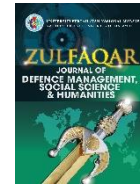




ZULFAQAR Journal of Defence Management, Social Science & Humanities

Journal homepage: <https://zulfaqarjdmssh.upnm.edu.my/index.php/zjdmssh>



EXAMINING STUDENTS' EVALUATION OF SIGNIFICANT TEACHING CRITERIA AT THE MALAYSIAN DEFENCE UNIVERSITY

Fatin Amirah Ahmad Shukri^{a*}, Mohd Syazwan Mohamad Anuar^b, Asniah Alias^c

^a fatin@upnm.edu.my

^b m.syazwan@upnm.edu.my

^c asniah@upnm.edu.my

*Corresponding author: fatin@upnm.edu.my

ARTICLE INFO

Article history:

Received

08-04-2022

Received in revised

17-04-2023

Accepted

19-04-2023

Available online

30-06-2023

Keywords:

exploratory factor analysis, knowledge evaluation, knowledge transfer, personal trait, student evaluations of teaching

e-ISSN: 2773-529X

Type: Article

ABSTRACT

This paper used an exploratory factor analysis (EFA) method to inquire about student evaluations of teaching (SET) among a batch of foundation students. A total of 504 students completed the lecturer evaluation questionnaire at the end of the semester. The current study was prompted by several factors, including teaching quality, lecturer characteristics, personal traits, knowledge evaluation and students' ratings which provided the impetus for the present study. Lecturers were evaluated using a Likert scale ranging from one to five, with the numbers labelled very poor, poor, medium, good, and excellent, respectively. Two phases of exploratory factor analysis conducted in the study to determine whether the importance ratings students regarded formed separate domains with the mean and the standard deviation of the questionnaire results. The first phase of exploratory factor analysis (Study 1) evaluated 23 criteria attributes listed by Vevere and Kozlinskis (2011). The attributes are seven personal trait items; 11 knowledge transfer items, followed by five knowledge evaluation items. Besides, participants were also asked to indicate any additional attributes they considered important to be possessed by their lecturers, the strand of the course they were enrolled in, and their gender. The purpose of the second study (Study 2) is to provide a contrast to the first study based on the data with an improved version of the SET instrument and with a different cohort of students. We discovered that all the proposed attributes of a lecturer were important to the students, and the categories proposed by Vevere and Kozlinskis (2011) for evaluating lecturers did not correspond to the importance of categories evaluated by the students.

© Fatin Amirah Ahmad Shukri 2023. All rights reserved.

DOI: <http://doi.org/10.58247/jdmssh-2023-0601-03>

Introduction

Student evaluation of teaching (SET) has become a common practice in universities. It is usually conducted by university administrations using generic instruments to provide a basis for performance reviews and decisions regarding staff development, employment, tenure, promotion, and awards (Boring, Ottoboni, & Stark, 2016; Spooren, Brockx, & Mortelmans, 2013); to provide teaching staff with feedback that may improve their performance and student outcomes; to provide students with information to assist them in course selection; and to provide metrics for comparisons between universities (Vidal & León, 2020; Alauddin & Kifle, 2014; Hazelkorn, 2009).

Several studies have looked into whether or not SET is valid and useful (see, e.g., Alauddin & Kifle, 2014; Benton & Kashin, 2012; Hornstein, 2017; Kornell & Hausman, 2016; Pounder, 2007; Marsh & Roche, 1997; Sánchez et al. 2020). Some aspects of SET have been heavily criticised. Mason et al. (1995), for example, considered the evaluation of students are not valid as they argued that they are not qualified to evaluate their teachers on certain items. Perhaps, the items on SET instruments were not perceived as relevant or important by students (Woodhouse, 2008). The results of SET are not the same as, or have little correlation with actual teaching quality of outcomes (Uttl, White, & Gonzalez, 2017; Burdsal & Harrison, 2008; Ginns & Barrie, 2004; Becker, 2000; Shevlin, Banyard, Davies, & Griffiths, 2000). Some teaching staff used inappropriate strategies to improve their SET scores (Jewell et al. 2013; Valsan & Sproule, 2008). There is biasedness against female academics, interactions between teachers and students, and an inverse relationship between SET results and actual teachers' effectiveness. More specifically, female academics are less likely to obtain more favourable scores compared to male teachers, especially when male teachers are evaluated by male students. Nonetheless, SET results are also positively correlated with grade expectations (Boring et al. 2016).

Analysis of the reliability of the items suggests that SET reveals more about student variability than about instructor characteristics in which students disagree about what they are asked to evaluate (Clayson, 2017). Teaching staff make little use of the results (Nasser & Fresko, 2002), which begs the question whether SET has any impact on the quality of teaching. SET results are frequently not presented in a way that they are easily interpretable or informative for university administrators, teachers, and others who receive SET results. Another issue is whether generic SET instruments are valid and useful when it is applied in a variety of context. The validity of generic SET instruments has been questioned (see Becker, 2000; Galbraith, Merrill, & Kline, 2012; Johnson, 2003; Spooen et al. 2013; Stark and Freishtat, 2014). For example, SET instruments offered to students within certain components of their overall courses may not be appropriate for other components; instruments offered early in students' overall courses may not be valid at other points in their courses; instruments appropriate for one overall course may not be appropriate for a different overall course; and instruments at one type of university may not be appropriate for another type of university. Despite these criticisms, arguments have been made that SET is not only useful (Feldman, 1997;1998), but the efforts to improve and explore it are also necessary (e.g., Marsh, 1987). For example, Burdsal and Harrison (2008) acclaimed that it is worth investigating whether SET results should be presented as a single figure or divided into different domains – multidimensional profile. Apart from that, Marsh and Roche (1997) suggested that SET should be considered as multidimensional. Even if SET does not indicate actual teaching outcomes in relation to students' performance (as indicated by Caddell & Newell 2019), it should be made as valid and useful as possible.

Several of abovementioned factors inspired the current study, which was conducted with first-year students at a defence university in Malaysia. In this university, most students are male and commence their studies between the ages of 18 and 20, after graduating from high school. They must pass physical fitness tests as well as meet the minimum entry requirements for Malaysian public universities. After completing an initial year of foundation academic study, the students will transfer to a 4-year degree program in defence engineering, defence management studies, defence science and technology, or medicine and defence health sciences. Throughout their studies, English, Bahasa Malaysia, or a combination of both languages is used as the medium of communication. SET is conducted at the end of every semester with an online evaluation form written in English. Students can easily understand what they are being asked to evaluate because of the language used. Teaching staff are subsequently given a set of feedback form, answered by the students, ranging from one to five, with the numbers labelled as very poor, poor, medium, good, and excellent, respectively.

Under certain circumstances, several features of SET might be regarded as assuming a particular salience. Lecturer attributes regarded as important or significant by the students at a defence university could be perceived insignificant by the students from other universities with different niche areas. Additionally, the SET instrument may not be presented in language that is easily understandable by students; and the results from SET might not be sufficiently helpful and meaningful, particularly for university administrators and the academicians themselves. Besides, the SET instrument are out of date and irrelevant to our current circumstance. Therefore, it is crucial to determine which unique criteria (if any) should be included in the evaluation of teaching. The research reported in this article began with an attempt to identify the qualities of lecturers that university students considered important. Veveve and Kozlinskis (2011) provided an appropriate foundation because they looked at a variety of instruments that had been developed for SET in teaching contexts including Australia, Russia, the United Kingdom, and the United States. Although these instruments tapped into noticeably discrepant aspects of teaching and

teacher attributes, Vevere and Kozlinskis (2011) proposed three teacher-related domains that they believed would have broad applicability and labelled these domains as personal traits, knowledge transfer, and knowledge evaluation. Within each of these domains, Vevere and Kozlinskis (2011) proposed specific attributes, which they referred to as criteria. We used these attributes as a basis for a questionnaire to assess how important students regarded each of them in their teachers. We then employed exploratory factor analysis to determine whether the important ratings formed a separate domain and, if so, whether those domains corresponded to the domains proposed by Vevere and Kozlinskis (2011).

This followed by the second study (study 2) to compare the first study (study 1) by using data based on an improved version of the SET instrument and with a different cohort of students. To minimise differences between the two cohorts, the questionnaire was given to foundation-year students from the year after those in Study 1 to minimise differences between the two cohorts. However, they were surveyed in the middle of the year, so they had less exposure to university teaching. Furthermore, data collection for this study occurred during the COVID-19 pandemic, whereby the new foundation students had been taught entirely online for the first five months of their course, which introduced an unanticipated variable into the comparison.

Methodology

i) Study 1

a) Participants

A total of 504 participants, representing 79% of the total cohort of 638 foundation-level students, provided data for this study. Questionnaires were administered in classes at the end of a teaching semester when many students had already returned home for semester break, but some students did not attend classes due to illness or other commitments.

b) Questionnaire

A set of questionnaire was administered based on the 23 attributes listed as criteria in Table 2 from Vevere and Kozlinskis (2011). The sequence of criteria in the table was retained, beginning with the seven personal trait items followed by the 11 knowledge transfer items and the five knowledge evaluation items. Subheadings of personal traits, knowledge transfer, and knowledge evaluation were placed above each group of items, highlighting the domains proposed by Vevere and Kozlinskis (2011).

Items were not reproduced accurately from the Vevere and Kozlinskis (2011) article, where major improvements in meaning or English expression were considered desirable. For all items, participants were asked to respond using five Likert-type response options of not important (1), slightly important (2), moderately important (3), important (4), and very important (5) to indicate how important each item was in terms of lecturer characteristics. Participants were also asked to indicate any additional attributes they considered important in their lecturers, the strand of the course they enrolled and their gender.

c) Procedure

Questionnaires were distributed to students in their regular mathematics classes at the end of their first year at the university. There were 26 of these classes, each comprising 15 to 30 students. These classes were conducted by 12 lecturers who spoke Bahasa Malaysia, English, or a combination of the two languages as their primary teaching languages.

The lecturers provided information about how they believed students should understand each item because the wording of some items, even after revision, was suspected of being either cryptic in content (e.g., Item 1 speech culture and Item 9 opposed different theories) or unusual (e.g., Item 8 provided with adequate workload and Item 17 offered different viewpoints of learning). Clarification was given if students needed more information.

d) Analyses

Data were analysed with SPSS Version 26® (IBM Corp., Armonk, NY, USA).

ii) **Study 2**

a) **Participants**

A total of 418 participants, representing 67% of the entire cohort of 626 foundation-level students, provided data for this study.

b) **Questionnaire**

An online English-language questionnaire was developed based on revisions to the questionnaire used in Study 1. The sequence of items was retained, but subheadings were not placed above each group of items proposed by Vevere and Kozlinskis (2011). Participants were offered the same five response options, presented as radio buttons. In addition to the 23 items on lecturer attributes, students were asked to indicate any additional attributes they considered important in their lecturers, the strand of the course they were enrolled in, and their gender.

c) **Procedure**

In regular online classes at the middle of the students' first year, students were told briefly about the research, invited to participate, and given a link to the questionnaire that was also made available to the students through their official WhatsApp and Teams groups. Students were asked to complete the questionnaire in their own time but within the next 3 days, and they were told that participation was not compulsory.

d) **Analyses**

Data were analysed with SPSS Version 26® (IBM Corp., Armonk, NY, USA).

Results

i) **Study 1**

a) **Participants**

Among the 418 participants, 56% of them enrolled in the engineering and technology foundation strand, 35% participants were in the management and strategic foundation strand, and remaining 9% participants in the medical foundation strand; 70% were male.

b) **Importance ratings**

The full range of response options was used on 16 of the 23 items. For the remaining seven items, the lowest option was not used. Most items had a small amount of missing data, with the most significant amount occurring on two items, each with nine omitted responses.

The mean and standard deviation (SD) of responses on each item are shown in Table 1. Twenty of the 23 items had mean responses that lay between the options of essential, and the three remaining items also had relatively high means, lying between the options of moderately essential. The bar charts (not shown), except for Item 8, indicate a strong negative skew in the students' responses. The proposed lecturer attributes were given a high level of importance for the most part, and there was a high degree of agreement. The 20 items with the highest importance ratings had the smallest SDs, while the three with the lowest had the largest SDs. Entries in Table 1 reveal that the five items with the highest mean (Items 4, 6, 11, 14, and 22) also had the narrowest SDs, whereas the item with the lowest mean (Item 8) had the broadest SD. The 20 items with the highest importance ratings had the narrowest SDs; conversely, the three items with the lowest importance ratings had the widest SDs.

Table 1: Means and Standard Deviations of Importance Items

Item	Variable	Mean	SD
Personal			
1	Speech culture	4.3	0.82
2	Respect for student	4.2	0.81
3	Appropriate appearance	4.2	0.86
4	Responsiveness	4.5	0.68
5	Punctuality	4.0	0.97
6	Good manners	4.5	0.67
7	Ability to control and discipline students	4.2	0.86
Knowledge transfer			
8	Supplied with adequate workload	3.5	1.01
9	Opposed different theories	3.7	0.98
10	Asked students about their goals	3.9	0.98
11	Introduced topics appropriately	4.4	0.72
12	Asked students about their learning interests	4.1	0.88
13	Encouraged students to focus on their interests and goals	4.3	0.79
14	Provided with appropriate practical examples	4.4	0.75
15	Explored learning issues fully	4.2	0.80
16	Ensured the required supply of literature and handout materials	4.0	0.85
17	Offered different viewpoints to the learning	4.1	0.82
18	Inspired students to further reading	4.1	0.87
Knowledge evaluation			
19	Offered students to evaluate themselves	4.1	0.83
20	Asked students how they intend to achieve the goals and tasks set	4.0	0.89
21	Offered students to share their ideas and knowledge	4.1	0.83
22	Explained to students why they are right or wrong	4.5	0.74
23	Opened new learning opportunities	4.4	0.76

^a Item wording in this table conforms with what appeared in the students' questionnaire. It does not always correspond with wording provided by Vevere and Kozlinskis (2011) because adjustments were made to improve comprehensibility.

In addition, the students also responded to the open-ended question about the qualities they believe lecturers should possess. Lecturers should be pleasant, friendly, caring, flexible, and approachable (88 mentions); have a good rapport with students and be supportive (47 mentions); they should be good communicators and teach in creative and exciting ways (63 mentions); be knowledgeable in their fields (24 mentions); and be reliable and fair (25 mentions). Three of these categories, with 160 mentions, referred to personality traits; two categories, with 87 mentions, referring to knowledge evaluation and knowledge transfer.

c) Exploratory factor analyses

None of the 253 inter-item correlations was <0.15 ; most were between 0.20 and 0.50; six were >0.50 , and none exceeded 0.62. The mean of these correlations was 0.31. These correlations, as well as results from Bartlett's test of sphericity ($p < 0.001$) and the Kaiser-Meyer-Olkin (KMO) index (0.92), proved that the data was suitable for Exploratory Factor Analysis (EFA) (0.92).

In the preliminary analyses, parallel analysis using principal components and 1,000 randomly generated matrices was indeterminate, implying the presence of two, three, or four factors. We used principal axis factoring as the extraction method, with Promax rotation to conduct EFAs constrained to each factor.

In the two-factor solution, two apparent factors emerged, one with 14 items based on how lecturers convey knowledge and the other with seven items based on the lecturers' characteristics, such as punctuality and good manners. Only 41.59% of the variance was accounted for in this solution, and although 21 of the 23 items were loaded at > 0.40 on their respective factors, the remaining two items were loaded at < 0.30.

In the three-factor solution, the first factor from the two-factor solution was divided into nine items related to delivering information and encouraging students' appreciation of complexity, and six items focused on students' own learning goals and interests. The third factor retained its seven items, focusing on lecturers' characteristics. However, five of the 23 items loaded < 0.40 (one < 0.30) and only 47.14% of the variance was accounted for.

The four-factor solution appeared to be superior to the other two solutions. In that solution, extraction communalities ranged from 0.29 to 0.55 (M = 0.43, SD = 0.07; only four > 0.50), and the factors accounted for were 34.24%, 7.34%, 5.56%, and 5.45% of the variance respectively, with a total of 52.59%. Results from this analysis are shown in Table 2. Within the first factor, seven items with loadings from 0.47 to 0.73 were focused on delivering information and encouraging an appreciation of theoretical complexity. Because of the general tenor of this factor, we labelled it Instructional Skill/Sophistication. Within the second factor, seven items, with loadings from 0.41 to 0.77, focused primarily on lecturers' characteristics. We labelled this factor Lecturer Personal Characteristics. Within the third factor were six items, with loadings from 0.45 to 0.72, focused primarily on students' learning goals and interests. We labelled this factor as Student Focus/ Orientation. The fourth factor had only two items, with loadings of 0.43 to 0.58, and was challenging to interpret. One of the 23 items loaded < 0.35 on all four factors, but the first factor had the highest loading (0.34).

Table 2: Factor Loadings from Exploratory Factor Analysis^a

Item	Item	Factors and item loadings ^b			
		1	2	3	4
17	Offer different viewpoints	.73			
15	Explore learning issues fully	.67			
14	Provide practical or real-life examples	.65			
18	Inspire students to do further reading	.58			
23	Open new learning opportunities for students	.50			
16	Ensure availability of literature and handout materials	.48			
22	Provide students with explanations about why they are right or wrong	.47			
11	Introduce topics effectively	.34			
5	Demonstrate punctuality		.77		
6	Be well mannered		.71		
3	Have appropriate self-presentation		.61		
4	Respond appropriately to students		.58		
2	Demonstrate respect for students		.49		
7	Control and discipline students appropriately		.44		
1	Use appropriate language and linguistic features when teaching		.41		
20	Ask students how they intend to achieve the set goals and tasks			.72	
10	Ask students about their goals			.72	
12	Ask students about their learning interests			.69	
13	Encourage students to focus on their interests and goals			.56	
21	Offer students opportunities for sharing their ideas and knowledge			.53	
19	Offer students opportunities for evaluating themselves			.45	
8	Assign a workload that promotes student learning				.58

^a Item wording in this table conforms with what appeared to be the intended meaning of the items as well as how the lecturers were most likely to provide explanations. This wording does not always correspond with wording in the students' questionnaire because lecturers were at liberty to try to clarify the meaning of items when administering the questionnaires.

^b In order to facilitate legibility, leading zeroes and loadings of < 0.30 are not shown.

Coefficient alpha values on the three main factors were 0.83 (7 items), 0.81 (7 items), and 0.82 (6 items), respectively. The inter-factor correlation, shown in Table 3, ranged from 0.31 to 0.69, with the lowest correlations associated with the fourth factor.

Table 3: *Inter-factor Correlation*

Factor	Lecturer Personal Characteristics	Student Focus/Orientation	Indeterminate
Instructional Skill/Sophistication	0.61	0.69	0.31
Lecturer Personal Characteristics		0.56	0.32
Student Focus/Orientation			0.32

ii) Study 2

a) Participants

Among the 418 participants, 56% of them enrolled in the engineering and technology foundation strand, 35% participants were in the management and strategic foundation strand, and remaining 9% participants were in the medical foundation strand; 70% were male.

b) Importance ratings.

The complete response options were used on 20 of the 23 critical items. On the remaining three items, the lowest option was not used. Most items had a small amount of missing data, with the most outstanding amount occurring on one item that had five omitted responses.

With the exceptions of Items 3 and 10, responses again exhibited negative skew. This was indicated by bar charts (not shown) and the item means lying above the option-range midpoint; almost all items have narrow SDs. Refer to Table 3. As evident in Study 1, a high level of importance was accorded to most items, and again there was a high degree of agreement. Entries in Table 3 reveal that the three highest means (on Items 1, 11, and 22) were associated with the narrowest SDs, and the two items with the lowest mean (Items 3 and 10) had the broadest SDs.

Students provided only a few responses to the open-ended question about additional attributes they believed were important in their lecturers. Predominant among these was that lecturers should be able to enliven the class atmosphere, even with jokes (7 mentions), check that students understand the material being taught (4 mentions) and have a good bond with students (3 mentions).

c) Exploratory factor analyses

Among the 253 interitem correlations, only one was < 0.15; most were between 0.20 and 0.50; 13 were > 0.50, and none exceeded 0.65. The mean of these correlations was 0.36. These results inclined us to regard the data as suitable for EFA, as did results on Bartlett's test of sphericity ($p < 0.001$) and the KMO index (0.94).

In preliminary analyses, parallel analysis using principal components and 1,000 randomly generated matrices indicated the presence of only two factors in the data. Therefore, we conducted an EFA constrained to two factors in which we used principal axis factoring as the extraction method, with Promax rotation.

In this solution, two apparent factors emerged. Extraction communalities ranged from 0.20 to 0.61 ($M = 0.41$, $SD = 0.10$; only four > 0.50), and the two factors accounted for 38.93% and 6.91% of the variance, respectively (total 45.84%). Results from this analysis are shown in Table 4. Twenty-one of the 23 items loaded at > 0.40, and the remaining two items at < 0.30.

Within the first factor, 13 items with loadings from 0.43 to 0.80 focused primarily on a combination of the Instructional Skill/Sophistication and Student Focus/Orientation items identified in Study 1, so we labelled it as Teaching. Within the second factor, eight items, with loadings from 0.42 to 0.89, focused primarily on lecturers' characteristics, including how lecturers interact with students and therefore corresponded to Lecturer Personal Characteristics in Study 1. However, the two items; one about providing students with explanations about why they are right or wrong and the other about communicating effectively when speaking; have little relationship with the other items, with loadings of < 0.35. These items were among the three with the highest importance rating, and their narrow SDs indicated a high degree of agreement among the students. Coefficient alpha values were 0.91 on Factor 1 (with 13 items) and 0.84 on Factor 2 (with eight items). The inter-factor correlation was 0.70.

Table 4: Means and Standard Deviations of Importance Items

Item	Variable	Mean	SD
1.	Communicate effectively when speaking	4.3	0.77
2	Demonstrate respect for students	4.0	0.82
3	Present themselves appropriately in terms of dress and grooming	3.2	1.01
4	Respond appropriately to students	4.1	0.79
5	Demonstrate punctuality	3.9	0.88
6	Be well mannered	4.0	0.82
7	Control and discipline students appropriately	3.9	0.89
8	Assign a workload that promotes learning	3.7	0.88
9	Present opposing theories or points of view effectively	4.0	0.77
10	Ask students about their goals	3.4	1.05
11	Introduce topics effectively	4.3	0.76
12	Ask students about their learning interests	3.8	0.94
13	Encourage students to focus on their interests and goals	4.1	0.81
14	Provide practical or real-life examples	4.1	0.80
15	Explore learning issues fully	4.0	0.81
16	Ensure availability of literature and handout materials	3.8	0.79
17	Offer different viewpoints	3.9	0.80
18	Inspire students to do further reading	3.9	0.89
19	Offer students opportunities for evaluating themselves	3.9	0.85
20	Ask students how they intend to achieve the set goals and tasks	3.7	0.95
21	Offer students opportunities for sharing their ideas and knowledge	3.8	0.82
22	Provide students with explanations about why they are right or wrong	4.3	0.73
23	Open new learning opportunities for students	3.9	0.88

^a Item wording in this table conforms with what appeared in the students' questionnaire.

Discussions

A major finding from this study is that the students regarded all the proposed lecturer attributes to be important. Furthermore, the pattern of means and SDs revealed a tendency for attributes considered most important to also attract the least divergence of opinion, and vice versa. An analysis of the items indicates that ambiguity might have contributed to greater diversity of opinions and lower ratings. Two items are noteworthy in this regard. Item 8, "Supplied with adequate workload", contains a term (supplied with) that is semantically inappropriate in relation to a workload, and the word adequate might be difficult to interpret in that context. In addition, Item 9, "Opposed different theories", is ambiguous because the word "opposed" could mean either contrasted or expressed a hostile attitude toward different theories. The ambiguity of these items might have prevented the lecturers who administered the questionnaires from providing clear or consistent explanations. This situation might not only have resulted in the more varied interpretations applied by the students, as evidenced in their wider SDs, but might also have contributed

to the lower levels of importance according to these items. This indicates two items (8 and 9) loaded together in the EFA, and on the fourth factor, it was worded differently, they might have loaded on one of the other three factors, probably Instructional Skill/Sophistication (Factor 1) because they inherent the thematic similarity with items on that factor.

The inter-factor correlations also raised some concern. The moderately high correlations between the first three factors might be expected, but the low correlations of those three factors with the two-item fourth factor (comprising Items 8 and 9) supports the possibility that rewording of the fourth factor's two items might move those items across to one of the other three factors. These results and conjectures indicate the importance of item wording when designing SET scales and, of more immediate relevance, that the items used in this study might benefit from rewording. The wording of items shown in Table 2 only partially moves toward that possibility. The EFAs revealed two main sets of insights about the foundations on which students evaluate their lecturers. First, some concern could be raised because the mean of the extraction communalities was only 0.43 in combination with only 52% of the variance in the data being accounted for by the four factors. If the wording of the items were improved, it would be possible to determine whether these characteristics were due to deficiencies in the wording, but some of the lack of consolidation could be due to students having had only 1 year of exposure to university-level studies. Additional lecturer attributes such as those suggested by the students, particularly attributes that did not appear on the questionnaire, might allow students to rate their importance in ways that are more meaningful to them and therefore increase the amount of shared variance in the data.

Secondly, the final EFA indicated that the categories proposed by Vevere and Kozlinskis (2011) for evaluating lecturers did not correspond to the categories of importance perceived by the students. Although all seven of the Vevere and Kozlinskis (2011) personal trait items loaded on Lecturer Personal Characteristics, only five of the 11 items that Vevere and Kozlinskis (2011) categorized as knowledge transfer grouped together, another three of those items loaded on Student Focus/Orientation, two loaded on the fourth factor that could not be characterized, and one of the 11 knowledge transfer items did not load adequately on any factor. Furthermore, three of the five Vevere and Kozlinskis (2011) knowledge evaluation items loaded on the five-item Student focus/Orientation factor but the other two loaded on Instructional Skill/Sophistication (items interspersed among each other). Given the prospect of improving outcomes by altering the wordings of some items, we decided to conduct a second study, while strictly adhering to the foundation laid by Vevere and Kozlinskis (2011) previously. Improvements to item wording could not only reduce students' uncertainty, but also remove the need for additional explanations concerning the meaning of the items, which could have added to the variety of meanings faced by, or generated by, respondents. There are notable similarities as well as notable differences between Study 1 and Study 2. The studies are reasonably similar in terms of the proportion of the cohorts that were represented among the respondents, with 79% and 67% of the students participating in the first and second studies, respectively. However, the reasons behind the participation similarity which was quite different in that physical attendance as classes were lower than usual in Study 1, largely attributable to students having gone home for vacation, whereas lower participation in Study 2 could be attributed to the questionnaire being delivered online.

The studies also gave the items a similarly high degree of importance, as shown by many items in Study 2 which have means for the important response option region and therefore have a negative data skew. There was also evidence that, as in Study 1, items with the highest importance ratings in Study 2 were associated with the greatest convergence of opinion. Attempts to improve wordings of some items appears to have had an effect. Particularly, Items 8 and 9, which had previously been characterized by ambiguity, were no longer among the items with the lowest ratings and widest SDs, implying that rewording of those items had sharpened the students' perceptions of them and, as a result, increased the importance given to those items. Inexplicably, both items were moved from the fourth, two-item, factor in Study 1 to the factor typified by lecturer personal attributes in this second study—rather than into a factor associated with teaching, as one might expect given their overt meaning. Four items received ratings that were noticeably lower in this second study. They were Item 3, *Present themselves appropriately in terms of dress and grooming*, $t(940) = 16.24, p < 0.001$; Item 6, *Be well mannered*, $t(940) = 10.19, p < 0.001$; Item 10, *Ask students about their goals*, $t(940) = 7.47, p < 0.001$; and Item 23, *Open new learning opportunities for students*, $t(940) = 9.26, p < .001$. Some of these differences might be attributed to the fact that teaching has moved from a face-to-face to an online environment, but others may be due to the difference in the length of time spent in university-level education (five versus ten months). In Study 2, Items 11 and 22 retained their high importance ratings and the narrow SDs, although they were joined by Item 1—perhaps because it had been reworded to produce greater clarity. Because these three items (1, 11, and 22) were among the

items with the lowest loadings among the factors in Study 1, they may belong to a separate domain of importance, and to create a domain with an adequate number of items, it would be necessary to create additional similarly themed items. The basis for those items might lie within the suggestions made by students in response to the open-ended question at the end of each study, particularly given the number and range of suggestions made by students in Study 1. If adding items raises concerns about the length of a SET scale, it may be countered by removing items that are overtly redundant. The high coefficient alpha of 0.91 on the first factor suggests that the items are redundant. This is possibility supported by Items 10 and 20 having similar overt content, as do Items 12 and 13. Furthermore, the inter-item correlations for these pairs of items were among the highest in both studies. At least one item of each pair could therefore be discarded.

Despite the expectation that rewording items would result in higher communalities and a greater percentage of variance accounted for in Study 2 EFA, the mean communality dropped from 0.43 in Study 1 to 0.41 in Study 2, and 52% of variance accounted for in Study 1 dropped to an unsatisfactory level of 45% in Study 2. The decrease in Study 2 suggests that unique features of the participants in that study, such as being earlier in their course and being taught online, resulted in a more diffused impression on what is important in lecturers. The substantial difference in factor structure between the two studies highlights the risk of assuming that the dimensions on which students evaluate lecturers are consistent from cohort to cohort and context to context.

Conclusion

This study might provide some foundation for increased effectiveness of SET in the university, as well as ideas for SET in other institutions. The students' perspective towards teaching staff may vary depending on the education institution, but this study may yield general attributes that may be added to the existing SET instrument in university. This can be seen in the second study, where some of the attributes are also accepted by the students as qualities, they want to see in their teaching staffs. As a conclusion, researchers should consider all the attributes in the Study 2, along with the attributes in Study 1, before omitting the attributes that is/are not selected by their respondents. Other studies will be useful if they can contribute to these general attributes, which may be unique to specific demographics, spatial and temporal studies.

ACKNOWLEDGEMENT

Authors gratefully thanked Robert Trevethan for helping with the data analysis and process of editing the paper. This research received no specific research grant from any funding agency in the public, commercial, or non-profit sectors.

REFERENCES

- Becker, W. E. (2000). Teaching economics in the 21st century. *J. Econ. Perspect.* 14 (1), 109–119.
- Burdsal, C. A., & Harrison, P. D. (2008). Further evidence supporting the validity of both a multidimensional profile and an overall evaluation of teaching effectiveness. *Assessment & Evaluation in Higher Education*, 33(6), 567–576.
- Ginns, P., & Barrie, S., (2004). Reliability of single-item ratings of quality in higher education: A replication. *Psychological. Reports.* 95, 1023–1030.
- Jewell, T. T., McPherson, M. A., & Tieslau, M. A. (2013). Whose fault is it? Assigning blame for grade inflation in higher education. *Applied Economics*, 45(9), 1185–1200.
- Johnson, V.E. (2003). *Grade inflation: A crisis in college education*. New York, NY: Springer.
- Mason, P. M., Steagall, J. W., & Fabritius, M. M. (1995). Student evaluation of faculty: A new procedure for using aggregate measures of performance *Economics of Education Review.* 14 (4), 403–416.

- Nasser, F., Fresko, B. (2002). Faculty views of student evaluation of college teaching. *Assessment & Evaluation in Higher Education*, 27(2), 187–198.
- Valsan, C., & Sproule, R. (2008). The invisible hands behind the student evaluation of teaching: the rise of the new managerial elite in the governance of higher education. *Journal of Economic Issues*, 42(4), 939–958.
- Woodhouse, H. (2008). Promoting quality literacy: Undoing the damage of rankings. *International Proceedings of the Australian Universities Quality Forum*, 16, 139–144.
- Alauddin, M., & Kifle, T. (2014). Does the student evaluation of teaching instrument really measure instructors' teaching effectiveness? An econometric analysis of students' perceptions in economics courses. *Economic Analysis and Policy*, 1–34.
- Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, 16(1), 78–117.
- Benton, S. L., & Cashin, W. E. (2012). Student ratings of teaching: A summary of research and literature. IDEA Paper 50. Manhattan, KS: The IDEA Center. https://www.ideaedu.org/Portals/0/Uploads/Documents/IDEA%20Papers/IDEA%20Papers/PaperIDEA_50.pdf
- Boring, A. (2015). Gender biases in student evaluations of teachers (working paper). Retrieved from Sciences Po publications. <https://www.ofce.fr/pdf/dtravail/WP2015-13.pdf>
- Boring, A., Ottoboni, K., & Stark, P. B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. Retrieved from Science Open Research. <https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.V1>
- Cadell, J. D., & Newell, D. (2019). Evaluating teacher impact on student performance: A case study at the United States Military Academy. *IEEE International Systems Conference*, pp.1-5. <https://doi:10.1109/SYSCON.2019.8836716>.
- Clayson, D. E. (2018). Student evaluation of teaching and matters of reliability. *Assessment & Evaluation in Higher Education*, 43(4), pp. 666–681. <https://doi.org/10.1080/02602938.2017.1393495>
- Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, 35(3), 124–129.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(1), 7.
- DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, 2(3), 292–307.
- Devlin, M. (2002). An improved questionnaire for gathering student perceptions of teaching and learning. *Higher Education Research & Development*, 21(3), 289–304.
- DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, 14(20), 1–11.
- Dunn, K. A., Hooks, K. L., & Kohlbeck, M. J. (2016). Preparing future accounting faculty members to teach. *Issues in Accounting Education*, 31, 155–170. <http://dx.doi.org/10.2308/iace-50989>
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage Publications: California.
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286–299.
- Galbraith, C. S., Merrill, G. B., & Kline D. M. (2012). Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business related classes? A neural network and Bayesian analyses. *Research in Higher Education*, 53(3), 353–374

- Gee, N. C. (2018). The impact of lecturers' competencies on students' satisfaction. *Journal of Arts and Social Sciences*, 1(2), 74–86.
- Gerritsen-van Leeuwenkamp, K. J., Joosten-ten Brinke, D., & Kester, L. (2017). Assessment quality in tertiary education: An integrative literature review. *Studies in Educational Evaluation*, 55, 94–116. <https://doi.org/10.1016/j.stueduc.2017.08.001>
- Hazelkorn, E. (2009). Rankings and the battle for world-class excellence: Institutional strategies and policy choices. *Higher Education Management and Policy*, 21(1), 1–22.
- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53–60.
- Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, 4(1), 1304016. <https://doi.org/10.1080/2331186X.2017.1304016>
- Ip, R. K. F., Iong, S. I. F., Wu, M. X. Y., & Wang, S. S. Y. (2017). A preliminary study on teaching quality assessment from the perspective of 'students as customers'. *Proceedings of 2017 IEEE International Conference on Teaching, Assessment and Learning for Engineering*, 444–448. <https://doi.org/10.1109/TALE.2017.8252376>
- Johnson, V.E., (2003). *Grade Inflation: A Crisis in College Education*. Springer: New York.
- Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press: New York.
- Kornell, N., & Hausman, H. (2016). Do the best teachers get the best ratings? *Frontiers in Psychology*, 7, 570. <https://doi.org/10.3389/fpsyg.2016.00570>
- Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research & Evaluation*, 12(2), 1–11.
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52(11), 1187–1197. <https://doi.org/10.1037/0003-066X.52.11.1187>
- Montanelli, R. G., & Humphreys, L. G. (1976). Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: A Monte Carlo study. *Psychometrika*, 41(3), 341–348.
- Newton, R. R., & Rudestam, K. E. (2012). *Your statistical consultant*. Sage Publications: CA.
- Osborne, J. W., & Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research & Evaluation*, 9(11). <https://doi.org/10.7275/ktzq-jq66>
- Pounder, J. S. (2007). Is student evaluation of teaching worthwhile? An analytical framework for answering the question. *Quality Assurance in Education*, 15(2), 178–191.
- Sánchez T, Gilar-Corbi R, Castejón J-L, Vidal J., & León J. (2020). Students' evaluation of teaching and their academic achievement in a higher education institution of Ecuador. *Frontiers in Psychology*, 11, 233. <https://doi.org/10.3389/fpsyg.2020.00233>
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323–338.
- Shevlin, M., Banyard, P., Davies, M.N., & Griffiths, M.D. (2000). The Validity of Student Evaluation of Teaching in Higher Education: Love me, love my lectures? *Assessment & Evaluation in Higher Education*, 25, 397 - 405. <https://doi.org/10.1080/713611436>

- Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the Validity of Student Evaluation of Teaching: The State of the Art. *Review of Educational Research*, 83(4), 598–642. <https://doi.org/10.3102/0034654313496870>
- Stark, P. B., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research*. 0(0), 1-7. <https://doi.org/10.14293/s2199-1006.1.sor-edu.aofrqa.v1>.
- Stewart, A. L., & Ware, J. E. (1992). *Measuring Functioning and Well Being: The Medical Outcomes Study Approach*. Durham and London: Duke University Press.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6th ed.). Boston, MA: Pearson.
- Thompson, P.A. (2016). *Perspectives of Teacher Quality: Perceptions from Secondary Educators in Private and Public Schools*. [Doctoral Dissertation, University of Portland]. University of Portland Pilot Scholars. <https://pilotscholars.up.edu/etd/27>
- Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22-42. <https://doi.org/10.1016/j.stueduc.2016.08.007>
- Vevere, N. & Kozlinskis, V. (2011). Students' evaluation of teaching quality. *US-China Education Review*, 5, 702–708.
- Westfall, P. H., & Henning, K. S. S. (2013). *Understanding advanced statistical methods*. Boca Raton: CRC Press.
- Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, 9(2), 79–94. <https://doi.org/10.20982/tqmp.09.2.p0>.