

Journal of Educational Research and Reviews Vol. 2(6), pp. 89-97, September 2014 ISSN: 2384-7301 Research Paper

Interrater reliability of University of Ibadan teaching practice scores

Fehintola, J. O.

Department of Guidance and Counselling, University of Ibadan, Ibadan, Nigeria.

E-mail: jof677@yahoo.com, joseph.fehintola@gmail.com. Tel: 08162023919.

Accepted 24th August, 2014

Abstract. This study investigated the relative effectiveness of interaction analysis on inter rater scores of the teaching practice of undergraduate students in University of Ibadan. The sample used for the study consisted of undergraduate students in the Faculty of Education, University of Ibadan which had undergone teaching practice exercise between 2011/2012 and 2012/2013 academic sessions for both regular and part-time students. One hundred and ninety-five students' scores for the teaching practice exercise for two raters were used for the study. To examine the extent to which the interrater scores influence the result of the exercise, Scott Reliability formular was used. Results showed that the reliability coefficient is r = 0.92. This means that the two raters' scores are relatively close. This study suggests that rating of teaching practice students by two or more raters (supervisors) enhances the ethics of the practice as the interaction agreement between raters score gives new dimension of obtaining true score of students and it prevents victimization of the student teachers in some quarters. It is recommended that stakeholders responsible for producing teachers in our schools make provision for enough supervisors that will supervise teaching practice students and this supervisor must be equipped and trained with the pedagogy to rate.

Keywords: Interrater scores, reliability, teaching practice, undergraduate students, faculty.

INTRODUCTION

Vision 20-20-20 of the Federal Government of Nigeria is a pragmatic step taken to guide the nation's course of scientific and technological development towards making Nigeria one of the best twenty world economies by the year 2020 (Chigbu and Idoko, 2013). Nigeria cannot achieve this without developing her human resources. This is because there must be relevant manpower to harness natural and human resources. Education is the best instrument for achieving the above mentioned goals. It is therefore very important that quality teachers should be prepared for this crucial role of human capital development (Chigbu and Idoko, 2013). No wonder, teaching as a profession is receiving recognition all over the world. The abilities of teachers are crucial determinants of the quality of education in any nation. Also, incompetent teachers may not help in training the youths to meet the challenges of modern life; hence, the technological growth of the nation suffers. The focus here is in the training of would be teachers; the issue now lies in effective supervising and rating of teaching practice students as many supervisors have failed in achieving this objective, thus discouraging practicing teachers. The National Policy of Education (Federal Republic of Nigeria, 2004) stipulated that, no educational system can rise above the quality of its teachers. Therefore, teachers ought to be sufficiently prepared to meet the modern demands of the teaching and learning processes. A look at inter rater score of teaching practice should give a reliable score but in many cases it is not so, making one to wonder the criteria these raters use in rating their students.

Teaching practice is a field experience course in Teacher Education Programme. Imogie (1990) defined teaching practice as a professional development exercise aimed at helping student teachers to bridge the gap between educational theories and practice. Maduabum and Abah (2004) defined teaching practice as a component of teacher education programme in which teacher trainees have the opportunity to put into practice all the teaching and classroom management theories they have acquired during the course of their studies with a view to attain professional competence.

De-ville (2010) stated that teaching practice is an internship experience, a part of formative training process for pre-service teachers in skills exploration and enhancement. In the context of this paper, teaching practice is a professional development course, a field experience for pre-service teachers' professional development. Teaching practice affords the student teachers opportunity to put into practice in actual classroom setting, theories and concepts learnt as curriculum contents in their academic programmes.

Philip et al. (2000) opined that supervision is a term that covers a wide range of teacher education processes and activities. It begins before a teaching practice, intensifies through the practice and continues after the practice has been completed. It includes aspects of advising, guiding, counselling, modelling, coaching, evaluating and assessing. It is a form of systematic, purposeful behaviour having clear aims and destructive content and activities. This study is hanged on the definition of Offorma (2005) which identifies technical competence, mental ability, human relation skills, high achievement and creativity as the attributes needed by supervisors. Supervisors of teaching practices programme are therefore expected to be equipped with adequate skills that will make them function maximally and prepare them to give qualitative supervision in any teaching practice programme in a bid to sustain high quality of education for the future development of Nigeria. Dandeago (2012) observes that supervision is a popular term in the public sector but poorly discharged. He remarks that what might be lacking is the expected 'close marking' to ensure that the supervisors do things according to plan/budget and the reporting of cases of laxity or ineffectiveness for appropriate remedial action to be taken. This idea was effectively utilized by the erstwhile.

Flanders (1967) commented that Interaction Analysis, originally used as a research tool, is employed by a trained observer to collect reliable data regarding classroom behavior as a part of a research project. The system is especially meaningful as an in service training device for teachers. It may be employed by a teacher either as he observes someone else teach or as he categorizes a tape recording of his own classroom behavior. In either case, the method is the same. In Interaction Analysis observation, all teacher statements are classified as either indirect or direct. This classification gives central attention to the amount of freedom granted the student by the teacher. Therefore, in a given situation, a teacher has a choice. He/she can be direct, minimizing the freedom of the student to respond, or he can be indirect, maximizing such freedom. His/her choice, conscious or unconscious, depends upon many factors, among which are his perceptions of the classroom

interaction and the goals of the particular learning situation. To make total classroom behavior or interaction meaningful, interaction analysis also provides for the categorizing of student talk. A third major section, that of silence or confusion, is included to account for the time spent in behavior other than that which can be classified as either teacher or student talk. Most of the contemporary writers agree that supervision is an organized, democratic process involving the supervisor and the persons or groups of persons supervised, for the achievement of desired objectives. Supervision of teaching practice should be based on the need that the student teacher does his/her teaching practice in schools and needs to be stimulated, motivated, supported, guided and inspired. This process could be called ideal supervision which could as well be referred to as mentoring. Supervisors of teaching practices programme are therefore expected to be equipped with adequate skills that will make them function maximally and prepare them to give qualitative supervision in any teaching practice programme in a bid to sustain high quality of education for the future development of Nigeria.

Aremu (2002) commented that supervisor needs to be aware that student teachers attitude in a teaching practice situation is guite different to that of normal classroom situation. There is what we popularly call 'teaching practice syndrome' which is noticeable among students of a school which is continuously used for demonstration purposes. Inter-rater agreement is the degree to which two or more evaluators give the same rating/close rating to an identical observable situation (e.g., a lesson, a video, or a set of documents) using the same rating scale. Inter-rater agreement is often confused with inter-rater reliability. The latter concerns the similarity in the ordering of ratings made by two or more evaluators. While both apply to any situation where multiple observers assess an action or object (such as a teacher's performance), it is important to recognize their differences. Agreement measures how frequently two or more evaluators assign the exact same rating: if both give a rating of "4," they are in agreement. Reliability measures the relative similarity between two or more sets of ratings, regardless of the absolute value of each evaluator's rating. Thus, even if two evaluators never assign the same numerical score, they could still have high inter-rater reliability if their ratings are in the same relative order (Tinsley and Weiss, 2000). Educator evaluation systems often designate certain rating levels as cutoffs for consequences or rewards, such as pay increases or retention decisions. In such situations, the educators being rated should probably be more concerned with inter-rater agreement than inter-rater reliability. Inter-rater reliability is still an important property, as a system with higher inter-rater reliability can more accurately determine a teacher's relative strengths and weaknesses. However, because reliability is concerned only with the relative order of rankings, it does not test for the possibility that some evaluators will be stricter than others.

Since the high-stakes decision is generally made based on the absolute value of the rating (e.g., a "2" versus a "3"), inter-rater agreement is likely to be of greater concern to parties that may reap rewards or consequences based on their rating.

It will suffice to note that inter-rater agreement and inter-rater reliability both measure aspects of how consistently raters apply knowledge of know-how, meaning awareness of the objectives of the construct to rate, each can be enhanced by reducing the variation between evaluators' ratings. Rater training is the most commonly identified method of decreasing the variability between raters, though studies have found that some variability can persist even after lengthy training (Hoyt and Kerns, 1999; Lumley and McNamara, 1995). Based on the finding that not all raters should reduce variability after training, some researchers have recommended recruiting more trainers than necessary and dismissing those who cannot pass agreement or reliability screenings (Henry et al., 2010; Lumley and McNamara, 1995). Research has identified a number of components of effective training. When designers fail to allot sufficient time for training, it is much less likely to improve evaluator performance (Barrett, 2001). Training typically involves viewing artifacts or videotaped lessons that designers have scored, group discussion of the standards and their meaning, instruction on how to interpret language, information about common sources of systematic rater error, and practice scoring sample videos (Bakker, 2008; Borko and Stecher, 2005; Clare, 2000).

Researchers often require that raters meet a predetermined minimum threshold of agreement or reliability in practice scoring sessions before they enter the field (Beesley, 2009; JAMB, 2005). Based on their qualitative analysis of evaluators, Henry et al. (2010) and Lumley and McNamara (1995) recommend that trainees discuss all of the relevant judgment processes, including personal biases and the mental processes needed to adjudicate between ratings or standards, so that raters will be more attuned to the ways in which their ratings might strav from objectivity and Rater selection is also an important component of the process. The American Educational Research Association's standards advise that raters understand the domain they are assessing as well as the subjects to be assessed (American Educational Research Association, 1999). This suggests that raters may need to be drawn from a pool of people with subject matter expertise and not just lecturers in a department. Finding such raters can be challenging when evaluating high and some middle school teachers of world languages or advanced science.

Rater attitudes and beliefs are also important. In their research on the Classroom Assessment Scoring System (CLASS), Henry at al. (2010) found that when raters' beliefs conflicted with the underlying theoretical foundation of the evaluation system, it was more difficult to bring their

ratings in line with those of other raters. As with other training issues, one possible solution is to recruit more evaluators than needed so that those who are difficult to train can be excused. Even if a good group of evaluators has been selected, maintaining inter-rater agreement and inter-rater reliability is not a one-shot deal, but an ongoing challenge. Rater agreement is not an inherent property of an evaluation system, but a facet that requires constant attention. Studies have shown that even reliable raters may change their rating behavior over time (Hoyt and Kerns, 1999; Lumley and McNamara, 1995). As a result of this finding, Lumley and McNamara (1995) recommend against the practice of certifying raters and then assuming that a single rater will produce consistent results — instead, they suggest periodic re-calibrations and the use of multiple raters.

The National Institute for Excellence in Teaching's evaluation system, The System for Teacher and Student Advancement, takes this approach (Dalev and Kim, 2001). Holding raters accountable for accurate rating is another potential way to improve agreement. Researchers have found that raters, even experienced ones, provide much lower quality data when they do not know that their performance is being monitored (Henry et al., 2010; Lumley and McNamara, 1995). This could be accomplished by reviewing some of the raters' scores on artifacts or classroom observations, possibly by randomly double-scoring videotaped observations or artifacts. Dymond (2008) found high levels of rater agreement between classroom observers and viewers of videotapes and contend that the use of videotaped observations could have the added benefit of reducing travel time and logistical issues, thereby making more observations possible.

As mentioned earlier, it is important to recognize that it is neither possible nor cost effective to achieve perfect agreement. Some degree of professional judgment is necessary if ratings are to represent different levels of complex behavior. However, evaluation system administrators can take many concrete steps to improve the consistency of evaluation results. This section discusses the major factors system designers should consider in order to maximize potential agreement.

Rater training is one of the most important tools system administrators have to improve agreement. Though some studies have found that some variability can persist even after lengthy training (Hoyt and Kerns, 1999; Lumley and McNamara, 1995), research has found that correctly designed training can improve agreement. Current thinking about rater training emphasizes developing a common understanding among evaluators so that they will apply the rating system as consistently as possible. This common understanding, is often called Frame of Reference (FOR) training, addresses the main sources of observer disagreement. Hoyt and Kerns (1999) identified: lack of overlap among what is observed, discrepant interpretations of descriptor meanings, and personal beliefs or biases. FOR training typically involves an explanation of the rating system, discussion of avoiding bias and common errors, advice on mental processes for observation and making judgments, and practice observations. Research shows that FOR training improves rating accuracy, reliability and validity (Henry et al., 2010; Lumley and McNamara, 1995).

While large-scale evaluation systems have the added challenge of ensuring that training sessions at different times or locations teach the same principles, administrators can take several steps to maximize uniformity. Dibu-Ojerinde and Jegede (1999) recommend monitoring the training sessions to ensure quality. Monitoring methods could include videotaping different training sessions, comparing how raters at different training sessions rate the same sample lessons, and conducting training for trainers using a FOR model that reinforces the common understandings to be developed among the raters. Although training many raters is more difficult, researchers have concluded that it is possible.

Henry et al. (2010) examined a sample of 2,093 Classroom Assessment Scoring System (CLASS) raters trained by the Office of Head Start and concluded that it was feasible to calibrate large pools of raters. Seventyone percent of those observers passed the initial screening on the first trial, which required 80% adjacent agreement (within one scale point) on a 7-point scale, as well as agreement requirements within each of the three dimensions CLASS assesses. The duration of training is also important. Research indicates that training needs to be more than an hour or two long to be effective. Researchers have found short training sessions to be ineffective at calibration (bringing evaluators in line with expert ratings or one another) and unlikely to produce consistent results (Barrett, 2001).

Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings 16 of those training sessions that lasted 5 h or more were significantly more effective than those lasting fewer than 5 h. They also suggested that for highly inferential measures (those that are not explicitly linked to a directly observable behavior), there was a large benefit for 25plus hours of training. This finding suggests that intermediate-length training sessions may be appropriate for more objective recording of behaviors, but that rating systems requiring more subjective judgments should use lengthier training. Many current systems require multi-day training, including CLASS, the Performance Assessment for California Teachers (PACT), Hillsborough County Schools' system in Florida, and the Early Childhood Environment Rating Scale (Henry et al., 2010).

Statement of the problem

The problem of teaching practice in terms of grading students' calls for concern as many raters (lecturers) often score some perceived brilliant students low and vice-versa, some students often complaint that the teaching practice score given to him/her is fictitious. At times, some lecturers will not go for supervision themselves they will send some postgraduate students to help them supervise. From aforementioned, the question that comes to mind is that could this be that teaching does not involve brilliancy or raters insensitivity to scoring or because of the teaching practice supervision by proxy? In some cases, a rater might not visit a practicing student nor send anybody to supervise the student-teacher and such supervisor will just decide to give any mark, indicating that score giving to individual under this condition is not a true representation of the performance of such student-teacher.

Purpose, significance and research question of the study

The study examines the implication of inconsistency in rating teaching practice students by their supervisors. It is obvious that rating students by a single rater does not provide a reliable score for students. This now calls for insisting that three or more supervisors should observe student-teachers and return appropriate scores to prevent the award of spurious, unwarranted or biased scores to students, resulting in undeserved grades. This study provides an avenue for teaching practice organizers on the need to have two or more raters for the teachers in training. This is expected to give insight to students to know their true scores on teaching practice when rated by two or more supervisors. This will also prevent the situation whereby the students will be feeling that he/she is being victimized by one supervisor or the other. The supervisor too will be free of harassment from the students in the sense that it is not a supervisor that determine his/her performance in teaching practice. The education stake holders will also know that the individual student-teacher is well supervised and that any of them that is able to pass the training are actual merit it. The only research question raised for the study is how reliable are the scores of the raters on the teaching practice exercise?

METHODOLOGY

The research design for this study is a descriptive research design of correlational type because it is quantitative in nature due to the fact that the performance of the student-teacher would be rated by the supervisor in line with observer rating scale. The population for this study includes students and lecturers in Faculty of Education, University of Ibadan, Ibadan, Oyo state, Nigeria in the 2010/2011 to 2012/2013 sessions for both regular and part-time students. Simple random sampling of one hundred and ninety-five (143) students out of which 75 of them are males and the rest 68 were females

which participated in the teaching practice exercise and were supervised by researcher in the 2010/2011 of 2012/2013 sessions were used for the study. The researcher and some of his colleagues generated the data used for this study. Each time the researcher went to the field to supervise he will record the proceedings before grading the student teacher and when the researcher gets back to school, he played it back to his colleagues for assessment and the scores are recorded. One hundred and forty-three students were used and two raters were used to rate the teachers in trainee. The data collected by the researcher from the two raters on the teaching practice in the Faculty of Education were subjected to Scott's inter-rater method. The study is conducted in the University of Ibadan, making use of lecturers as respondents from the Faculty of Education. The lecturers that are involved are lecturer II and above with education background and with at least five years working experience.

RESULTS

The results in Table 1 are as follows: Column 1 simply stands for the number of teaching practice students used in the study, that is, serial number 1 to 143 while Columns 2 and 3 stands for the number of raters used in rating the performance of the teaching practice students and the scores given to each student. Columns 4 and 5 stand for the proportion of the score given to each student over the sum total of the scores in each column multiply by 100. Column 6 is the difference between the columns 4 and 5 and column 7 is the squared of column 6.

Scott formular is given as:

$$\Pi = \frac{P_0 - P_e}{100 - P_e}$$

$$P_0 = (100 - 7.860) = 92.14$$

$$P_e = 1.045$$

$$\Pi = (92.14 - 1.045)/100 - 1.045$$

$$= 91.095/98.955$$

= 0.92

Table 1 presents a summary of the scores awarded by the supervisors. Values in the second and third columns represent information on the scores awarded by the first and second supervisors respectively. Thus, the mean score of the first supervisor was 61.9 for all the 143 student-teachers, and the mean score of the second supervisor was 62.4 for the 143 student-teachers who were observed by two different supervisors, and hence had two teaching practice scores. The inter-rater reliability of the two raters using Scotts method showed that r = 0.92. This shows that the scores of the raters were much closed and in agreement.

DISCUSSION

The result of the research question in this study showed that the two scores given to the student-teachers independently by the two raters correlated significantly with each other. Some past studies equally established the fact that there was significant correlation between two observers that observed classroom activities during teaching-learning process of university lecturer (Tinsley and Weiss, 2000). Also, Flanders (1967) in his study discovered that when two or more trained observers were used to observe classroom behaviour as part of a research project they collected data that result to high interrater reliability. Furthermore, the result of this study corroborate the study of Aremu (2002), who discovered after using three trained raters in micro teaching exercise obtained high and significant reliability coefficient. The result of this study is contrary to the findings of Gorman and Rentsch (2009) that confirmed that there is no significant relationship in the grade awarded by both WAEC and NECO in their public examination if the two examining bodies were seen as two observers or raters. Therefore, an interrater reliability study requires a carefully planned design that ensures the data gathered is relevant and valid. A poorly designed case study or question may be misunderstood by the raters and result in poor outcomes. It requires an experienced reviewer to interpret and analyze results compared with the plan's standard criteria.

RECOMMENDATIONS

For educational practice, this study suggests that rating of teaching practice students by two or more raters (supervisors) will enhance the ethics of the practice as the interaction agreement between raters score gives new dimension of obtaining true score of students and it will put an end to the complaints of the students teachers that the supervisor deliberately under rated him/her. It is therefore recommended that stakeholders responsible for producing teachers in our schools make provision for enough supervisors that will supervise teaching practice student and these supervisors must be equipped and trained with the pedagogy to rate students.

CONCLUSION

The results of this study showed a considerable concordance between the scores awarded by different supervisors to student-teachers in the teaching practice exercise conducted in University of Ibadan, Ibadan,

Category		Observer B	9/ A	% B	% Difference	(Λ)
		CDServer D	0.40	0.55	0.12	
1	51	64 60	0.42	0.55	0.13	0.004
2	60	60	0.55	0.55	0.00	0.006
3	63	62	0.53	0.52	0.01	0.006
4	64 52	60	0.54	0,55	0.01	0.005
5	52	00	0.44	0.57	0.13	0.005
6 7	62	00	0.52	0.55	0.03	0.005
/	60 CC	70	0.50	0.59	0.09	0.006
8	60	63	0.55	0.53	0.02	0.006
9	63	62	0.53	0.52	0.01	0.005
10	64	72	0.54	0.60	0.06	0.007
11	53	54	0.44	0.45	0.01	0.004
12	63	52	0.53	0.44	0.09	0.005
13	61	50	0.51	0.42	0.09	0.004
14	64	60	0.54	0.50	0.04	0.005
15	60	60	0.50	0.50	0.00	0.005
16	63	61	0.53	0.51	0.02	0.005
17	64	58	0.54	0.49	0.05	0.005
18	62	56	0.52	0.47	0.05	0.005
19	63	64	0.53	0.54	0.01	0.005
20	65	63	0.55	0.53	0.01	0.006
21	66	50	0.55	0.42	0.13	0.004
22	68	61	0.57	0.51	0.06	0.006
23	67	65	0.56	0.55	0.01	0.006
24	60	60	0.50	0.50	0.00	0.005
25	65	53	0.55	0.44	0.11	0.005
26	61	51	0.51	0.43	0.08	0.004
27	63	68	0.53	0.57	0.04	0.006
28	61	62	0.51	0.52	0.01	0.005
29	65	63	0.55	0.53	0.02	0.005
30	57	64	0.48	0.54	0.06	0.005
31	62	65	0.52	0.55	0.03	0.006
32	70	66	0.59	0.55	0.04	0.006
33	63	64	0.53	0.54	0.01	0.005
34	62	65	0.52	0.55	0.03	0.006
35	63	65	0.53	0.55	0.02	0.006
36	57	60	0.48	0.50	0.02	0.005
37	63	60	0.53	0.50	0.03	0.005
38	50	62	0.42	0.52	0.10	0.004
39	58	60	0.49	0.50	0.01	0.005
40	65	60	0.55	0.50	0.05	0.005
41	62	50	0.52	0.42	0.10	0.004
42	45	62	0.38	0.52	0.14	0.004
43	57	61	0.48	0.51	0.03	0.005
44	61	62	0.51	0.52	0.01	0.005
45	55	61	0.46	0.51	0.05	0.005
46	65	60	0.55	0.50	0.05	0.006
47	60	68	0.50	0.57	0.07	0.006
48	63	60	0.53	0.50	0.03	0.005
49	66	60	0.55	0.50	0.05	0.006
50	66	58	0.55	0.49	0.06	0.005
51	65	50	0.55	0.42	0.13	0.004

Table 1. Data presentation and analysis.

Table 1. Contd.

52 65 62 0.55 0.52 0.03 0.006 53 68 56 0.57 0.47 0.10 0.005 54 67 50 0.56 0.57 0.13 0.009 55 50 61 0.42 0.51 0.09 0.004 56 50 66 0.42 0.55 0.13 0.005 57 52 60 0.44 0.50 0.06 0.004 58 60 58 0.50 0.49 0.02 0.005 60 58 62 0.49 0.52 0.03 0.006 61 60 68 0.52 0.07 0.006 63 62 58 0.49 0.49 0.00 0.005 64 64 65 0.54 0.50 0.04 0.005 67 63	
53 68 56 0.57 0.47 0.10 0.005 54 67 50 0.56 0.57 0.13 0.005 55 50 61 0.42 0.51 0.09 0.004 56 50 66 0.42 0.55 0.13 0.005 57 52 60 0.44 0.50 0.06 0.004 58 60 58 0.50 0.49 0.01 0.005 59 61 58 0.51 0.49 0.02 0.005 60 58 62 0.49 0.52 0.03 0.005 61 60 68 0.50 0.57 0.07 0.006 62 62 63 0.52 0.53 0.01 0.006 63 62 58 0.49 0.49 0.00 0.005 64 64 65 0.54 0.54 0.00 0.006 65 66 65 0.55 0.54 0.01 0.006 66 65 0.52 0.50 0.03 0.005 68 47 65 0.39 0.54 0.15 0.04 70 60 60 0.50 0.50 0.00 0.005 71 57 57 0.48 0.48 0.00 0.005 74 45 56 0.38 0.47 0.09 0.004 75 55 51 0.46 0.43 0.03 <	
54 67 50 0.56 0.57 0.13 0.005 55 50 61 0.42 0.51 0.09 0.04 56 50 66 0.42 0.55 0.13 0.005 57 52 60 0.44 0.50 0.06 0.004 58 60 58 0.50 0.49 0.01 0.005 59 61 58 0.51 0.49 0.02 0.005 60 58 62 0.49 0.52 0.03 0.005 61 60 68 0.50 0.57 0.07 0.006 62 62 63 0.52 0.53 0.01 0.006 63 62 58 0.49 0.49 0.00 0.005 64 64 65 0.54 0.54 0.00 0.006 65 66 65 0.54 0.50 0.04 0.005 67 63 60 0.54 0.50 0.04 0.005 68 47 65 0.39 0.54 0.15 0.004 70 60 62 0.50 0.50 0.00 0.005 71 57 57 0.56 0.48 0.02 0.005 74 45 56 0.38 0.47 0.99 0.004 75 55 51 0.46 0.48 0.00 0.005 74 45 56 0.38 0.49 <td< td=""><td></td></td<>	
55 50 61 0.42 0.51 0.09 0.004 56 50 66 0.42 0.55 0.13 0.005 57 52 60 0.44 0.50 0.06 0.004 58 60 58 0.50 0.49 0.01 0.005 59 61 58 0.51 0.49 0.02 0.005 60 58 62 0.49 0.52 0.03 0.005 61 60 68 0.50 0.57 0.07 0.006 62 62 63 0.52 0.53 0.01 0.006 63 62 58 0.49 0.49 0.00 0.005 64 64 65 0.54 0.54 0.01 0.006 65 66 655 0.54 0.01 0.006 66 65 60 0.54 0.50 0.04 0.005 67 63 60 0.53 0.50 0.03 0.004 70 60 60 0.50 0.50 0.00 0.005 71 57 55 51 0.48 0.02 0.005 74 45 56 0.38 0.47 0.09 0.004 75 55 51 0.46 0.43 0.03 0.004 75 55 51 0.46 0.43 0.03 0.004 76 64 57 0.54 0.56 0.02 <td< td=""><td></td></td<>	
56 50 66 0.42 0.55 0.13 0.005 57 52 60 0.44 0.50 0.06 0.004 58 60 58 0.50 0.49 0.01 0.005 59 61 58 0.51 0.49 0.02 0.005 60 58 62 0.49 0.52 0.03 0.005 61 60 68 0.50 0.57 0.07 0.006 62 62 63 0.52 0.53 0.01 0.006 63 62 58 0.49 0.49 0.00 0.005 64 64 65 0.54 0.54 0.00 0.006 65 66 65 0.55 0.54 0.01 0.006 66 65 60 0.54 0.50 0.04 0.005 67 63 60 0.53 0.50 0.00 0.005 68 47 65 0.39 0.54 0.15 0.004 69 50 60 0.42 0.50 0.08 0.004 70 60 60 0.50 0.50 0.00 0.005 71 57 57 0.48 0.48 0.00 0.004 72 60 57 0.50 0.48 0.00 0.005 74 45 56 0.38 0.47 0.09 0.004 75 55 51 0.46 0.48 <t< td=""><td></td></t<>	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	
58 60 58 0.50 0.49 0.01 0.005 59 61 58 0.51 0.49 0.52 0.03 0.005 60 58 62 0.49 0.52 0.03 0.005 61 60 68 0.50 0.57 0.07 0.006 62 62 63 0.52 0.53 0.01 0.006 63 62 58 0.49 0.49 0.00 0.005 64 64 65 0.54 0.54 0.00 0.006 65 66 65 0.55 0.54 0.01 0.006 66 65 60 0.54 0.50 0.04 0.005 67 63 60 0.53 0.50 0.03 0.004 69 50 60 0.42 0.50 0.08 0.004 70 60 60 0.50 0.50 0.00 0.005 71 57 57 0.48 0.48 0.02 0.005 73 60 62 0.50 0.52 0.02 0.005 74 45 56 0.38 0.47 0.99 0.004 75 55 51 0.46 0.43 0.03 0.004 76 64 57 0.54 0.48 0.06 0.005 77 58 58 0.49 0.46 0.03 0.004 79 64 61 0.54 <t< td=""><td></td></t<>	
59 61 58 0.51 0.49 0.02 0.005 60 58 62 0.49 0.52 0.03 0.005 61 60 68 0.50 0.57 0.07 0.006 62 62 63 0.52 0.53 0.01 0.006 63 62 58 0.49 0.49 0.00 0.005 64 64 65 0.54 0.54 0.00 0.006 65 66 65 0.55 0.54 0.01 0.006 66 65 60 0.54 0.50 0.04 0.005 67 63 60 0.53 0.50 0.03 0.005 68 47 65 0.39 0.54 0.15 0.004 69 50 60 0.42 0.50 0.004 0.005 71 57 57 0.48 0.48 0.00 0.004 72 60 57 0.50 0.52 0.02 0.005 73 60 62 0.50 0.52 0.02 0.005 74 45 56 0.38 0.47 0.09 0.004 76 64 57 0.54 0.48 0.06 0.005 77 58 58 0.49 0.46 0.03 0.004 79 64 61 0.54 0.51 0.03 0.006 81 63 65 0.53 0.54 <	
60 58 62 0.49 0.52 0.03 0.005 61 60 68 0.50 0.57 0.07 0.006 62 62 63 0.52 0.53 0.01 0.006 63 62 58 0.49 0.49 0.00 0.005 64 64 65 0.54 0.54 0.00 0.006 65 66 65 0.55 0.54 0.01 0.006 66 65 60 0.54 0.50 0.04 0.005 67 63 60 0.53 0.50 0.03 0.005 68 47 65 0.39 0.54 0.15 0.004 69 50 60 0.42 0.50 0.08 0.004 70 60 60 0.50 0.50 0.00 0.005 71 57 57 0.48 0.48 0.00 0.004 72 60 57 0.50 0.48 0.02 0.005 74 45 56 0.38 0.47 0.09 0.004 75 55 51 0.46 0.43 0.03 0.004 76 64 57 0.54 0.48 0.06 0.005 77 58 58 0.49 0.46 0.03 0.004 79 64 61 0.55 0.51 0.04 0.006 81 63 65 0.53 0.54 <t< td=""><td></td></t<>	
61 60 68 0.50 0.57 0.07 0.006 62 62 63 0.52 0.53 0.01 0.006 63 62 58 0.49 0.49 0.00 0.005 64 64 65 0.54 0.54 0.00 0.006 65 66 65 0.55 0.54 0.01 0.006 66 65 60 0.54 0.50 0.04 0.005 67 63 60 0.53 0.50 0.03 0.005 68 47 65 0.39 0.54 0.15 0.004 69 50 60 0.42 0.50 0.08 0.004 70 60 60 0.50 0.50 0.00 0.005 71 57 57 0.48 0.48 0.00 0.004 72 60 57 0.50 0.48 0.02 0.005 73 60 62 0.50 0.52 0.02 0.005 74 45 56 0.38 0.47 0.09 0.004 75 55 51 0.46 0.43 0.03 0.004 76 64 57 0.54 0.48 0.06 0.005 77 58 58 0.49 0.49 0.00 0.005 78 58 55 0.49 0.46 0.03 0.004 79 64 61 0.55 0.51 <t< td=""><td></td></t<>	
62 62 63 0.52 0.53 0.01 0.006 63 62 58 0.49 0.49 0.00 0.005 64 64 65 0.54 0.54 0.00 0.006 65 66 65 0.55 0.54 0.01 0.006 66 65 60 0.54 0.50 0.04 0.005 67 63 60 0.53 0.50 0.03 0.004 69 50 60 0.42 0.50 0.08 0.004 70 60 60 0.50 0.50 0.00 0.005 71 57 57 0.48 0.48 0.00 0.004 72 60 57 0.50 0.48 0.02 0.005 73 60 62 0.50 0.52 0.02 0.005 74 45 56 0.38 0.47 0.09 0.004 75 55 51 0.46 0.43 0.03 0.004 76 64 57 0.54 0.48 0.06 0.005 78 58 55 0.49 0.46 0.03 0.004 79 64 61 0.55 0.51 0.04 0.006 81 63 65 0.53 0.54 0.01 0.006 81 63 65 0.55 0.04 0.006 83 56 58 0.47 0.49 0.02 <t< td=""><td></td></t<>	
63 62 58 0.49 0.49 0.00 0.005 64 64 65 0.54 0.54 0.00 0.006 65 66 65 0.55 0.54 0.01 0.006 66 65 60 0.54 0.50 0.04 0.005 67 63 60 0.53 0.50 0.03 0.005 68 47 65 0.39 0.54 0.15 0.004 69 50 60 0.42 0.50 0.08 0.004 70 60 60 0.50 0.50 0.00 0.005 71 57 57 0.48 0.48 0.00 0.004 72 60 57 0.50 0.52 0.02 0.005 73 60 62 0.50 0.52 0.02 0.004 75 55 51 0.46 0.43 0.03 0.004 76	
64 64 65 0.54 0.54 0.00 0.006 65 66 65 0.55 0.54 0.01 0.006 66 65 60 0.54 0.50 0.04 0.005 67 63 60 0.53 0.50 0.03 0.005 68 47 65 0.39 0.54 0.15 0.004 69 50 60 0.42 0.50 0.08 0.004 70 60 60 0.50 0.50 0.00 0.005 71 57 57 0.48 0.48 0.02 0.005 73 60 62 0.50 0.52 0.02 0.005 74 45 56 0.38 0.47 0.09 0.004 75 55 51 0.46 0.43 0.03 0.004 76 64 57 0.54 0.48 0.06 0.005 78	
6566650.550.540.010.0066665600.540.500.040.0056763600.530.500.030.0056847650.390.540.150.0046950600.420.500.080.0047060600.500.500.000.0057157570.480.480.000.0047260570.500.520.020.0057360620.500.520.020.0057445560.380.470.090.0047555510.460.430.030.0047664570.540.480.060.0057758580.490.490.000.0057858550.490.460.030.0047964610.540.510.030.0068064670.540.560.020.0068163650.530.540.010.0068266610.550.510.040.0068356580.470.490.020.0058470600.590.550.040.0068570610.590.510.080.006	
6665600.540.500.040.0056763600.530.500.030.0056847650.390.540.150.0046950600.420.500.080.0047060600.500.500.000.0057157570.480.480.020.0057360620.500.520.020.0057445560.380.470.090.0047555510.460.430.030.0047664570.540.480.060.0057858550.490.460.030.0047964610.540.510.030.0068064670.540.560.020.0068163650.530.540.010.0068266610.550.510.040.0068356580.470.490.020.0058470600.590.550.040.0068570610.590.510.080.006	
6763600.530.500.030.0056847650.390.540.150.0046950600.420.500.080.0047060600.500.500.000.0057157570.480.480.020.0057360620.500.520.020.0057445560.380.470.090.0047555510.460.430.030.0047664570.540.480.060.0057858550.490.460.030.0047964610.540.510.030.0068064670.540.560.020.0068163650.530.540.010.0068266610.550.510.040.0068356580.470.490.020.0058470600.590.550.040.0068570610.590.510.080.006	
6847650.390.540.150.0046950600.420.500.080.0047060600.500.500.000.0057157570.480.480.020.0057360620.500.520.020.0057445560.380.470.090.0047555510.460.430.030.0047664570.540.480.060.0057858550.490.460.030.0047964610.540.510.030.0068064670.540.560.020.0068163650.530.540.010.0068356580.470.490.020.0058470600.590.550.040.0068570610.590.510.080.006	
6950600.420.500.080.0047060600.500.500.000.0057157570.480.480.000.0047260570.500.480.020.0057360620.500.520.020.0057445560.380.470.090.0047555510.460.430.030.0047664570.540.480.060.0057858580.490.490.000.0057858550.490.460.030.0047964610.540.510.030.0068064670.540.560.020.0068163650.530.540.010.0068356580.470.490.020.0058470600.590.550.040.0068570610.590.510.080.006	
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	
7157570.480.480.000.0047260570.500.480.020.0057360620.500.520.020.0057445560.380.470.090.0047555510.460.430.030.0047664570.540.480.060.0057758580.490.490.000.0057858550.490.460.030.0047964610.540.510.030.0068064670.540.560.020.0068163650.530.540.010.0068266610.550.510.040.0068356580.470.490.020.0058470600.590.550.040.0068570610.590.510.080.006	
7260570.500.480.020.0057360620.500.520.020.0057445560.380.470.090.0047555510.460.430.030.0047664570.540.480.060.0057758580.490.490.000.0057858550.490.460.030.0047964610.540.510.030.0068064670.540.560.020.0068163650.530.540.010.0068266610.550.510.040.0068356580.470.490.020.0058470600.590.550.040.0068570610.590.510.080.006	
7360620.500.520.020.0057445560.380.470.090.0047555510.460.430.030.0047664570.540.480.060.0057758580.490.490.000.0057858550.490.460.030.0047964610.540.510.030.0068064670.540.560.020.0068163650.530.540.010.0068266610.550.510.040.0068356580.470.490.020.0058470600.590.550.040.0068570610.590.510.080.006	
7445560.380.470.090.0047555510.460.430.030.0047664570.540.480.060.0057758580.490.490.000.0057858550.490.460.030.0047964610.540.510.030.0068064670.540.560.020.0068163650.530.540.010.0068266610.550.510.040.0068356580.470.490.020.0058470600.590.550.040.0068570610.590.510.080.006	
7555510.460.430.030.0047664570.540.480.060.0057758580.490.490.000.0057858550.490.460.030.0047964610.540.510.030.0068064670.540.560.020.0068163650.530.540.010.0068266610.550.510.040.0068356580.470.490.020.0058470600.590.550.040.0068570610.590.510.080.006	
7664570.540.480.060.0057758580.490.490.000.0057858550.490.460.030.0047964610.540.510.030.0068064670.540.560.020.0068163650.530.540.010.0068266610.550.510.040.0068356580.470.490.020.0058470600.590.550.040.0068570610.590.510.080.006	
7758580.490.490.000.0057858550.490.460.030.0047964610.540.510.030.0068064670.540.560.020.0068163650.530.540.010.0068266610.550.510.040.0068356580.470.490.020.0058470600.590.550.040.0068570610.590.510.080.006	
7858550.490.460.030.0047964610.540.510.030.0068064670.540.560.020.0068163650.530.540.010.0068266610.550.510.040.0068356580.470.490.020.0058470600.590.550.040.0068570610.590.510.080.006	
7964610.540.510.030.0068064670.540.560.020.0068163650.530.540.010.0068266610.550.510.040.0068356580.470.490.020.0058470600.590.550.040.0068570610.590.510.080.006	
8064670.540.560.020.0068163650.530.540.010.0068266610.550.510.040.0068356580.470.490.020.0058470600.590.550.040.0068570610.590.510.080.006	
8163650.530.540.010.0068266610.550.510.040.0068356580.470.490.020.0058470600.590.550.040.0068570610.590.510.080.006	
82 66 61 0.55 0.51 0.04 0.006 83 56 58 0.47 0.49 0.02 0.005 84 70 60 0.59 0.55 0.04 0.006 85 70 61 0.59 0.51 0.08 0.006	
8356580.470.490.020.0058470600.590.550.040.0068570610.590.510.080.006	
8470600.590.550.040.0068570610.590.510.080.006	
85 70 61 0.59 0.51 0.08 0.006	
86 64 61 0.54 0.51 0.03 0.005	
87 60 61 0.55 0.51 0.04 0.005	
88 62 53 0.52 0.44 0.08 0.005	
89 60 55 0.50 0.46 0.04 0.005	
90 70 64 0.59 0.54 0.05 0.006	
91 70 61 0.59 0.51 0.08 0.006	
92 61 57 0.51 0.49 0.02 0.005	
93 65 65 0.54 0.54 0.00 0.006	
94 61 63 0.51 0.53 0.02 0.006	
95 60 64 0.50 0.54 0.04 0.005	
96 62 65 0.52 0.54 0.02 0.005	
97 61 66 0.51 0.55 0.04 0.005	
98 62 60 0.52 0.50 0.02 0.005	
99 61 55 0.51 0.46 0.05 0.005	
100 57 48 0.48 0.40 0.08 0.004	
101 62 60 0.52 0.50 0.02 0.005	
102 63 65 0.53 0.54 0.01 0.006	
<u>103</u> 64 66 0.54 0.55 0.01 0.006	

Tab	le 1	. Cor	ntd.

104	62	53	0.52	0.44	0.08	0.005
105	64	50	0.53	0.42	0.11	0.005
106	60	68	0.50	0.57	0.07	0.006
107	56	66	0.47	0.55	0.08	0.005
108	61	70	0.51	0.58	0.07	0.006
109	46	56	0.39	0.47	0.08	0.004
110	61	56	0.51	0.47	0.04	0.005
111	67	65	0.56	0.54	0.02	0.006
112	61	61	0.51	0.51	0.00	0.005
113	50	60	0.41	0.50	0.09	0.005
114	52	62	0.43	0.52	0.09	0.004
115	60	66	0.50	0.55	0.05	0.005
116	61	67	0.51	0.56	0.05	0.006
117	60	68	0.50	0.57	0.07	0.006
118	57	60	0.48	0.50	0.02	0.005
119	60	70	0.50	0.58	0.08	0.006
120	62	66	0.43	0.55	0.07	0.005
121	53	61	0.44	0.51	0.06	0.005
122	55	65	0.46	0.55	0.09	0.005
123	60	64	0.50	0.54	0.04	0.005
124	58	61	0.49	0.51	0.02	0.005
125	65	60	0.55	0.50	0.05	0.005
126	65	54	0.55	0.45	0.10	0.005
127	67	55	0.56	0.46	0.10	0.005
128	62	60	0.52	0.50	0.02	0.005
129	62	57	0.52	0.48	0.04	0.005
130	62	55	0.52	0.46	0.06	0.005
131	60	60	0.50	0.50	0.00	0.005
132	63	64	0.53	0.54	0.01	0.006
133	65	54	0.54	0.45	0.09	0.005
134	66	63	0.55	0.53	0.02	0.006
135	60	68	0.50	0.57	0.07	0.006
136	62	66	0.52	0.55	0.03	0.006
137	63	65	0.53	0.55	0.02	0.006
138	65	63	0.54	0.53	0.01	0.006
139	61	64	0.51	0.54	0.03	0.005
140	63	65	0.53	0.55	0.02	0.006
141	62	54	0.52	0.45	0.07	0.005
142	61	64	0.51	0.54	0.03	0.006
143	61	61	0.51	0.51	0.00	0.006
Total	8852	8923			7.860	1.045

Nigeria. This tends to lend support to the value and efficacy of the procedure followed in arriving at the final teaching practice (TP) scores for each student-teacher. Other institutions that are training teachers should try as much as possible to make use of this approach. It is therefore obvious from the results of the study that inter-rater of teaching practice students through interaction analysis of raters or observers is not an exercise in futility as it provides a means of obtaining the true score of these practicing teachers. This in no doubt will encourage the students to put their best effort in the teaching practice. A reliability of 0.92 obtained from the result shows a valid and reliable score of the process.

REFERENCES

American Educational Research Association (1999). Standards for educational and psychological testing. Washington, DC: AERA.

Aremu A (2002). Teaching Strategies for Nigerian Secondary Schools. Published by Power- House.

- **Bakker N (2008).** Design and evaluation of video portfolios: Reliability, Generalizability and Validity of an authentic performance assessment for teachers. ICLON PhD Dissertation Series. Netherlands Organization for Scientific Research. Leiden: Mostert & Van Onderen.
- Barrett S (2001). The impact of training on rater variability. Int. Educ. J. 2(1):49-58.
- Beesley A (2009). Measuring classroom assessment with a work sample. Denver, Co: Mid-Continent Research for Education and Learning.
- Borko H, Stecher BM (2005). Using classroom artifacts to measure instructional practices in middle school mathematics: A two-state field test. Los Angeles, CA: National Centre for Research on Evaluation, Standards and Student Testing.
- Chigbu ED, Idoko NA (2013). The role of academic libraries and librarians in knowledge Management and realization of vision 20:2020 in Nigeria. Int. J. Lib. Inform. Sc. 5(10):351-361.
- Clare L (2000). Using teachers' assignments as an indicator of classroom practice: CSE Technical Report. Los Angeles, CA: Centre for Research on Evaluation, Standards and student testing
- Daley A, Kim H (2001). A study of the Bloom's Strategy for Mastery learning Seoul: Korean Institute for Research into the Behavioural Science.
- **Dandeago KI (2012).** Management Professionalism: "The Missing Link for Nigeria's Sustainable Development" Glob. J. Manage. Sci. Technol. (1):5-9.
- **De-ville L (2010).** Perceived Relevance of Teaching Practice Exercise in the Teacher Training Program by Nigerian University Undergraduates: Afr. J. Cross-cultural Psychol. Sport facilit. 9:49-60.
- **Dibu-Ojerinde OO, Jegede PO (1999).** Estimating Essay Scoring Reliability by combining Experimental Design and Scores' Resampling. Ife J. Behav. Res. 1(1):41-48.
- **Dymond J (2008).** Becoming a Teacher. Buckingham (Philadepia): Open University Press.
- Federal Republic of Nigeria (2004). National Policy on Education. Lagos: NERDC Press.
- Flanders NA (1967). Interaction Analysis: Theory, Research and Application. Addison-Wesley Publishing Company.
- Gorman RL, Rentsch O (2009). Estimation of the reliability of ratings. *Psychometrika*, 16:407-424.

- Haggard EA (1958). Intra-class correlation and the analysis of variance. NY: Dryden.
- Henry S, Grimm L, Pianta B (2010). "Knowledge Management for higher education" Erik Digest, Virginia University. http://www.emwrald.insight.com
- Hopkins KD, Moore WS (1993). Clinical supervision: A practical guide to student teacher Supervision. Madison, WI: Brown & Benchmark Publishers.
- Hoyt H J, Kerns KL (1999). Research in Educational Assessment in Africa: Challenges for the 21st Century in Lekofa, J. and Matsoro, L. M. (eds) Educational Research for Development Challenges for the 21st Century. Roma, Lesotho, University of Leosotho.
- Imogie AJ (1990). Counselling for Quality Assurance in Education. A key Address being delivered on the occasion of 26th Annual Conference of CASSON, University of Benin, Benin City.
- Joint Admissions and Matriculations Board (2005). UME/DE Brochure: Guidelines for Admissions to First Degree Courses in Nigerian Universities, 2005/2006 Session. Lagos: Author.
- Lumley J, McNamara K (1995). Inter-rater Reliability. Res. Measure. Trans. 5(3):166.
- Maduabum L, Abah O (2004). Lecturers' Assessment of Teaching practice Exercise in Nigerian Universities. J. Educ. Res. 6:65:74.
- Offorma GC (2005). Curriculum for Wealth Creation: Seminar Paper presented at World Council for Curriculum and Instruction Federal College of Education, Kano.
- Philip S, Francis L, David B, Dennis WC, Atara S, Micheal W (2000). Better Supervision, Better Teaching. A Handbook for Teaching Practice Supervision Hong Kong: Hong Kong University Press.
- **Tinsley L, Weiss K (2000).** A Guide to Teaching Practice. London: Routledge Falmer.

http://sciencewebpublishing.net/jerr