales in the questionnaire to investigate differences between students' and teachers' ratings.

Table 1: The descriptive statistics on students' and teachers' ratings on different scales in the questionnaire

	Students		Teachers		Students-teachers		
Statements	**AVG.**	**SD**	**AVG.**	**SD**	**SD of AVG.**	**Significance F p-value**	**Multiple R correlation**
Good teaching scale	4.2*	0.13	4.64	0.267	0.307		
Clear goals and standards scale	4.05	0.296	4.67	0	0.387		
Appropriate assessment scale	3.56	0.107	3.83	0.333	0.267		
Appropriate workload scale	3.07	0.271	4.04	0.699	0.716		
Generic skills scale	3.47	0.391	4.17	0.236	0.479		
Overall rating							
Overall teachers rating	4.16	0**	4.33	0**	0.135	0.058	0.739
Overall course rating	3.32	0**	3.33	0**	0.01		

* 1 = strongly disagree, 2 = disagree, 3 = not sure, 4 = agree and 5 = strongly disagree.
** SD = 0 of dependent variables because they are evaluated using one single item; p≤.05.

Generally, students' and teachers' responses did not vary much throughout the five scales and the overall rating. The highest SD of the average of their responses was in the appropriate workload scale of 0.716, followed by a generic skills scale of 0.479. Interestingly, the SDs of overall ratings of both students and teachers were significantly low: SD = 0.01 for course rating and SD = 0.135 for teachers' rating. These results indicate a significant agreement between teachers' and students' evaluations of course content over teaching performance. The coefficient of multiple correlations of the average of both teachers' and students' responses indicates a significant positive correlation as r = 0.739, more than the 0.29 found by Feldman (1989). This can be interpreted as the linear relationship of variables indicating a strength of 73.9%. Moreover, the p-value of 0.058 ≥ 0.05 does not show any statistical difference between the average of both ratings.

In order to provide a clear insight into the significance of the variance between teachers' and students' evaluations, both ratings were compared using an independent t-test, the Mann-Whitney U Test1. The test results (Table 2) indicate a significant agreement between the two evaluations since the values of the 2-tailed (medians) of different groups ranged from 0.081 to 0.946 ≥ P. However, in the sixth sub-group, the teachers and students rated two items under the excellent teaching scale differently, with 0.009 for both.

Scales/Statements*	Asymp. Sig. (2-tailed) – P-value						
	T-SS 1**	T-SS 2	T-SS 3	T-SS 4	T-SS 5	T-SS 6	Overall
Good Teaching Scale							
1. Lecturer uses different teaching styles	0.312	0.54	0.229	0.694	0.317	0.159	0.724
2. Lecturer is willing to help	0.578	0.614	0.365	0.344	0.876	0.009	0.625
3. Lecturer is excellent in explaining the course content	0.081	0.344	0.183	0.903	0.838	0.009	0.835
4. lecturer encourages students to express themselves freely	0.475	0.497	0.214	0.606	0.122	0.643	0.085
5. lecturer motivates students	0.518	0.554	0.258	0.254	0.193	0.612	0.031
6. Lecturer gives helpful feedback	0.594	0.367	0.199	0.902	0.146	0.185	0.208
Clear Goals and Standards Scale							
7. Course objectives are clear	0.829	0.203	0.176	0.191	0.189	0.771	0.022
8. Course objectives were discussed at the beginning of the semester	0.21	0.381	0.233	0.349	0.628	0.69	0.498
9. The standards of the expected work in the course are clear	0.935	0.262	0.211	0.206	0.928	0.335	0.056
Appropriate Assessment Scale							
11. The assessment methods are appropriate	0.345	0.936	0.177	0.697	0.378	0.901	0.666
10. Exams questions are appropriate and related to course content	0.574	0.332	0.737	0.785	0.946	0.657	0.799
12. The feedback on exams and quizzes is valuable	0.489	0.159	0.552	0.625	0.626	0.863	0.114
Appropriate Workload Scale							
13. The workload of the course is heavy	0.367	0.311	0.093	0.404	0.412	0.124	0.797
14. Many hours are required to study	0.119	0.401	0.095	0.36	0.367	0.116	0.002
15. The course pace is fast	0.467	0.654	0.161	0.365	0.479	0.165	0.021
16. Enough time is given to comprehend the course content.	0.397	0.221	0.104	0.31	0.226	0.106	0.002
Generic Skills Scale							
17. The course develops teamwork skills.	0.473	0.226	0.188	0.453	0.263	0.626	0.086
18. The course develops analytical skills.	0.665	0.711	0.137	0.116	0.4	0.767	0.036
19. The course develops problem-solving skills.	0.831	0.736	0.108	0.345	0.267	0.522	0.121
20. The course develops writing skills.	0.698	0.603	0.269	0.129	0.926	0.552	0.571
Overall Teachers Rating							
21. Overall satisfaction with teaching performance	0.088	0.499	0.222	0.987	0.567	0.057	0.909
22. Overall satisfaction with course content	0.368	0.736	0.379	0.509	0.442	0.223	0.997

95% Confidence Interval, p≤.0.05
*statements were phrased to indicate the idea of each in short
**The test results are of the six sub-groups (teachers and students) based on their evaluation of the adopted SCEQ
*** T-SS: Teacher + his/her students (group-wise)

Table 2: The results of 2-tailed values of teacher-students groups (mapped) according to the Mann-Whitney U Test

The overall evaluations of the students and their teachers revealed a consensus, although certain discrepancies were observed. On the appropriate workload scale, participants' perceptions were different about course peace (p≥0.021), the time required to study outside (p≥0.002), and the class and time needed to comprehend the course content (p≥0.002). The workload scale needs to be investigated more, as the difference occurred in three items out of four. In addition to these, participants' evaluations varied in three other items related to the good teaching scale (p≥0.031), clear goals and standards scale (p≥0.022), and generic skills scale

1 The results of the Mann-Whitney U Test in details are in Appendix I

(p≥0.036). Interestingly, participants' perspectives on overall teaching performance and course content evaluation did not indicate any significant difference with the 2-tailed values of 0.909 and 0.997, respectively; instead, they showed solid agreement. Evidently, participants'perspectives differed in some aspects related to the different scales but not in the overall ratings, either group-based or as a whole. Participants may have lacked awareness regarding the multidimensional questionnaire and its coverage of various scales.

The findings revealed that students' evaluations of teaching quality significantly matched teachers' evaluations, with a strength of 73.9% found between the two evaluations and a p-value = 0.058 ≥ 0.05, indicating a significant parallel between both ratings. In addition, exact matching was exposed based on the Mann-Whitney U Test; however, there were variances in some scales, especially the workload scale. According to the findings of the average, SD, and t-test, the ratings of both groups presented a clear mismatch on this particular scale, indicating the need to examine the topic further.

4.2 Academics' perceptions of SET to improve teaching practises

Teachers were asked to rank five different indicators of teaching performance in terms of importance to provide details about their perceptions of using SET for teaching quality improvement. The frequencies of teachers' responses to different evaluation methods of teaching quality were calculated to obtain the ranking (Table 3).

Statements	Ranking
Students' Evaluation	1
Self-evaluation	2
Peer's Evaluation	3
External Visitor Evaluation	4
Administrator's Evaluation	5

Table 3: Teachers' ranking of preferred ways to evaluate teaching performance and quality

The results in Table 3 indicate that teachers agreed and preferred students' evaluations compared to other evaluation methods, as they were ranked first, followed by teachers' self-evaluation. In contrast, adminstration evaluation was the least preferred. These results align with Boysen (2016), who claimed that academics tend to favour the use of SET due to its positive impact on improving teaching quality. When asked how they would use students' feedback to improve their teaching performance, all teachers agreed that SET is essential for developing and improving teaching practises. First, SET helps identify different students' needs. This is grounded in the belief that students "are aware more than anyone else about their needs", as reported by Teacher A. Another reason is that students' different learning styles can be easily met using certain teaching styles; without students' feedback, these styles would not be recognised. As Teacher B stated, teaching is all about communicating "knowledge the way the students want to learn".

Some academics have expressed their interest in SET by providing valuable suggestions. Teachers were unanimous in their desire to implement SET every semester. One of the proposals was to develop a mobile application for students to evaluate their teachers at the end of each semester. Another suggestion was arranging for a meeting between students and teachers so that students could express their concerns and opinions about the teaching quality. Due to limited resources, one teacher suggested that students evaluate their teachers by noting their strengths and weaknesses. These intentions indicate how teachers are inspired by the advantages of using SET for pedagogical development.

On the other hand, the honesty of students' feedback was disputed among teachers. Some teachers argue that students are honest in their feedback because they seek a better teaching-learning experience. On the contrary, some teachers claimed that students may lack the knowledge and technical aspects of SET, which might result in biased feedback. Moreover, students' evaluations can be affected by their grades, levels, or different academic experiences.

4.3 Students' perspectives on using SET to improve teaching quality

Based on students' responses, many students claimed that they are "mature enough" and able to or have to evaluate teachers fairly since they are the ones who will benefit if teaching performance improves. From students' perceptions, there are different reasons why SET should be implemented effectively. First, and according to the majority, SET is students' right, as learners are part of the teaching-learning process, and they are the ones whom teachers teach; thus, teachers' performance must be evaluated by them. They know that their productive feedback enhances teaching, which will benefit them. Furthermore, SET might encourage

academics to do their best because of the expected evaluation at the end of the semester. Some students explained that as they are evaluated by the end of the academic semester, teachers should expect the same.

Moreover, some students pointed out that SET is a way to express their opinions and concerns to administrators about academic performance and to take necessary actions if needed. These findings are consistent with those of Chen and Hoshower (2003), who proved that students are optimistic about using SET to develop teaching effectiveness. Referring to students' perceptions of using SET, in this study, students' feedback can be categorised as formative and summative teaching indicators. As a formative assessment, SET helps teachers improve their teaching practises and provides valuable feedback for administrators and stakeholders as a summative assessment. On the other hand, many students claimed that the administration had not taken their evaluation seriously. Based on their feedback, they are unsure if any action has been taken to improve the teaching practises. Ignoring students' evaluations might prevent students from providing valuable input to improve teaching effectiveness (ibid, 2003).

4.4 The practical implications of SET in HE

Based on the study findings, the following are essential implications for SET policy and practise. First, students' evaluations should be given more weight when evaluating teachers' performance since students' and teachers' perspectives about teaching quality revealed a significant agreement, demonstrating students' ability and readiness to evaluate teaching effectively. In addition, teachers and students supported the use of SET to improve teaching practises. Thus, SET has been regarded as an influential measure of teaching effectiveness, and "an active participation of meaningful input from students can be critical in the success of such teaching evaluation systems" (Chen & Hoshower, 2003:71). Applying students' feedback to improve teaching practises is more convincing and meaningful for teachers than any other method of evaluation that directly assists academics in implementing quality in their teaching practises. In this way, teachers will establish a rapport between academic practises and QA, which is challenging in many HE contexts (Newton, 2000). Thus, a higher percentage should be allocated to students' evaluations to provide constructive feedback on teaching quality. For example, instead of minimising SET value in academic staff appraisal reports to 30% for new staff and 20% for experienced staff, more percentage should be allocated to students' evaluation.

In addition, teachers and students should be oriented and trained about the technical use of SET to improve teaching quality since their teaching quality ratings reveal inconsistencies in some scales. This might be because they lack knowledge about the nature of multidimensional evaluation. The training should explain the aspects of using SET and its multidimensionality. Thus, students can develop a better understanding of how to evaluate teaching effectiveness honestly. Teachers will also be aware of how they are evaluated and will be more confident in dealing with students' feedback

The empirical results reveal that SET is valuable for improving teaching quality in HEIs. However, students argued that their feedback had not been taken seriously and that evaluating teaching effectiveness in the college is pointless when no action is taken. Because the quality of students' feedback should be anticipated, there is an urgent need to construct a trusted relationship between students and the administration. If students' evaluation is ignored and unconsidered, they will be demotivated to provide a meaningful evaluation.

5. Conclusion

Different empirical studies, including this one, have shown the valuable input of SET in constructing normative data for teaching improvement. Based on the results of the present study, students' and teachers' perspectives have demonstrated a significant agreement on implementing SET to improve teaching effectiveness.

The statistical findings indicate that students' and teachers' responses have no significant differences, and a general matching between the two ratings was interpreted. In addition, teachers' perspectives about SET were encouraging, as they rated it as the most effective and preferred instrument for improving teaching practises. Similarly, students' perceptions were similar to those of their teachers. Based on the study's findings, some implications can be implemented. First, the effectiveness of SET should be considered, and more weight should be allocated to it compared to other evaluation methods. In addition, a training programme could be proposed for students and teachers to learn about different aspects of SET. Moreover, the administration must demonstrate a genuine interest in receiving student feedback to foster an environment encouraging them to provide valuable input in their evaluations. This feedback can then be utilised to enhance pedagogical practises.

The present research is not without its limitations, several of which became evident during the study. The first limitation is that the sample is small and taken from a single university in a particular geographic location due to time and capacity constraints. Moreover, dealing with writing courses may restrict the generalisation of the results to other courses and contexts. In addition, my status as a researcher and insider in my research context as a colleague might prevent participants from expressing themselves freely. Some might be defensive to comment or share any concern on the investigated topic, which could affect the results' transparency (Creswell, 2018). One significant limitation was the role of the researcher as a participant due to the limited number

of teachers who could participate. In order to avoid biased data, I tried to remain neutral as much as possible in answering the questionnaire. Also, my students' feedback was collected by another teacher to ensure unbiased data.

Overall, using SET is promising for improving teaching quality in HE. The study findings could be extended by highlighting the need for additional research on some related topics:

- The gap between SET policies and their implementation procedures
- SET and its stability and usefulness
- Practical interpretation of SET results for pedagogical and administrative decisions
- How to decide the scales to be evaluated in SET in an institutional context

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
ORCID: https://orcid.org/0009-0009-8744-3456
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Alhija, F. N.-A., & Fresko, B. (2009). Student evaluation of instruction: What can be learned from students' written comments? *Studies in Educational Evaluation*, *35*(1), 37–44.
[2] Al-Hinai, N. S. (2011). *Effective college teaching and students' ratings of teachers: What students think, what faculty believe, and what actual ratings show: implications for policy and practice in teaching quality assurance and control in higher education in Oman* [Ph.D., Durham University].
[3] Boysen, G. A. (2016). Using student evaluations to improve teaching: Evidence-based recommendations. *Scholarship of Teaching and Learning in Psychology*, *2*(4), 273–284.
[4] Burke, J. C. (1998). Performance Funding: Arguments and Answers. *New Directions for Institutional Research*, *1998*(97), 85–90.
[5] Burke, J. C., & Minassians, H. P. (2002). Reporting indicators: What do they indicate? *New Directions for Institutional Research*, *2002*(116), 33–58.
[6] Cashin, W. E., & Downey, R. G. (1992). Using global student rating items for summative evaluation. *Journal of Educational Psychology*, *84*(4), 563–572.
[7] Chen, Y., & Hoshower, L. B. (2003). Student Evaluation of Teaching Effectiveness: An assessment of student perception and motivation. *Assessment & Evaluation in Higher Education*, *28*(1), 71–88.
[8] Cohen, L., Manion, L., & Morrison, K. (2017). *Research Methods in Education*. Routledge.
[9] Costin, F., Greenough, W. T., & Menges, R. J. (1971). Student Ratings of College Teaching: Reliability, Validity, and Usefulness. *Review of Educational Research*, *41*(5), 511–535.
[10] Creswell, J. W. (2013). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson Education Limited.
[11] Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in Second Language Research: Construction, Administration, and Processing*. Routledge.
[12] Drew, C. J. (2008). *Designing and conducting research in education*. SAGE.
[13] Dunlop, N. (2015). *Beginning Big Data with Power BI and Excel 2013 Big Data Processing and Analysis Using PowerBI in Excel 2013* (1st ed. 2015.). Apress : Imprint: Apress.
[14] Duy, B., Çivitçi, A., & Aksu, M. B. (2008). College Students' Perceptions about Teaching Practices, Classroom Behaviors and Attitudes of the Faculty Members. *Journal of Inonu University Faculty of Education*, *9*(16), 17–42.
[15] Feldman, K. A. (1989). Instructional effectiveness of college teachers as judged by teachers themselves, current and former students, colleagues, administrators, and external (neutral) observers. *Research in Higher Education*, *30*(2), 137–194.
[16] Ferguson, R. F. (2012). Can Student Surveys Measure Teaching Quality? *Phi Delta Kappan*, *94*(3), 24–28.
[17] Friedrich, J. (1998). Teaching evaluations: Concerns for psychologists? *American Psychologist*, *53*(11), 1226–1227.
[18] Gay, L. R. (1987). *Educational research: Competencies for analysis and application* (3rd ed.). UK: Merrill Publishing Co.
[19] Gay, L. R., & Diehl, P. L. (1992). *Research methods for business and management*. Canada: Maxwell Macmillan International.
[20] Harris, A., Bennett, N., & Preedy, M. (Eds.). (1997). *Organisational Effectiveness and Improvement in Education*. Open University Press.
[21] Hay, H., & Van der Merwe, B. (2010). The role of student evaluation in improving the Quality of teaching and learning practices at the Central University of Technology, Free State: A case study. *South African Journal of Higher Education*, *21*(5).
[22] Hong, L. K. (1984). List processing free responses: Analysis of open-ended questions with word processor. *Qualitative Sociology*, *7*(1), 98–109.
[23] Johnson, R. N. (2000). *A qualitative study of student feedback: Lecturers' and Students' perception experiences* [Doctoral, Sheffield Hallam University].
[24] Koon, J., & Murray, H. G. (1995). Using Multiple Outcomes to Validate Student Ratings of Overall Teacher Effectiveness. *The Journal of Higher Education*, *66*(1), 61–81.
[25] Krantz-girod, C., Bonvin, R., Lanares, J., Cueánot, S., Feihl, F., Bosman, F., & Waeber *, B. (2004). Stability of repeated student evaluations of teaching in the second preclinical year of a medical curriculum. *Assessment & Evaluation in Higher Education*, *29*(1), 123–133.

[26] Kwan, Kam-Por. (2000). *How university students rate their teachers: A study of the attitudes and rating behaviors of university students in teaching evaluations.* [Doctoral, Durham University].
[27] Mackey, A., & Gass, S. M. (2015). *Second Language Research: Methodology and Design*. London: Taylor & Francis Group.
[28] Marginson, S., & van Der Wende, M. (2007). Globalisation and Higher Education. *OECD Education Working Papers*, 1–83.
[29] Marsh, H. W. (1983). Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics. *Journal of Educational Psychology*, *75*(1), 150–166.
[30] Marsh, H. W., Hau, K.-T., Chung, C.-M., & Siu, T. L. P. (1997). Students' evaluations of university teaching: Chinese version of the Students' Evaluations of Educational Quality Instrument. *Journal of Educational Psychology*, *89*(3), 568–572.
[31] 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.
[32] Marsh, H. W., & Roche, L. (1993). The Use of Students' Evaluations and an Individually Structured Intervention to Enhance University Teaching Effectiveness. *American Educational Research Journal*, *30*(1), 217–251.
[33] Marsh, H. W., & Roche, L. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, *52*(11), 1187–1197.
[34] Mårtensson, K., Roxå, T., & Stensaker, B. (2014). From quality assurance to quality practices: An investigation of strong microcultures in teaching and learning. *Studies in Higher Education*, *39*(4), 534–545.
[35] Newton, J. (2000). Feeding the Beast or Improving Quality?: Academics' perceptions of quality assurance and quality monitoring. *Quality in Higher Education*, *6*(2), 153–163.
[36] Oman Academic Accreditation Authority. (2016). *Institutional Standards Assessment Manual Institutional Accreditation: Stage 2*. Muscat, Oman.
[37] Otani, K., Kim, B. J., & 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*, *18*(3), 531–544.
[38] Prøitz, T. (2010). Learning outcomes: What are they? Who defines them? When and where are they defined? *Educational Assessment, Evaluation and Accountability*, *22*(2), 119–137.
[39] Stensaker, B. (2008). Outcomes of Quality Assurance: A Discussion of Knowledge, Methodology and Validity. *Quality in Higher Education*, *14*(1), 3–13.
[40] Teddlie, C. (2009). *Foundations of mixed methods research integrating quantitative and qualitative approaches in the social and behavioral sciences*. London: SAGE.
[41] Walliman, N. (2006). *Social research methods*. London: SAGE.
[42] Wilson, C. (2013). *Credible Checklists and Quality Questionnaires: A User-Centered Design Method*. Newnes.
[43] Zerihun, Z., Beishuizen, J., & Van Os, W. (2011). Conceptions and practices in teaching and learning: Implications for the evaluation of teaching quality. *Quality in Higher Education*, *17*(2), 151–161.

Appendices

*Appendix I: The results of the Mann-Whitney U Test**

	T-SS 1**		T-SS 2		T-SS 3		T-SS 4		T-SS 5		T-SS 6		Overall	
	U-value***	(2-tailed)****	U-value	(2-tailed)	U-value	(2-tailed)	U-value	(2-tailed)	U-value	(2-tailed)	U-value	(2-tailed)	U-value	(2-tailed)
Good Teaching Scale														
1	16	0.312	25	0.54	11	0.229	23	0.694	17.5	0.317	12.5	0.159	1141	0.724
2	25	0.578	27	0.614	17.5	0.365	14	0.344	36.5	0.876	4.5	0.009	1114	0.625
3	8.5	0.081	18	0.344	8.5	0.183	27	0.903	35.5	0.838	4.5	0.009	1183	0.835
4	21.5	0.475	23	0.497	10.5	0.214	23	0.606	5	0.122	29.5	0.643	771.5	0.085
5	23	0.518	25	0.554	12.5	0.258	11	0.254	10.5	0.193	28.5	0.612	663	0.031
6	25.5	0.594	18	0.367	9.5	0.199	27.5	0.902	7.5	0.146	13.5	0.185	894	0.208
Clear Goals and Standards Scale														
7	29	0.829	10	0.203	8	0.176	8	0.191	10.5	0.189	30.5	0.771	597	0.022
8	13	0.21	19	0.381	11.5	0.233	15	0.349	29.5	0.628	31	0.69	1062	0.498
9	31.5	0.935	13	0.262	10	0.211	9	0.206	38	0.928	17.5	0.335	705	0.056
Appropriate Assessment Scale														
10	15.5	0.345	33	0.936	8	0.177	23	0.697	20.5	0.378	33.5	0.901	1117	0.666
11	22.5	0.574	16	0.332	27.5	0.737	25	0.785	38.5	0.946	27	0.657	1167	0.799
12	20	0.489	7	0.159	22.5	0.552	21.5	0.625	29	0.626	32.5	0.863	788.5	0.114
Appropriate Workload Scale														
13	16	0.367	15	0.311	1.5	0.093	15.5	0.404	21.5	0.412	4.5	0.124	1166	0.797
14	4	0.119	18	0.401	2	0.095	14.5	0.36	19.5	0.367	4	0.116	370.5	0.002
15	19.5	0.467	25	0.654	7.5	0.161	14.5	0.365	24	0.479	7.5	0.165	583	0.021
16	17	0.397	11	0.221	2.5	0.104	12.5	0.31	12.5	0.226	3	0.106	347	0.002
Generic Skills Scale														
17	19.5	0.473	11	0.226	8.5	0.188	17	0.453	14.5	0.263	26	0.626	746	0.086
18	25	0.665	27	0.711	5.5	0.137	3.5	0.116	21	0.4	30	0.767	645	0.036
19	29	0.831	28	0.736	3	0.108	14	0.345	15	0.267	23	0.522	797	0.121
20	26	0.698	25	0.603	13	0.269	6	0.129	38	0.926	24.5	0.552	1083	0.571
Overall Teachers Rating														
21	9	0.088	23	0.499	10.5	0.222	28	0.987	27	0.567	8.5	0.057	1208	0.909
Overall Course Rating														
22	16	0.368	27	0.736	17	0.379	18.5	0.509	22.5	0.442	11	0.223	1238	0.997

95% Confidence Interval, $p \leq .0.05$

- * The test results are of the six sub-groups (teachers and students) based on their evaluation of the adopted SCEQ
- ** T-SS: Teacher + his/her students (group wise)
- *** Mann-Whitney U Value
- **** Asymp. Sig. (2-tailed)