

## A REVALIDATION OF STUDENTS' EVALUATION OF TEACHING EFFECTIVENESS RATING SCALE

**Bamidele Abiodun Faleye  
Olajide Aremu Awopeju  
Obafemi Awolowo University**

### **Abstract**

*This study conducted a confirmatory factor analysis of the Students' Evaluation of Teaching Effectiveness Rating Scale (SETERS). The study examined the factor structure of SETERS; it also investigated its convergent validity and determined the coefficients of its internal consistency and stability reliability. The study population comprised all undergraduate students of Obafemi Awolowo University, Ile-Ife, Nigeria. A sample of 1,000 undergraduate students was selected by systematic sampling. The selected students completed SETERS developed by Toland and Ayala (2005). The collected data were subjected to confirmatory factor analysis, correlation analysis and factor analysis (using unrotated principal component analysis. Results emanating from principal component analysis suggested five factors with eigenvalue greater than 1 that accounted for 59.967% of the total scale variance. Only the values of item loadings on factor one were significant. Scree plot also suggested only one factor, Instructor's Delivery of Course Information, and this was exemplified by the significant values of item loadings on the factor. Results also confirmed SETTERS as demonstrating strong convergent validity ( $r = 0.52$ ,  $p = 0.02$ ) against Student Evaluation of Education Quality scale developed by Marsh (1984). Finally, SETTERS was found to be reliable with a Cronbach Alpha coefficient of 0.91 and a split-half reliability coefficient of 0.98. It was concluded that SETTERS in its present form (i.e. in this study) is suitable for use in the evaluation of teaching effectiveness of lecturers in Nigerian universities.*

**Keywords:** SETTERS, Validation, Construct Validation, Scaling

### **Introduction**

Effectiveness generally refers to the extent to which somebody achieves his purpose. Relating this to the teacher therefore, it is about doing the right things in the teaching process so that at the end of teaching, the goals and objectives of the teachers can be achieved (Awotua-Efebo, 2004). Research findings have demonstrated that effective teaching leads to good academic performance in courses (Abdulkadir, 2006). Also, parents, practitioners and policy makers agree that the key to improving public education and teaching effectiveness is by placing highly skilled and effective teachers in all classrooms. A key problem is that current measures for assessing academics for promotion in most Nigerian Universities are not often linked to their capacity to teach effectively. Existing Federal University policies for measuring teacher effectiveness either rely almost exclusively on perception by heads of departments, or focus on teachers' course-taking records and on paper-and-pencil tests of basic academic skills and subject matter knowledge. Also, criteria for assessing academics for promotion in most Nigerian universities include qualifications, publications and community service at the local, national or international levels (Adomi and Mordi, 2003), all these are poor predictors of teaching effectiveness.

In recent years there has been growing interest in moving beyond these traditional measures of teachers' qualifications, such as completion of a preparation programme, number of degrees, years of experience, or the number of published paper, in order to evaluate teachers' actual performance and effectiveness in the classroom as the basis for making decisions for promotion and selection for leadership roles. Some private universities like Covenant University, Ota, Nigeria, have put some measures in place in assessing teachers' effectiveness. In a practical situation, however, such measures of teaching effectiveness lack both reliability and validity to a degree that probably makes them indefensible as indicators of teaching effectiveness. Research has shown that rating scale is a feasible approach to the evaluation of teachers' effectiveness in educational institutions in western counties (Marsh, 1987).

Scaling involves a high degree of operationalization and allows researchers to measure complex issues (Sarantakos, 2003). Furthermore, it enables researchers to summate values of several variables into one score and this with a relatively high degree of reliability. In general, it offers respondents a choice of picking their answers out of given sets of alternatives, which, as well shall see, are established in a very careful but also a cumbersome way.

University teaching involves diverse modes of instruction, including: lectures, seminars, laboratory and mentoring. Disciplines, courses, and instructors also vary widely in their emphasis on such different educational objectives such as learning new knowledge, stimulating student's interest, developing cognitive skills, and leading students to question established tenets. One important distinction in determining effective teaching may be the relative importance of cognitive and affective goals. That is, some qualities of teaching may be more strongly related to such cognitive objectives as knowledge and skills, whereas other qualities are more strongly related to affective objectives as fostering student's interest and curiosity.

For many years, educators and researchers have debated over which variables influence students' achievement. Schools can make a great difference in terms of student achievement, and a substantial portion of that difference is attributable to teachers, class size, class heterogeneity, intellectual capability and hard work. Specifically, differential teacher effectiveness is a strong determinant of differences in students' learning, far outweighing the effects of differences in class size and class heterogeneity (Darling-Hammond, 2000). Students who are assigned to ineffective teachers, one after another, have significantly lower achievement and learning (that is, gains in achievement) than those who are assigned to a sequence of several highly effective teachers (Sanders and Rivers, 1996). Thus, the impact of effective teaching (or ineffective teaching) seems to be additive and cumulative.

Students' evaluation of their instructors has been suggested to begin at the Universities of Medieval Europe (Centra, 1993). In the modern era, there has been a tremendous increase in interest regarding students' ratings of instruction and this topic has been the subject of a substantial body of research spanning approximately 70 years (Areola, 1995). Students' ratings were used in the North American Universities in the mid-1920s as basis for promoting their teachers (D'Apollonia & Abrami, 1997).

Students' Evaluations of Teaching Effectiveness Rating Scale (SETERS) has been useful in assessing teaching effectiveness in Western Countries. However, its suitability for use in Nigeria has not been established. Even though there is still low comparable, and validated instrument for measuring teaching effectiveness using Nigerian samples. Research

and theory have shown that teaching effectiveness as measured by students' rating of teaching is multidimensional in nature. Previous research by Toland and De Ayala (2005), conducted a multilevel factor analysis using samples from two mid-western universities in the United States of America. This analysis suggested that one or three factors between and within levels were a plausible representation of SETERS scores; that is, a one-factor model performed about as well as a three-factor model. These results contradict previous work that has found that SET measures are multidimensional. Hence, the conflicting factor structures of student evaluation of teaching also the need for this study.

The SETERS was developed in the United State of America, a culture and environment different from Nigeria's; there is need for cross-cultural validation of SETERS. As recommended by many authors that scales be cross-validated before use once such scales are to be used in different cultures and national boundaries (Tschannen-Moran, Woolfolk Hoy and Hoy, 1998 and Henson, 2001 and Brouwers, 2003). Moreover, previous researchers, Toland and De Ayala (2005) saw a need for additional empirical research on the SETERS before widespread use, hence this study.

Arising from the foregoing, one may wish to ask some pertinent questions such as:

- i. how reliable is SETERS?
- ii. are the items of SETERS significant in measuring teaching effectiveness?
- iii. how will sex of students and teachers affect students' ratings of teaching effectiveness?

The specific objectives of this study were to:

- i. examine the factor structure of SETERS
- ii. investigate the convergent validity of SETERS
- iii. determine the internal consistency and stability reliability coefficients of SETERS
- iv. find out if there is any significant relationship between male and female students' evaluation of teaching effectiveness of university teachers.

In order to be able to carry out this study effectively, the following research questions were raised:

- i. What is the factor structure of SETERS?
- ii. What is the convergent validity of SETERS?
- iii. What are the coefficients of SETERS' internal consistency and stability reliabilities?

## **Methodology**

The research design for this study is a survey. The population for this study consisted of undergraduate students of Obafemi Awolowo University, Ile-Ife. Five Faculties were selected by simple random sampling. They are Social Sciences, Education, Environmental Design and Management (EDM), Pharmacy and Science. The sample for this study comprised of 1,000 undergraduate students that were selected by systematic sampling. Courses were selected within each faculty through random sampling technique. The number of participants that were randomly selected from each course within the faculty was equal to the proportion in the population that was present for the class at the period of the administration of this instrument. That means that the samples were not equal for the courses.

'Students' Evaluation of Teaching Effectiveness Rating Scale' (SETERS) developed by Toland & DeAyala (2005) was adopted for the study. *SETERS questionnaire* is made up of two sections, namely Personal data and Inventory. The first section (i.e. Personal Data) consists of items on student's faculty, department, part, course title, course code, sex, age, level of interest in the subject matter prior to the course (1= very low to 5=very high) and lecturer's sex.

The second section consists of 34 items measuring three dimensions of teaching effectiveness, namely: Instructor's Delivery of Course Information (12 items), Teacher's Role in Facilitating Instructor/Student Interaction (10 items) and Instructor's Role in Regulating Student's Learning (12 items). The response is of the Likert-type format ranging from Strongly Disagree to Strongly Agree, on a rating scale of 1 to 5.

The personal data variables in the questionnaire include items 1-8, which were scored according to the number of the sub-group under each of the items. For example, in the case of 'sex', male is assigned one and female, two. In section B, there are five possible responses, 1-5, to each of the items. The response format for this scale was patterned after that of Likert (1932) in which the response options were five. They are strongly disagree (1), disagree (2), neither agree nor disagree (3) agree (4) and strongly agree (5). The highest obtainable point which a student's response could attract in particular course is 170 while the least is 34. However, an affirmative response to any of the statement indicates the level of agreement on that aspect of dimensions of teaching effectiveness. The scores of subject on each item were added together to obtain the course total score of each student on the instrument. The SETERS has internal consistency reliability with alphas ranging from 0.92 to 0.93 across the factors and total scores with a correlation between forms of 0.79.

In order to establish evidence of convergent validity for scores on the SETERS, data were collected from the sample using *Student Evaluation of Educational Quality* (SEEQ) questionnaire (Marsh, 1984). The SEEQ is the instrument which seems to have been most widely used. It has been adapted and utilized in Australia, New Zealand, Papua New Guinea and Spain (Marsh, 1984 and 1987; Marsh and Roche, 1997). A trial conducted by Curtin University of Technology Teaching Learning Group in 1997 confirmed that the SEEQ was far more acceptable than other in-house instruments. The SEEQ requires that students complete thirty four statements based on their perceptions of their teacher and the course materials and presentation. It is however perceived to have its limitations in that it presents student evaluations as a reflection of the teacher taking the unit of study and not on the actual unit itself. The results are therefore a function of the students' reactions and attitudes towards a particular teacher and not the materials.

The SEEQ is a 34-item (scored on a 5-point, Likert-type scale) questionnaire designed for use with undergraduate and graduate-level students. The nine SEEQ factors are Learning/Value (4 items), Instructor Enthusiasm (4 items), Organization/Clarity (4 items), Group Interaction (4 items), Individual Rapport (4 items), Breadth of Coverage (4 items), Examination/Grading (3 items), Assignments/Readings (2 items), Workload/Difficulty (2 items) and Overall (3 items). The consistency of student's ratings on the basis of previous research has shown subscale reliability estimates (coefficient  $\alpha$ ) ranging from 0.87 to 0.98 (Marsh, 1987) and subscale inter-rater reliability estimates for class average responses ranging from 0.90 to 0.95 (Marsh and Hocevar, 1994).

The questionnaires were administered by the researcher with the permission of the course coordinators (and the receipt of willingness to participate by those involved through informed consent) in the five faculties selected for this study. Data were collected two weeks before the rain semester examinations from students enrolled in classes. At the beginning of each administration, students were told of the purpose of the study and given the SETERS questionnaire. One week after, another similar instrument, SEEQ, was distributed to each student in the same courses as before. Course representatives assisted in distribution and collection of the questionnaires. The researchers distributed 1000 questionnaires by hand with the assistance of the class representatives, the administration and collection of the questionnaires took five days in all the faculties. Precisely, 735 questionnaires were retrieved which represented 73.5% return. A total of 725 questionnaires were found to be useful and this forms 72.5%. That is across the five faculties selected for sample, there were 735 undergraduate students, in 95 courses. Students ranged in age from 17 to 35 years. Out of 735 questionnaires that were retrieved, 10 were rejected as they were not properly completed.

The consistency of students' ratings was estimated for the total scale on the SETERS by computing Cronbach's alpha coefficient. Convergent evidence was established when positive correlation existed between scores on the SETERS and SEEQ. Confirmatory factor analysis was computed to determine the inter-correlation among the items on SETERS and to determine how much each item contributes to each factor and the total measurement instrument.

## Results

The data collected was subjected to Kaiser-Mayer-Olkin (KMO) measure of sampling adequacy test to determine the appropriateness of factor analysis for these data. For these data, the KMO value is 0.96, hence, factor analysis is appropriate for the collected data.

**Research Question 1:** What is the factor structure of SETERS?

The original SETERS administered contains 34 items with three factors (Toland and Ayala, 2005). The collected data were subjected to confirmatory factor analysis, intercorrelation of the items and factor analysis using principal component analysis (unrotated factor). Results emanating from principal component analysis (unrotated factor) suggested five factors with 'eigenvalue greater than one' that accounted for 59.967% of the total scale variance. This is presented in Table 1.

Table 1 listed the eigenvalues associated with each linear component (factor). It has identified 34 linear components within the data set. The eigenvalue associated with each factor represent the variance explained by that particular linear component and the eigenvalue was displayed in terms of the percentage of variance explained. Thus, factor1 explained 40.238% of the total variance, while subsequent factors explained only small amount of total variance. The analysis then extracted all factors with eigenvalues and left us five factors. These eigenvalues associated with the factors are again displayed (and the percentage of variance explained) in the columns labeled 'Extraction sums of squared loadings'. In determining the number of factors, only factors with eigenvalues greater than 1.1 were considered as significant (Rummel, 1970). Hence, five factors were suggested in Table 1. In order to confirm the result in Table 1, the scree plot in Figure 1 relates the eigenvalues with the component numbers, those values in the first two columns of Table1.

**Table 1: Initial Eigen Values of Items on SETTERS**

Comp onent	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	13.68	40.24	40.24	13.68	40.24	40.24
2	2.36	6.93	47.17	2.36	6.93	47.17
3	1.92	5.65	52.82	1.92	5.65	52.82
4	1.31	3.84	56.66	1.31	3.84	56.66
5	1.12	3.31	59.97	1.12	3.31	59.97
6	0.96	2.83	62.80			
7	0.85	2.50	65.30			
8	0.74	2.17	67.47			
9	0.69	2.02	69.49			
10	0.66	1.94	71.43			
11	0.64	1.89	73.32			
12	0.63	1.86	75.17			
13	0.58	1.69	76.86			
14	0.55	1.61	78.47			
15	0.54	1.58	80.05			
16	0.52	1.51	81.56			
17	0.49	1.43	83.00			
18	0.49	1.43	84.43			
19	0.47	1.39	85.82			
20	0.43	1.27	87.09			
21	0.41	1.22	88.31			
	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Comp onent	Total	% of Variance		Total	% of Variance	
22	0.41	1.20	89.50			
23	0.39	1.14	90.65			
24	0.380	1.12	91.76			
25	0.369	1.09	92.85			
26	0.348	1.02	93.87			
27	0.322	0.95	94.82			
28	0.314	0.92	95.74			
29	0.291	0.86	96.60			
30	0.263	0.78	97.38			
31	0.26	0.75	98.13			
32	0.24	0.69	98.82			
33	0.20	0.60	99.42			
34	0.20	0.58	100.0			



In figure 1, from the fourth component on, the line is almost flat, meaning that each successive component from the fourth is accounting for small and small amounts of the total variance. Hence, the scree plot suggested only one factor, and that is the first factor

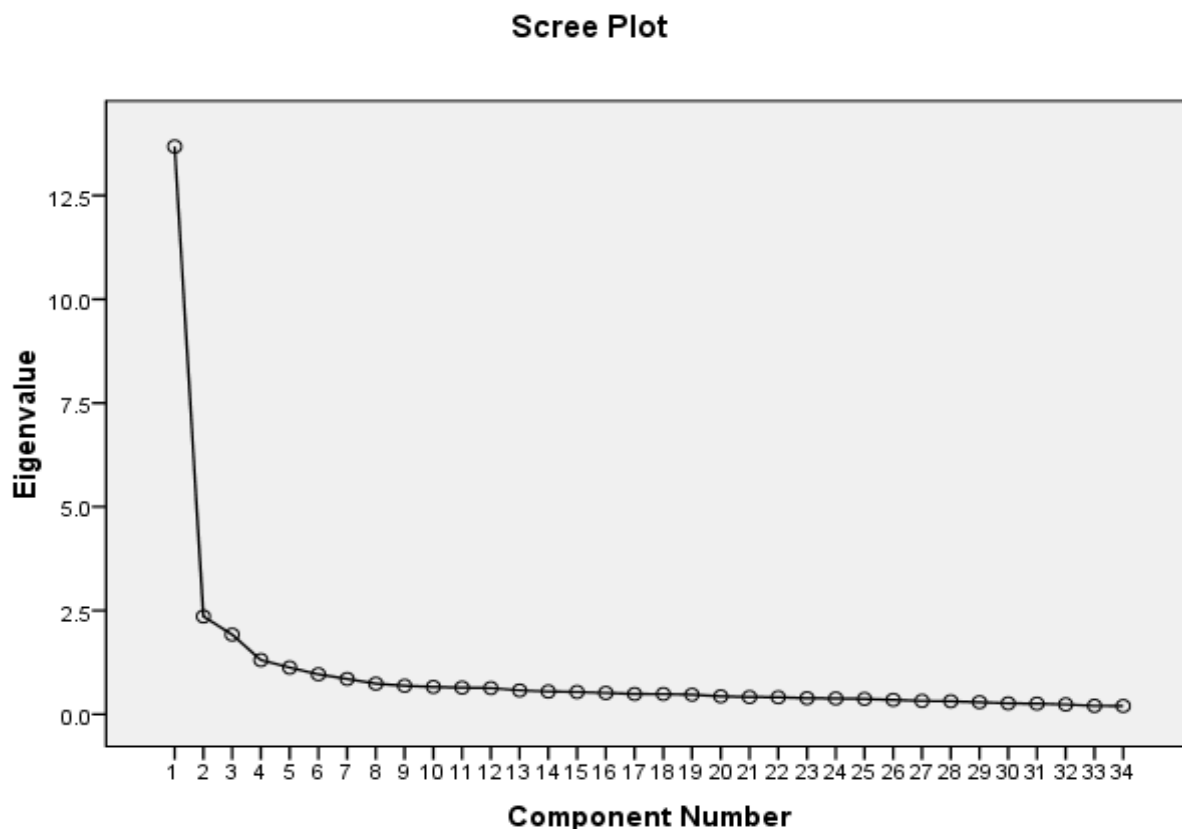


Figure 1: Scree Plot of the 34 item version of SETTERS

Looking at the results in Figure 1, the plot suggested that there was one factor on SETTERS. This is evident in the factor listed on the steep side of the plot, just before its elbow. Therefore, the numbers of factors on which the items on SETTERS loaded were again investigated and the results are presented in Table 2.

Table 2 showed that all the 34 items on SETTERS loaded on factor 1, Instructor’s Delivery of Course Information. A factor loading with absolute value greater than 0.50 was considered sufficiently high to assume a strong relationship between a variable and a factor. Factor loadings less than 0.50 in absolute value were regarded as insignificant. The item labeled “The lecturers promoted a comfortable learning atmosphere” has the highest coefficient loading. While the item labeled “the lecturer clearly outlined the direction of the course through a syllabus” has the lowest coefficient loading. Therefore, SETTERS contain only one factor, Instructor’s Delivery of Course Information.

Table 2: Item Loadings on each of the Suggested Five Factors

S/N	Items	Factors				
		1	2	3	4	5
1	The lecturer promoted a comfortable learning atmosphere	0.73	-0.08	0.01	0.16	-0.13
2	The lecturer was informative when responding to students' questions in class	0.70	-0.15	-0.08	0.20	-0.08
3	The lecturer encouraged me to share my knowledge about course subject matter	0.70	0.10	-0.42	-0.04	0.09
4	The lecturer encouraged me to express my opinions about course subject matter	0.69	0.06	-0.43	-0.05	0.09
5	The lecturer increased my interest in the course subject matter	0.68	-0.23	0.14	-0.20	0.07
6	The lecturer understood my pace for learning course content	0.67	0.28	0.04	-0.28	-0.37
7	The lecturer acknowledged my individual learning achievements/accomplishments	0.67	0.32	0.02	-0.31	-0.37
8	The lecturer helped me understand the course subject matter	0.67	-0.34	0.15	-0.17	0.11
9	The lecturer respected my opinions about course subject matter	0.66	0.10	-0.35	-0.05	0.01
10	The lecturer encouraged me to participate in class discussions	0.66	0.06	-0.47	0.10	0.12
11	The lecturer presented the course subject matter with enthusiasm	0.66	-0.24	0.22	-0.07	-0.10
12	The lecturer encouraged me to ask questions during classes	0.65	-0.01	-0.43	0.11	0.02
13	The lecturer helped me with my individual learning needs	0.65	0.37	0.03	-0.31	-0.28
14	I was motivated to learn the course subject matter	0.64	-0.18	0.09	-0.16	0.06
15	The lecturer made the class practical and participatory	0.64	0.00	-0.11	-0.10	-0.07
16	The lecturer's explanation of the course subject matter were clear	0.63	-0.26	0.09	0.02	-0.02
17	The lecturer related the course subject matter to my future needs	0.62	-0.25	0.16	-0.02	-0.18
18	The lecturer encouraged class discussions	0.62	0.07	-0.45	0.02	0.14
19	The lecturer clarified course material by reviewing concepts	0.62	-0.35	0.22	-0.05	-0.11
20	The lecturer appeared knowledgeable in the course content area	0.62	-0.32	0.04	0.32	-0.05
21	The lecturer related the course subject matter to my present needs	0.62	-0.14	0.08	-0.32	0.32
22	The lecturer used a variety of approaches/strategies when delivering lectures	0.62	-0.36	0.17	0.13	-0.09
23	The lecturer encouraged me to interact with other students in class	0.61	0.17	-0.30	0.05	-0.12
24	The lecturer's feedback on my learning progress was valuable	0.61	0.51	0.28	-0.04	0.10
25	The lecturer was an effective instructor compared with other College or university lecturers that taught me	0.61	-0.25	0.05	-0.23	0.27
26	The lecturer presented the course subject matter at an appropriately pace sequence	0.60	-0.16	0.04	-0.14	0.45
27	The lecturer provided feedback on my assigns/exams/tests that helped me learn from my mistakes	0.60	0.48	0.26	0.19	0.24
28	The lecturer explained the situations under which course subject matter could be applied	0.60	-0.29	0.15	-0.04	-0.10
29	The lecturer could be contacted outside the class	0.59	0.04	-0.14	0.30	-0.15
30	The lecturer's feedback on my assign/exams let me know what I knew and did not know	0.59	0.47	0.27	0.18	0.26
31	The lecturer provided me with feedback on my learning progress	0.59	0.50	0.32	-0.10	0.07
32	The lecturer graded my assigns/exams according to prescribed standard for grading	0.58	0.15	0.29	0.33	0.02
33	The assignment/examinations covered course content as emphasized by the Lecturer	0.57	0.12	0.22	0.36	0.08
34	The lecturer clearly outlined the direction of the course through a syllabus	0.54	-0.10	0.08	0.34	-0.20



**Research Questions 2:** What is the convergent validity of SETERS? To answer this question, the scores of SETERS and SEEQ were correlated to demonstrate convergent validity and the summary of the result is presented in Table 3.

**Table 3: Correlation between the scores on SETERS and the scores on SEEQ**

	$\bar{X}$	SD	$r$	P
SEEQ	1.19	10.63	0.52	0.02
SETERS	1.22	13.10		

*Correlation is significant at the 0.05 level*

Table 3 showed the result of Pearson correlation between the scores of SETERS and the scores of SEEQ where correlation coefficient was 0.52 and  $p = 0.02$ . This result suggested that there is significant correlation between the two instruments at the 0.05 level. Hence, there was a positive relationship between scores on the SETERS and SEEQ. Thus, convergent validity evidence for scores on the SETERS was established.

**Research Question 3:** What are the coefficients of SETERS' internal consistency and stability reliabilities?

To determine the internal consistency and stability reliability of SETERS, Cronbach's Alpha and split-Half analyses were performed on the 34 items of SETERS. The results showed that SETTERS has a Cronbach Alpha coefficient of 0.91 and a split-half reliability coefficient of 0.97. These coefficients showed that SETTERS is reliable.

SETERS comprising of 34 items was split into odd number items and even number items, the odd number items were correlated with even number items and the result showed that Guttman Split-Half coefficient (equal length) was 0.97. The split-half method treated the two halves of this measure as alternate forms. This "halves reliability" estimate is then stepped up to the full test length using Spearman-Brown Propesy formular. Thus,  $r_{xx'} = \frac{2r_{0e}}{1+r_{0e}}$  where  $r_{0e}$  = correlation between odd and even items, and  $r_{xx'}$  = estimated reliability for full test.

$$r_{xx'} = \frac{2 \times 0.97}{1 + 0.97}$$

$$r_{xx'} = 0.98$$

Therefore, the result, high  $r_{xx'}$  confirmed that SETERS is reliable and suitable for use. From the results, it was noticed that the data produced high coefficients of intercorrelation. This showed that the 34 items on SETTERS are probing (i.e. converging on) the same construct, which is teaching effectiveness.

## Discussion

The analysis of data collected on the Students' Evaluation of Teaching Effectiveness Rating Scale (SETERS) and Students' Evaluation of Educational Quality (SEEQ) among undergraduate students of Obafemi Awolowo University showed that correlation coefficient between the scores on SETERS and SEEQ was only fairly high but demonstrated that the scores on all three of the factors of the SETERS were positively related to the nine factors of

Marsh's (1983) SEEQ. Although it is expected that the scores on the SETERS would have strong positive relationship with the scores on the SEEQ because the items on them were suppose to be measuring teaching effectiveness. In other words, there are many reasons why this relationship was not positively strong; the scales were designed in developed country to measure teaching effectiveness in which the culture of teaching were different from Nigeria culture. The facilities that both the lecturers and students were exposed to and the conditions of teaching and learning environments are not comparable. Also, the students in the developed country are exposed to evaluating their teachers every semester or session unlike Nigerian students that do not have such background, although the undergraduate students of Obafemi Awolowo University were adequately informed of the importance and reason for the administration of these questionnaires.

Findings from this study also revealed that the high values of Cronbach's Alpha and Guttman Split-Half coefficient demonstrated that SETERS is highly reliable, stable and measuring teaching effectiveness. In other words, this scale is suitable for use in Obafemi Awolowo University. This is not different from the findings of Toland & De Ayala (2005) when they used samples from two mid-western universities in the United States of America to investigate the psychometric properties of SETTERS. Their findings showed that the items on SETERS could function on a one-factor model. These results contradict previous work that has found that SET measures are multidimensional.

Another major finding of this study is that, when SETERS was subjected to confirmatory factor analysis and its factors were conducted through the scree plot, it was revealed that all the 34 items of SETERS were considered to be significant to measure teaching effectiveness and one factor was more prominent than any other on the scale and this factor was Lecturer's Delivery of Course Information. All the items on the scale however were loaded on this only one factor. These results appeared to indicate that a one-factor of teaching effectiveness using the SETERS would be more suitable for this sample than a three-factor model as designed by Toland & Ayala (2005).

## Conclusion

SETTERS in its present form is suitable for use in Nigerian universities, given its high coefficients of reliability and the coefficients of its factor loadings.

## References

- Abdulkadir, H. S. (2006). Strategies for improving the efficiency of teachers in Jigawa State. Unpublished M.Ed. Thesis, University of Nigeria, Nsukka, Nigeria.
- Adomi, E. E. & Mordi, C. (2003). Publication in foreign journals and promotion of academics in Nigeria. *Learned Publishing*, 16(4), 259-63.
- Aleamoni, L. M. & Hexner, P. Z. (1980). A review of the research on student evaluation and a report on the effect of different sets of instructions on student course and instructor evaluation. *Instructional Science*, 9, 67-84.
- Areola, R. A. (1995). *Developing a comprehensive faculty evaluation system: A handbook for college faculty and administrators on designing and operating a comprehensive faculty evaluation system*. Bolton: Anker Publishing.
- Awotua-Efebo, E. B. (2004). *Effective teaching: Principle and practice*. Port-Harcourt, (Nigeria): Paragraphic Publishers.

- Brouwers, A. (2003). A test of the factorial validity of the teacher efficacy scale. *Research in Education*. Retrieved 12<sup>th</sup> February, 2007 from [http://www.findarticles.com/particles/mi+qa3765/is\\_200305/ai\\_n919453/print](http://www.findarticles.com/particles/mi+qa3765/is_200305/ai_n919453/print).
- Centra, J. A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. San Francisco, CA: Jossey-Bass.
- D'Apollonia, S. & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist*, 52(11), 1198-1208.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives*, 8(1).
- Doyle, K. O. & Whitely, S. E. (1974). Student ratings as criteria for effective teaching. *American Educational Research Journal*, 11, 259-274.
- Henson, R. K. (2001). Teacher efficacy scale: Substantive implications and measurement dilemmas. Keynote Address, Annual Meeting of the Educational Research Exchange, Texas A & M University, College Station, Texas. Retrieved 14<sup>th</sup> February, 2005 from <http://www.css.edu/users/dswenson/web/scaledevstat.html>
- Likert, R. (1932). *A technique for the measurement of attitudes*. New York, Columbia University Press.
- Marsh, H. W. & Roche, L. A. (1997). Making students' evaluation teaching effectiveness effective: The critical issues of validity, bias and utility. *American Psychologist*, 52, 1187-1197.
- Marsh, H. W. (1977). The validity of students' evaluations: Classroom evaluation of instructors independently nominated as best and worst teachers by graduating seniors. *American Educational Journal*, 14, 441-447.
- 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.
- Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76(5), 707-754.
- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11(2), 253-388.
- Marsh, H. W. & Hocevar, D. (1994). The factorial invariance of student evaluations of college teaching. *American Educational Research Journal*, 21, 341-366.
- Rumel, R. J. (1970). *Applied factor analysis*. Evanston: Northwestern University Press.
- Sanders, W. L. & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville: University of Tennessee Value-Added Research and Assessment Center.
- Sarantakos, S. (2003 ed.). *Social research*. South Yarra (Australia): Macmillan.
- Toland, M. D. & De Ayala, R. J. (2005). A multilevel factor analysis of students' evaluation of teaching. *Educational and Psychological Measurement*, 65(2), 272-296.
- Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, 68(2), 202-248.
- Young, S.; Rush, L. & Shaw, D. (2009). Evaluating gender bias in ratings of university instructors' teaching effectiveness. *International Journal for the Scholarship of Teaching and Learning*, 3(2), 631-684.

**Author Information**

Bamidele Abiodun FALEYE,  
Department of Educational  
Foundations and Counseling,  
Faculty of Education,  
Obafemi Awolowo University,  
Ile-Ife, Nigeria  
E-mail: [bamidelefaleye@yahoo.com](mailto:bamidelefaleye@yahoo.com)

and

Olajide Aremu AWOPEJU,  
Department of Educational  
Foundations and Counseling,  
Faculty of Education  
Obafemi Awolowo University,  
Ile-Ife, Nigeria