

QUESTIONNAIRE DESIGN AND DATA ANALYSIS: AN ALTERNATIVE APPROACH IN STUDENT EVALUATION OF TEACHING (SET)

LIM TICK MENG
CHIAM CHOOI CHEA
Open University Malaysia

PHANG SIEW NOOI
Sunway University College

ABSTRACT

Student evaluation of teaching (SET) has been widely used in institutions of learning to assess teaching effectiveness. However, most of the SET questionnaires are designed by “experts” without taking students’ view into consideration. Besides that, the traditional approach in SET analysis is to obtain a global rating as a measurement of teaching effectiveness based on the average or mean scores of the items of SET. This paper identifies the weaknesses of the traditional approach in SET questionnaire development and data analysis. It then describes an alternative approach to SET items development which involved students- collaboration. The paper also describes the establishment of various scoring matrices for the measurement of overall teaching effectiveness as well as teaching effectiveness from three different domains, namely: knowledge, pedagogical skills and attitude.

Keywords: Student evaluation of teaching (SET), knowledge, pedagogical skills, attitude, scoring matrices

INTRODUCTION

Students are the core stakeholders of any learning institution. They are directly involved in the teaching and learning process. As such, it appears reasonable to assert that they are one of the most important sources of information about the teaching ability of their lecturers (Aleamoni, 1981). Ironically, the use of student evaluation of teaching (SET) has always been controversial (Spooren & Mortelmans, 2006; Hess, et al., 2005; England, Hutchings & McKeachie, 1996). There were concerns regarding student evaluation forms which ask questions about instructors that students are not in a position to answer (Scriven, 1995). There are also issues related to objectivity and biases in students’ evaluation of their (Haladyna & Hess, 1994; Thorpe, 2002; Merritt, 2007). Despite these controversies, the use of SET as a tool to evaluate teaching in higher institutions remains popular. Besides being used for the purpose of improving teaching and learning, SET is also used for administrative decision-making purposes such as determining tenures or promotions.

Teaching is a multi-dimensional construct (Berk, 2005; Markley, 2004; Pagani & Seghieri, 2002; Marsh, 1992; Cashin, 1988). If SET were to be used to assess teaching performance, then the SET analysis carried out should provide adequate information regarding the teaching performance of the lecturers being evaluated. Subsequently, those lecturers evaluated by students need to be informed of their specific areas of weakness, if any. Thus the use of just an overall rating may not be sufficient. On the other hand, if SET

results were to be used for the purpose of appraisal, then it would be truly necessary to take appropriate steps to ensure that the evaluations by students are conducted fairly and that the evaluation results are reflective of their actual performances (Emery, Tracy and Robert, 2003; Scriven, 1995; VanLeeuwen, Dormody & Seevers, 1999).

With the above issues in mind, this research study focuses on two major tasks. While the first task involves lecturers and learners in determining the weighted key items as an effort towards designing a fairer instrument of SET, the second is to propose a method for SET score analysis which would depict as much information about the teaching ability of a lecturer as possible.

BACKGROUND OF THE PAPER

SET has been widely used in tertiary institutions to measure teaching performances. It is also a common practice for administrators in those institutions to use the same kind of instrument as one of the measures to evaluate their academic staff. Students' achievements to a large extent are governed by the quality of instruction provided. Thus, it appears reasonable for SET to be used as a means to appraise teaching performance as well as to identify the strengths and weaknesses in teaching, and subsequently to seek improvements.

For SET to be used for the purpose of improving teaching, data obtained and the analysis carried out should provide adequate information to the instructor regarding his teaching effectiveness. From the summative perspective, the instructor needs to know his overall performance as perceived by students. It has however been mentioned that teaching is a multi-dimensional activity and teaching performance is a product of various factors. As such, it may be inadequate to provide just a global rating of teaching performance. More importantly, the instructor needs to be informed of his specific areas of weakness so that remedial work can be carried out.

Looking from another angle, if SET analysis is to be used for administrative decision-making purposes, such as in determining the tenure and promotions of teaching staff, then the instrument and its data interpretations must fulfil the following requirements:

- The students who conduct the evaluation and the lecturer to be evaluated should agree upon the same set of criteria that constitutes teaching effectiveness.
- Students need to understand clearly the questions or items in the SET questionnaire.
- The data analysis should provide information regarding the staff members' current performance as compared to his past performances. It should also be able to fairly and objectively measure the performance of an instructor as compared to other instructors.

Tagomeri (1994) conducted an extensive content analysis of evaluation instruments used by schools of education accredited by the National Council for Accreditation of Teacher Education (NCATE). A total of 4028 evaluation questions from the 200 evaluation instruments were analyzed. The analysis reveals that 54.6% of the questions were ambiguous, unclear, or subjective in content. Another 24.5% of the questions did not correlate with classroom teaching performance. A total of 79.1% of the questions were either flawed or failed to identify the teaching behaviour. The study indicates that despite the emphasis on SET, very little effort has been made to ensure the quality of the

instruments used. Cashin (1988) holds the view that the quality of the items used to depict opinions determines the reliability of the instrument. As such, steps must be taken to ensure that the questionnaire used to measure teaching effectiveness should comprise parameters and criteria that truly measure teaching effectiveness.

Currently, students are rarely given the opportunity to decide on the items used in SET. Such practice may result in the following situation:

- Students do not fully understand the questions in the questionnaire, or they are asked questions that they are not in a position to answer.
- Students may be answering questions regarding factors that they themselves view as incidental to teaching effectiveness

Another possible issue is the over-concern about the multi-dimensionality of teaching effectiveness resulting in experts designing too many items for the SET instrument. For example, The SEEQ (Student Evaluation of Education Quality) which was developed by Marsh in 1987 consists of more than 40 Likert-scaled questions (Marsh, 1992). The CEQ (Course Experience Questionnaire) initially developed by Ramsden and revised in 2002 has 25 items (McInnis, Griffin, James, & Coates, 2001). The claim for developing a large number of items is that the analysis would then yield a more comprehensive measurement of teaching performance (Santhanam, Ballantyne, Mulligan, de la Harpe, & Ellis, 2000). Such argument fails to consider an important factor, that is, many students may be discouraged from answering honestly as they view the task to answer that many questions as an extra burden. It needs to be noted that in many institutions, the administering of SET is not made compulsory for students. As such, it is necessary to take measures that would encourage more student participation in the evaluation process. Furthermore, if ratings were to be conducted online on a voluntary basis without much monitoring, then the response rate would be expected to be low (Gamliel & Davidovitz, 2005; Anderson, Cain & Bird, 2005). This may in turn reduce the reliability of the study.

Eggen and Kauchak (2001) assert that teachers' subject knowledge, pedagogical skills, and positive attitudes towards teaching are core factors which contribute to teaching effectiveness. Knowledge in the subject area is obviously the fundamental requirement of an effective instructor. Without sound knowledge of the subject to be taught, it is unlikely that students will benefit much from the instructional process. Besides the subject knowledge, the ability to deliver effectively and to establish an environment conducive to learning is also essential. It is futile to have deep knowledge that cannot be transmitted effectively and efficiently. In addition, it is important for teachers to acquire the skills to motivate students as this is fundamental to effective learning. However, it is unlikely for teachers to be able to motivate students to learn unless they themselves are motivated to teach or to make learning happen. In short, effective teaching may not occur unless the teacher has positive attitudes towards teaching and related activities. Adediwura and Tayo's (2007) study investigate the effect of these factors on students' academic performance, revealed that there were significant positive correlations between students' perceptions of these factors in their teachers and their academic achievements.

If it is agreed that subject knowledge, pedagogical skills and positive attitudes towards teaching are the three core determinants to teaching effectiveness, then, it is reasonable to suggest the regrouping as well as analysis of items of SET instrument under these three domains. This would help to identify the strengths and weaknesses of the lecturers being

evaluated in one or more of these domains. However, it needs to be noted that items in different classifications may not necessarily be mutually exclusive.

It is a common practice for institutions of higher learning to find the mean score for all items of SET, as a way to determine teaching effectiveness (Liaw & Goh, 2003; Santhanam, et al., 2000; Vanleeuwen, et al., 1999). The product is a global rating indicating overall teaching effectiveness which by itself may not be useful for the purpose of improvement as the information is too general. Another significant aspect is that the calculation of mean score implies that all items in the questionnaire are regarded as of equal importance as factors determining teaching performance. This is in fact unrealistic. As an example, “knowledgeable in the subject area” is certainly a more important contributing factor to teaching effectiveness as compared to another factor such as “can interact well with students”. Based on this point, the use of mean score for the purpose of comparing teaching performance among lecturers may not be a fair measure. In establishing a fairer scoring system, it may be essential to take into consideration the degree of importance of the item as a contributing factor.

OBJECTIVE OF THE STUDY

Essentially, the research study has two main objectives, which are:

1. To develop a SET instrument with weighted items for the purpose of fairer assessment.
2. To formulate an objective and fair data analysis system which consists of quantitative scoring matrices that will allow for the measurement of teaching performance both from a global perspective as well from three different facets, namely, the subject knowledge, the pedagogical skills and the attitude towards teaching.

METHODOLOGY

In establishing the method of research, certain factors have been taken into consideration, and they are:

1. What are the aspects of teaching that students are capable of assessing?
2. What should be the respective weighting of items in the SET questionnaire to reflect the degree of importance of each of these items in the assessment of teaching effectiveness?
3. What are the items that need to be included in the student rating form so that the analysis provides sufficient information to help lecturers improve their instructional skills?

With these factors in mind, the research process involved qualitative data collection for the purpose of formulating and developing the items of SET instrument. For the first phase, the participants comprising 52 lecturers and 60 students from local universities and colleges were identified. The method of sampling adopted was judgement sampling, an extension of convenient sampling. Samples were chosen from universities and colleges, both public and private and the sample comprises the three main races of Malaysia, namely, Malay, Chinese

and Indian. These participants were requested to list down as many characteristics and attributes of lecturers (from the perspective of teaching and learning) that constitute teaching effectiveness. A total of 52 responses from the lecturers and 60 responses from the students were collected. The criteria were then analysed, re-classified and re-phrased using observable behavioural terms which are simple, direct, and familiar to students. The synthesis yields 20 items (Table 1).

Table 1. Items of SET

No.	Factors contributing to effective teaching
1	Clear and systematic presentation
2	Voice can be heard clearly
3	Provide adequate notes and study materials
4	Knowledgeable in the subject matter
5	Good time management in the delivery of content
6	Able to motivate students
7	Show enthusiasm in teaching
8	Always encourage students to ask questions
9	Always involve students in classroom activities
10	Good sense of humour
11	Grade students' assignments fairly
12	Punctual for class
13	Write legibly
14	Interact well with students
15	A good listener
16	Well prepared before teaching
17	Easily available for consultation
18	Provide clear feedback on assignments
19	Able to make the subject interesting
20	Always ask questions that require students' thinking

SAMPLING

The second phase of the study was carried out with the intention of determining the relative importance or weighting of each item. The study involved 99 lecturers and 155 students from two local universities, a foreign university and a local college. Convenient sampling was also adopted for the selection of the sample. The sampling may not be representative of any particular learning institution but it serves the purpose of prototyping a data analysis system and establishing a model of quantitative scoring matrices. The participants were requested to rank the items from 1 to 20 in order of importance, based on their own perceptions. The most important item (or factor contributing to teaching effectiveness) was

assigned a value of 1 and the least important factor was assigned a value of 20. Responses from a total of 99 lecturers and 155 students with complete entries were collected. Complete entries refer to responses that ranked all the twenty items, ranging from 1 to 20, without duplications or omissions.

THE ESTABLISHMENT OF ITEMS IN THE SET QUESTIONNAIRE

The mean of the rankings of each item was calculated for three categories of data, namely, the means of the rankings of items by students, the means of the rankings of items by lecturers, and the means of the rankings of items by the whole group. The result is shown in Table 2.

Table 2. Means of Rankings of Items

No	Important Characteristics of an Effective Lecturer	Means of Rankings by		
		Lecturers	Students	All
1	Clear and systematic presentation	5.10	6.09	5.70
2	Voice can be heard clearly	9.58	7.12	8.08
3	Provide adequate notes and study materials	9.74	6.91	8.01
4	Knowledgeable in the subject matter	3.52	6.48	5.33
5	Good time management in the delivery of content	11.30	10.97	11.10
6	Able to motivate students	8.36	8.90	8.69
7	Show enthusiasm in teaching	6.74	9.39	8.35
8	Always encourage students to ask questions	11.66	13.35	12.69
9	Always involve students in classroom activities	11.59	12.75	12.30
10	Good sense of humour	13.09	11.07	11.86
11	Grade students' assignments fairly	13.80	12.05	12.94
12	Punctual for class	13.44	12.48	12.86
13	Write legibly	17.05	14.40	15.43
14	Able to interact well with students	9.40	9.99	9.76
15	A good listener	13.56	12.89	13.15

16	Well prepared before teaching	5.42	8.12	7.07
17	Easily available for consultation	13.87	12.63	13.44
18	Provide clear feedback on assignments	13.53	12.35	12.81
19	Able to make the subject interesting	7.09	8.61	8.02
20	Always ask questions that require thinking	10.73	13.41	12.36

Figure 1 shows the respective line graphs for the two categories drawn on the same axis to compare the means of rankings of every item by lectures and by students. It is interesting to note that there are not many variations in terms of the relative weights of the items as perceived by lecturers and students.

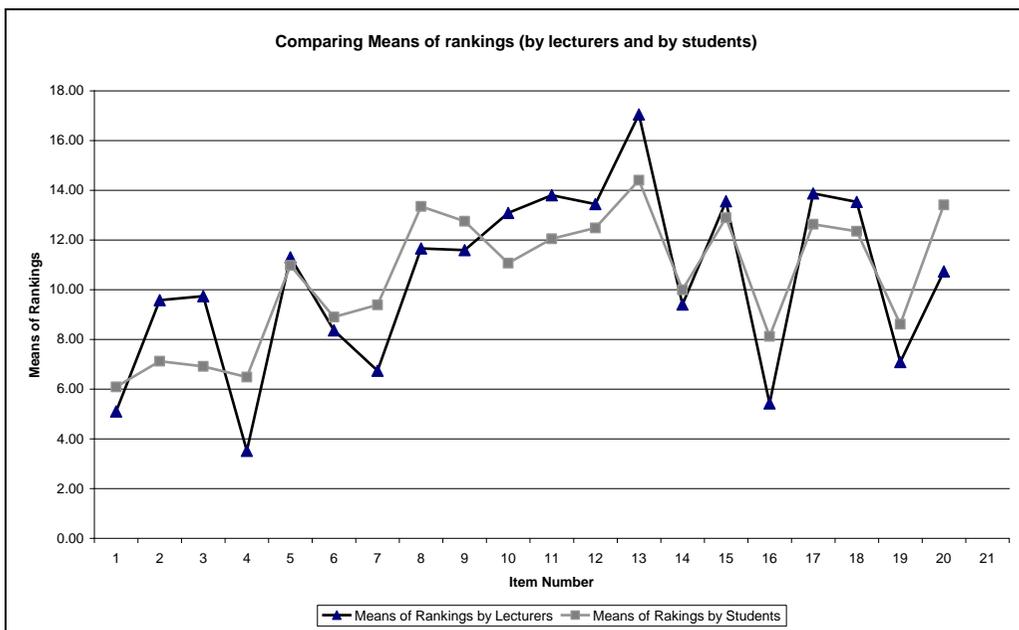


Figure 1. Comparison of the means of rankings (by lecturers and by students)

A paired t-test carried out for the means of rankings of each item by the two groups of participants (lecturers and students) returned a t value of -0.161 with a probability of 0.874. This indicates that there is no significant difference between the average rankings of items between lecturers and students.

Considering the concern that many students are unwilling to spend too much time in answering questions about lecturers, and too many questions may lead to biases, it was decided that there should not be more than ten items in the SET questionnaire. Thus only

nine items with means of rankings rated by the whole group below 10 were selected. The nine items were then classified into three domains, as follows:

1. Knowledge attributes
2. Pedagogical attributes
3. Attitudinal attributes

Some of these items belong to more than one domain (see Table 3)

Table 3. Revised Items for Student Evaluation on Teaching (SET) Questionnaire

No	Item	Domain			Overall Means of Ranking
		Knowledge	Pedagogy	Attitude	
1	Knowledgeable in the subject matter	√			5.33
2	Clear and systematic presentation	√	√		5.70
3	Well prepared before teaching	√		√	7.07
4	Provide adequate notes and study materials		√		8.01
5	Able to make the subject interesting	√	√		8.02
6	Voice can be heard clearly			√	8.08
7	Show enthusiasm in teaching			√	8.35
8	Able to motivate students		√	√	8.69
9	Able to interact well with students			√	9.76

To facilitate the formulation of scoring matrices, which will be discussed in the next section, the following Likert scale would be used for all the nine items:

- 1 = strongly disagree
- 2 = disagree
- 3 = neutral
- 4 = agree
- 5 = strongly agree

THE ESTABLISHMENT OF SCORING MATRICES

As mentioned earlier, teaching effectiveness is a multi-dimensional construct. Therefore, it is measured as a product of several factors. However, it should not be assumed that all factors contribute equally to teaching effectiveness. Some factors may have a greater impact than others. For example, the factor “knowledgeable in the subject area” is certainly more important than the factor “able to interact well with students”. As such, to establish a quantitative score as a measurement of teaching effectiveness based on the ratings given by students for all factors (items), the traditional way to calculate the mean score of all items may not be a fair measure of teaching effectiveness.

The establishment of scoring matrices for the SET instrument developed in this study was guided by the principle that the higher the perceived ranking of an item, the more weight should be assigned to that item in the calculation of a score for teaching effectiveness. The authors are aware that the adherence to such principle may be subjective and arguable, noting the fact that there might be disparity between the outcome of the rankings and the changing context of teaching today. Nevertheless, we are of the opinion that if SET is an assessment by the students on the teaching staff, then it is reasonable to adhere to the abovementioned principle. It needs however to be emphasized that SET is not the one and only mode of assessment of the lecturers. In this study, a smaller mean value of the rankings indicates a higher degree of ranking. Therefore, the weight assigned to each item in the determination of a quantitative teaching score for teaching effectiveness should be reciprocally proportional to the mean of the rankings of that item (by both and students). In short, the assigned weight of an item in the establishment of the final score is proportional to the reciprocal of the means of the ratings. Furthermore, it has been observed that both lecturers and students generally view ratings in the form of percentage as a more meaningful figure as compared to the 5-point rating. Therefore, in the scoring matrices established, the weighted Likert-scaled scores have been converted to their equivalence in the percentage form. Table 4(i), Table 4(ii), Table 4(iii) and Table 4(iv) show the scoring matrices for the measurement of teaching effectiveness from the knowledge domain, pedagogical domain, attitude domain and the overall teaching effectiveness respectively.

Table 4(i). Teaching Effectiveness Scoring Matrix (Knowledge Domain)

Item Description	Overall means of rankings	Reciprocal of means of rankings	Scoring Matrix Based on Scale of 1- 5				
			1	2	3	4	5
Knowledgeable in the subject matter	5.33	0.19	5.96	11.92	17.89	23.85	29.81
Clear and systematic presentation	5.70	0.18	5.58	11.15	16.73	22.30	27.88
Well prepared before teaching	7.07	0.14	4.49	8.99	13.48	17.98	22.47
Provide adequate notes and study materials	8.01	0.13	3.97	7.93	11.90	15.87	19.84
		Sum	20	40	60	80	100

Table 4(ii). Teaching Effectiveness Scoring Matrix (Pedagogical Domain)

Item Description	Overall means of rankings	Reciprocal of Means of Rankings	Scoring Matrix Based on Scale of 1- 5				
			1	2	3	4	5
Clear and systematic presentation	5.70	0.18	6.50	12.99	19.49	25.99	32.49
Provide adequate notes and study materials	8.01	0.13	4.62	9.25	13.87	18.49	23.12
Able to make the subject interesting	8.02	0.13	4.62	9.24	13.85	18.47	23.09
Able to motivate students	8.69	0.11	4.26	8.52	12.79	17.05	21.31
		Sum	20	40	60	80	100

Table 4(iii). Teaching Effectiveness Scoring Matrix (Attitude Domain)

Item Description	Overall Means of rankings	Reciprocal of Means of Rankings	Scoring Matrix Based on Scale of 1- 5				
			1	2	3	4	5
Well prepared before teaching	7.07	0.14	5.79	11.58	17.37	23.16	28.95
Provide adequate notes and study materials	8.01	0.13	5.11	10.22	15.33	20.45	25.56
Show enthusiasm in teaching	8.35	0.12	4.90	9.81	14.71	19.61	24.52
Interact well with students	9.76	0.10	4.19	8.39	12.58	16.78	20.97
		Sum	20	40	60	80	100

Table 4(iv). Overall Teaching Effectiveness Scoring Matrix

Item Description	Overall Means of rankings	Reciprocal of Means of Rankings	Scoring Matrix Based on Scale of 1- 5				
			1	2	3	4	5
Knowledgeable in the subject matter	5.33	0.19	3.09	6.18	9.26	12.35	15.44
Clear and systematic presentation	5.70	0.18	2.89	5.78	8.66	11.55	14.44
Well prepared before teaching	7.07	0.14	2.33	4.66	6.98	9.31	11.64
Provide adequate notes and study materials	8.01	0.13	2.05	4.11	6.16	8.22	10.27
Able to make the subject interesting	8.02	0.13	2.05	4.10	6.16	8.21	10.26
Voice can be heard clearly	8.08	0.12	2.04	4.07	6.11	8.15	10.19
Show enthusiasm in teaching	8.35	0.12	1.97	3.94	5.91	7.88	9.86
Able to motivate students	8.69	0.11	1.89	3.79	5.68	7.58	9.47
Interact well with students	9.76	0.10	1.69	3.37	5.06	6.75	8.43
		Sum	20	40	60	80	100

AN ILLUSTRATION

The SET questionnaire was administered in a postgraduate class on a trial basis, and one of the responses was used for the purpose of illustration. In the feedback form selected, five items were rated a score of 3 each, two items were rated a score of 4 each and the other two were rated a score of 5 each. The results generated using the proposed scoring matrices are shown in Table 5(i), Table 5(ii), Table 5(iii) and Table 5(iv) respectively.

Table 5(i). Overall Rating of Teaching

Item Description	Rating	Score
Knowledgeable in the subject matter	3	9.26
Clear and systematic presentation	3	8.66
Well prepared before teaching	3	6.98
Provide adequate notes and study materials	3	6.16

Able to make the subject interesting	3	6.16
Voice can be heard clearly	5	10.19
Show enthusiasm in teaching	4	7.88
Able to motivate students	4	7.58
Interact well with students	5	8.43
Total Score		71.31

Table 5(ii). Rating of Teaching (Knowledge Domain)

Item Description	Rating	Score
Knowledgeable in the subject matter	3	17.89
Clear and systematic presentation	3	16.73
Well prepared before teaching	3	13.48
Provide adequate notes and study materials	3	11.90
Total Score		60.00

Table 5(iii). Rating of Teaching (Pedagogical Domain)

Item Description	Rating	Score
Clear and systematic presentation	3	17.89
Provide adequate notes and study materials	3	16.73
Able to make the subject interesting	3	13.48
Able to motivate students	4	15.87
Total Score		63.97

Table 5(iv). Rating of Teaching (Attitude Domain)

Item Description	Rating	Score
Well prepared before teaching	3	17.89
Provide adequate notes and study materials	3	16.73
Show enthusiasm in teaching	4	17.98
Interact well with students	5	19.84
Total Score		72.43

It is interesting to note that although the particular lecturer obtained a global rating of 71.31 for teaching effectiveness, the scores he obtained which measure his teaching effectiveness from the knowledge domain (60.00) and pedagogical domain (63.97) indicate that he needs to place more effort towards the improvement of these particular domains. On

the other hand, a score of 72.43 for the attitude domain (Table 12) indicates that the lecturer was perceived by students to have a reasonably positive attitude towards teaching.

LIMITATIONS OF THE STUDY

In the above study, the overall average ranking for each item in the questionnaire was obtained by aggregating the rankings assigned by every subject and then calculates their means. The rankings obtained were then used in the calculation of the weighted coefficients, which in turn determine the weighted score of each item. Such a method raises the issue of external validity. Will the relative rankings and the ranking values be the same again? The answer is likely to be negative. There is a possibility of a slight variation in the relative rankings of the items if the research is to be conducted again by varying some variables such as the sample size, the targeted group. The ranking values will certainly vary. This in turn changes the relative weight of each factor. In other words, the scoring matrices established using this method is not meant to be an absolute index. It merely serves to establish a fairer system of evaluation.

It needs to be iterated that SET is not and should not be the only instrument for assessing teaching effectiveness of lecturers. Using SET to evaluate teaching by itself has limitations. But SET is still necessary as this is probably the easiest way to depict students' views about lecturers' teaching. Hence it may be worthwhile to improve the process of SET using the suggested means.

FUTURE RESEARCH AREA

The sample selected (52 lecturers and 60 students) to help determine the items of the SET questionnaire may not be a convincing sample. The categorisation of items into its respective categories (knowledge, pedagogical skills and attitude) is again arguable. It needs however to be noted that the focus of this paper is to propose an alternative process and scoring procedure for SET. The authors are fully aware that a more stringent validation process needs to be carried out to ensure the reliability and validity of the instrument. The authors are also aware that the categorisation of teaching into teaching effectiveness needs to be supported by sound theoretical framework and also tested through appropriate statistical means.

At this stage, the established items of the questionnaire which have been classified into three domains have not been subjected to the validation of their internal consistency within each classification. To ascertain that the items used in each domain truly measures that specific attribute (knowledge, pedagogical skills, or attitudes), factor analysis should be conducted. This is probably another area of research that needs to be looked into in the future.

The collaborative effort of and students in the establishment of questionnaire items is aimed at improving the validity of the items. In this respect, one basic assumption has been made, that is, and students know the factors that constitute effective teaching. However, it needs to be noted that the ultimate aim of teaching is to improve learning. As such, effective teaching should be portrayed in students' learning performance. Based on this argument, it

may be necessary to carry out another test of validity to look into the correlations between the evaluation outcomes of lecturers and the performances of the students.

CONCLUSION

An effective SET measures what it is supposed to measure. Effective SET begins with an appropriately designed questionnaire, with items relevant to the purpose and agreed upon by the student evaluators and the lecturers to be evaluated. An effective SET should also provide adequate information on different aspects of teaching effectiveness. This research study focused on these issues. Students were involved in the questionnaire design process. In ensuring a fairer system of evaluation, another survey was carried out to establish the rankings of every item in the questionnaire, according to the order of importance, as perceived by students and lecturers. Scoring matrices were established with these rankings being taken into consideration. The scoring matrices allow the measurement of overall teaching performance, as well as teaching effectiveness from the perspective of knowledge, pedagogical skills and knowledge. This research is an attempt to produce a more objective SET instrument for academic purpose as well as for administrative decision-making purposes. It is hoped that the idea would inspire other research to explore the possibility of producing a truly valid, reliable and fair system of evaluation for assessing teaching.

REFERENCES

- Adediwura, A. A., & Tayo, B. (2007). Perception of teachers' knowledge, attitude and teaching skills as a predictor of academic performance in Nigerian secondary schools. *Educational Research and Reviews*, 2(7), 165-171.
- Aleamoni, L. M. (1981). Student ratings of instruction. In J. Millman (Ed.), *Handbook of teacher evaluation* (pp.110-45). Beverly Hills, CA: SAGE Publication.
- Anderson, H. M., Cain, J., & Bird, E. (2005). Online student course evaluation: Review of literature and a pilot study. *American Journal of Pharmaceutical Education*, 69(1), 34-43.
- Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17(1), 48-62.
- Cashin, W. E. (1988). *Student ratings of teaching: A summary of the research*. IDEA Paper No. 20. Retrieved from http://www.theideacenter.org/sites/default/files/Idea_Paper_20.pdf
- Cashin, W. E. (1995). *Student ratings of teaching: The research revisited*. IDEA Paper No. 32. Retrieved from http://www.theideacenter.org/sites/default/files/Idea_Paper_32.pdf
- Eggen, P., & Kauchak, D. (2001). *Educational psychology: Windows on classrooms*. 5th ed. Upper Saddle, NJ: Prentice Hall.
- Emery, C. R., Kramer, T. R., & Tian, R. G. (2003). Return to academic standards: a critique of student evaluations of teaching effectiveness. *Quality Assurance in Education*, 11(1), 37-46.
- England, J., Hutchings, P., & McKeachie, W. J. (1996). *The professional evaluation of teaching*. American Council of Learned Societies Occasional Paper 33. Retrieved from http://archives.acls.org/op/33_Professional_Evaluation_of_Teaching.htm
- Haladyna, T., & Hess, R. K. (1994). The detection and correction of biases in student ratings of instruction. *Research in Higher Education*, 35(6), 669-687.
- Gamliel, E., & Davidovitz, L. (2005). Online versus traditional teaching evaluation: Mode can matter. *Assessment & Evaluation in Higher Education*, 30(6), 581-592.

- Hess, M., Barron, A. E., Carey, L., Hilbelink, A., Hogarty, K., Kromrey, J. D., et al. (2005, June). *From the learners' eyes: Student evaluation of online instruction*. Paper presented at the National Educational Computing Conference, Philadelphia, Pennsylvania, June 27-30, 2005. Retrieved from citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101
- Liaw, S. H., & Goh, K. L. (2003). Evidence and control of biases in student evaluations of teaching. *The International Journal of Educational Management*, 17(1), 37-43.
- Markley, T. (2004). *Defining the effective teacher: Current arguments in education*. Essays in Education Vol. 11. Retrieved from <http://www.usca.edu/essays/vol112004/markey.pdf>
- Marsh, H.W. (1992, April). *A longitudinal perspective of students' evaluations of university teaching: Ratings of the same teachers over a 13-year period*. Paper presented at the 1992 Annual Meeting of the American Educational Research Association, San Francisco, April 20-24.
- McInnis, C., Griffin, P., James, R., & Coates, H. (2001). *Development of the Course Experience Questionnaire (CEQ)*. Canberra: Australian Government Publishing
- Merritt, D. J. (2007). Biases, the brain, and student evaluations of teaching. *St. John's Law Review*, 82(1), 235-287. Retrieved from <http://scholarship.law.stjohns.edu/lawreview/vol82/iss1/6>
- Pagani, L., & Seghieri, C. (2002). A statistical analysis of teaching effectiveness from students' point of view. *Developments in Statistics (Metodoloski Zvezki)*, 17, 197-208.
- Ryan, S. M. (1998). Student evaluation of teachers. *The Language Teacher Online*, 22.(9). Retrieved from http://jalt-publications.org/old_tlt/files/98/sep/ryan.html
- Santhanam, E.; Ballantyne C., Mulligan, D., de la Harpe, B. & Ellis, R. (2000). Student questionnaires and teaching evaluation: Cutting the cloth to fit the purpose. In A. Herrmann and M.M. Kulski (Eds), *Flexible futures in tertiary teaching: proceedings of the 9th Annual Teaching Learning Forum, 2-4 February 2000*. Perth: Curtin University of Technology. Retrieved from <http://lsn.curtin.edu.au/tlf/tlf2000/santhanam2.html>
- Scriven, M. (1995). Student ratings offer useful input to teacher evaluations. *Practical Assessment, Research & Evaluation*, 4(7). Retrieved from <http://www.pareonline.net/getvn.asp?v=4&n=7>
- Spooren, P., & Mortelmans, D. (2006). Teacher professionalism and student evaluation of teaching: Will better teachers receive higher ratings and will better students give higher ratings? *Educational Studies*, 32(2), 201-214.
- Tagomori, H. T. (1994, April). *Content analysis of evaluation instruments used for student evaluation of classroom teaching performance in higher education*. Paper presented at the Annual Meeting of American Educational Research Association, New Orleans, April 4-8, 1994.
- Thorpe, S. W. (2002, June). *Online student evaluation of instruction: An investigation of non-response bias*. Paper presented at the 42th Annual Forum of the Association for Institutional Research.
- VanLeeuwen, D. M., Dormody, T. J., & Seevers, B.S. (1999). Assessing the reliability of student evaluations of teaching (SETS) with generalizability Theory. *Journal of Agricultural Education*, 40(4). 1-9.