European Journal of Educational and Development Psychology
Vol.12 No.3, pp.1-11, 2024
2024ISSN: 2055-0170(Print),
Online ISSN: 2055-0189(Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development-UK

Comparative Analysis of Reliability Estimates of Assessment Methods by Teachers in Secondary Schools in Port Harcourt Metropolis

[1]Amadioha, A. (Ph.D) and [2]Uwah, I., V. (Ph.D)
[1&2]Department of Educational Psychology, Guidance & Counselling
Faculty of Education, University of Port Harcourt
Email: 2idongesituwah6@gmail.com

doi: https://doi.org/10.37745/ejedp.2013/vol12n3111 Published August 11, 2024

Citation: Amadioha, A. and Uwah, I., V. (2024) Comparative Analysis of Reliability Estimates of Assessment Methods by Teachers in Secondary Schools in Port Harcourt Metropolis, *European Journal of Educational and Development Psychology*, Vol.12 No.3, pp.1-11

Abstract: *The present study was on comparative analysis of reliability estimates of assessment methods by teachers in secondary schools in Port Harcourt Metropolis. Three objectives and research questions guided the study. Comparative research design was adopted in the study. The population of the study consisted of 3,340 SS2 students drawn across public schools in Port Harcourt metropolis. A sample of 100 students drawn across 10 schools in the area with the help of multi-stage sampling process was used in the study. The researcher developed two forms multiple-choice test as measuring instruments. The first was a 60 item "Formative Assessment Test" (FAT) while the second was a 50 item "Summative Assessment Test" (SAT). The FAT was administered on the respondents during the course of the instruction within the term while SAT was administered at the end of the third term as part of the promotional examinations. Test blueprint was adopted to ensure that the two instruments have content validity while factor analysis was used in determining the construct validity. Administrations of the instruments were done face to face in the class with the help of the classroom teachers. Data analysis method included KR_{20}, split-half and test retest methods of reliabilities. For the test-retest, the researchers repeated the process for both the formative (FAT) and summative test (SAT) after a period of two weeks interval. Result of the study showed that when KR_{20} was used, summative assessment had higher reliability that formative assessment (SAT-0.81>FAT-0.52). When split-Half was used, summative assessment also had higher reliability that formative assessment (SAT-0.87>FAT-0.58). However, when test retest was applied, formative assessment had higher reliability than summative assessment (FAT-0.78>SAT-0.55). Based on this, it was recommended among others that teachers and test developers should rely more of summative assessment in the course of establishing the cognitive abilities of students.*

Keywords: Reliability, Assessment, Formative assessment, summative assessment. KR_{20}, Split-half, Test-retest.

European Journal of Educational and Development Psychology
Vol.12 No.3, pp.1-11, 2024
2024ISSN: 2055-0170(Print),
Online ISSN: 2055-0189(Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development-UK

INTRODUCTION

Education is the backbone of a country's development (Bhuiyan, 2019). Education plays an important role in public health, social mobility, equity, and better opportunities for employment. Rivers state and Port Harcourt metropolis is currently taking different measures to improve the quality of education. Universities are expanding all over the country. The number of private colleges is also increasing from time to time. Etzkowitz (2002) noted that the expansion of universities by themselves can't support and facilitate the development of the country. This is due to the fact that it is not all about the quantity, it is also about the quality of the education. The education quality of most African countries including Port Harcourt is at a lower level. Skutnabb-Kangas and Heugh (2013). A sound education is a powerful means to achieve national integration, industrial and economic progress, and social awareness and commitment to a nation. Education can create a great sense of responsibility among citizens toward the state and help them develop attitudes that convey significant values to the youths, enabling them to become functional members of society. Basic education in Nigeria prepares learners for practical and functional living in the society. Education is a system that involves teaching and learning activities, and it can be examine through the system of examination or better still as will be used in the course of this paper, "assessment".

Assessment is the systematic collection, review, and use of information about educational programs undertaken for the purpose of improving student learning and development (Palomba and Banta 1999). *Assessment* is used to understand the state or condition of learning. An instructor assesses learning through both observation and measurement in an attempt to better understand students' learning in a course. This includes collecting evidence, both graded and non-graded, about a students' progression in the course.

Ukwuije and Opara (2013) defined assessment as "a systemic process of determining the extent to which instructional objectives are met". Salvia and Ysseldyke (cited in Onunkwo 2002) defined assessment as "a process of collecting information for the purpose of specifying and verifying problems as well as for making decision about students". To Ukwuije (2012), it is the procedure of assigning value to the learning achieved during and at the end of a course, it can attempt by the teacher to gain knowledge of his students' competencies. Therefore, assessment is a process of determining how the learners have mastered the stated instructional objectives which will be used for decision making on the student. Assessment is classified into two types; formative and summative assessment.

Assessment as noted by Neukrug and Fawcett (2019) helps to give feedback on the teaching activities to determine the level of knowledge attainment or the rate of retention by the learners. Also, it serves as a standard for evaluating the effect of the learning process. In Nigeria, assessment is an indispensable instrument used to appraise learners at the basic level of education and to

determine their academic achievements, thereby giving room for self-assessment (Adisa, 2023). In addition, according to Johnson (2020), assessment helps to decide whether the set educational objectives in the school have been achieved. According to Opara and Uwah (2017), there are two basic forms of assessment. This includes formative assessment as well as summative assessment. Formative Assessment is any task or activity that produces feedback for students about their learning in a course. It typically does not contribute to the grade in a course (Irons, 2008). This is often referred to as 'assessment for learning'. The focus of formative assessment is on student learning. These activities provide students a chance to check their understanding of concepts, reflect on these understandings, and identify areas they may need to clarify in the course. In this type of assessment, an entire examination is done throughout a course or project. This also involves the process of getting detailed information in order to specifically verifying problems as well as for making decisions about students during a course of study or programme. Ukwuije (2009) opined that observational techniques, oral questions and answers are frequently used in formative assessment the purpose of formative assessment is for improving instruction, the results are not used to compute the final grades of students. It could be done unit by unit on daily or weekly basis. As stated earlier, formative assessment is done at the school level. This involves a more comprehensive, continuous systematic and diagnostic and integrative assessment process. This pattern evolves from classroom situation that requires an active involvement of the students. This also lays much emphasis on learning (Oviawe & Ojo 2008). Again, Khodabakhshzadeh, Kafi and Hosseinnia (2018) noted that some review studies have been done by researchers such as Crooks (1988) and Black (1993b) which have revealed some disadvantages regarding the formative assessment done by teachers in classroom settings. Among all it could be mentioned that classroom evaluation practices generally encourage artificial, rote learning, concentrating on recall of isolated details, usually items of knowledge which students soon afterwards forget. In addition, instructors do not usually review the assessment questions that they use and do not discuss them critically with their peers, so there so little reflection on what is being assessed. Also, the grading function is mostly often over-emphasized and the learning function as a matter of fact under-emphasized. Moreover, to examine the reliability of the scale, KR_{20} was used. The reliability coefficient was 0.855 for total scale, which shows the scale enjoys high reliability (Khodabakhshzadeh, Kafi & Hosseinnia, 2018).

On the other hand, in summative assessment, individuals carry thus out at the end of a particular project or course. It focuses on assessing students at the end of a term and of a class. This assessment is used to assign results of summative assessment are used for placement to higher classes. Summative assessment focuses on achievement. "This is the type of assessment in which the students provide genuine response in the area that is closest to the intended criteria for the examiner to make inferences" (Popham, 2000). Through summative assessment, it is an assessment which is set and marked by the school (teacher). Smith and Kubacka (2019) stated that teachers are considered as the key actors at schools, contributing and shaping the students' development and learning. It is a common belief that good teachers are good test developers.

However, in preparing a test, some of the following obstacles are found by the teachers Bijsterbosch (2016) averred that there is a tendency for teachers to use test items in class summative tests that focus on memory and memorization; therefore, it influenced the teacher's disposition towards summative assessment as well. Second, under increasing pressure to improve the students' scores, teachers are more likely to use shortcuts or limit the instruction to test certain contents and activities. Third, teachers rarely discuss or share their practices with colleagues at the same school. Scores obtained from summative assessments tend to predict invalid and unreliable scores about students (Ahmad, 2020). According to Ukwuije & Opara (2013), students get the marks and feedback regarding the assessment as against the external assessment where the students only get a mark. They have no idea how they actually performed. It is summited by Opara and Uwah (2018) that the method of assessment adopted by the classroom teacher may determine to some extent the reliability of such test

Hence, Li (2016) stated that reliability is the degree to which an instrument yields consistent results. Common measures of reliability include internal consistency, test-retest, and inter-rater reliabilities. Internal consistency reliability looks at the consistency of the score of individual items on an instrument, with the scores of a set of items, or subscale, which typically consists of several items to measure a single construct.

Test-retest is one of the methods of determining reliability of any assessment. It measures the correlation between scores from one administration of an instrument to another, usually within an interval of 2 to 3 weeks. Unlike pre-posttests, no treatment occurs between the first and second administrations of the instrument, in order to test-retest reliability. A similar type of reliability called "*alternate forms"*, involves using slightly different forms or versions of an instrument to see if different versions yield consistent results. Finding a correlation coefficient for the two sets of data is one of the most common ways to find a correlation between the two tests. Test-retest reliability coefficients (also called coefficients of stability) vary between 0 and 1. According to Glen, (2016), measuring reliability for two tests, individuals should use the Pearson Correlation Coefficient.

Kuder-Richardson Formula 20, or KR-20, is a measure of reliability for a test with binary variables (i.e. answers that are right or wrong). Reliability refers to how consistent the results from the test are, or how well the test is actually measuring what you want it to measure. The scores for KR-20 range from 0-1, where 0 is no reliability and 1 is perfect reliability. The closer the score is to 1, the more reliable the test. Just what constitutes an "acceptable" KR-20 score depends on the type of test. In general, a score of above 0.5 is usually considered reasonable

Split-half reliability is another form of reliability. It tests for a single knowledge area is split into two parts and then both parts given to one group of students at the same time. The scores from both parts of the test are correlated. A reliable test will have high correlation, indicating that a

student would perform equally well (or as poorly) on both halves of the test. Split-half testing is a measure of internal consistency. This is how well the test components contribute to the construct that's being measured. It is most commonly used for multiple choice tests you can theoretically use it for any type of test even tests with essay questions. From these parameters, it is the belief of the researchers that the methods of assessment which the teachers adopt command some level of reliability.

Based on these results of Ekolu and Quainoo (2019), it may be deduced that the split-half method is sensitive to similar factors as Cronbach's alpha, unlike the KR 21 coefficient, the behaviour of which is quite different. The factors responsible for the different behaviour of KR 21 relative to the alpha and split-half methods are not fully understood and require further investigation. Hence, the researchers reasons that some of the methods of assessment adopted by teachers may have varied level of reliability even to a significant basis which in turn can affect the outcome of the test.

Aim and Objectives

The study aimed at investigating comparative analysis of reliability estimates of assessment methods by teachers in secondary schools in Port Harcourt Metropolis**.** Specifically, the study intends to

1. Compare the reliability indices of formative and summative assessment by teachers using KR_{20} method in secondary schools in Port Harcourt Metropolis.
2. Compare the reliability indices of formative and summative assessment by teachers using split-half method in secondary schools in Port Harcourt Metropolis.
3. Compare the reliability indices of formative and summative assessment by teachers using test retest method in secondary schools in Port Harcourt Metropolis.

Research Questions

The following research questions were asked to guide the researchers in the study.

1. What are the reliability index of formative assessment and summative assessment comparatively as determined using KR_{20} method in secondary schools in Port Harcourt Metropolis?
2. What are the reliability index of formative assessment and summative assessment comparatively as determined using split-half method in secondary schools in Port Harcourt Metropolis?
3. What are the reliability index of formative assessment and summative assessment comparatively as determined using test-retest method in secondary schools in Port Harcourt Metropolis?

METHODS

The present study used the comparative research design in the study. comparative design is a type of design that involves comparing two or more groups, cases or phenomena in order to identify similarities and differences. It aims to identify and understand trends, patterns and relationship between variables. According to Iranifard and Roudsari (2022) stated that comparative research is the study of similarities and differences between two or more cases. This study used the design because it will compare the reliability indices of formative assessment technique and that of summative assessment using various methods of reliability in order to see which yields more reliability. The population of the study includes SS2 students drawn across public schools in Port Harcourt metropolis. As at the time of the study, there were 3,340 SS2 across public schools in the area. A sample of 100 students drawn across 10 schools in the area with the help of multi-stage sampling process was used in the study. The researchers at stage one used simple random sampling by ballot to draw 10 public schools from the area. At stage two stratified non-proportionate sampling technique was used to draw 10 students across the ten schools drawn for the study. This gave a total of 100 students. Finally, purposive sampling techniques was used at stage three to focus only on SS2 students because the researcher believes they are not busy with the preparations of public SSCE examinations compared to SS3 students neither are they notice to the assessment process. The researcher developed two forms testing instruments. The first was the Formative Assessment Test (FAT) while the second was the Summative Assessment Test (SAT). The FAT is a 60 items 4-option multiple-choice mathematics which was designed by the teacher and administered on the respondents during the course of the instruction within the term. On the other hand, the SAT was equally designed using the 4 option multiple choice mathematics test. On the contrary, the researcher administered this at the end of the third term as part of the promotional examinations. The test contained 50 items still on a similar topics and syllabus as that of FAT. Test blueprint was adopted to ensure that the two instruments have content validity while factor analysis was used in determining the construct validity. Administration of the instruments were done face to face I the class with the help of the classroom teachers while KR_{20}, split-half and test retest methods were used in analysis of the test scores. For the test-retest, the researchers repeated the process for both the formative (FAT) and summative test (SAT) after a period of two weeks interval.

RESULTS

Out of the 100 respondents assessed, 94 representing 94% were successfully retrieved. The reason for this short in retrieval was as a result of loss of six scripts during the retest process.

Research Question One: What are the reliability index of formative assessment and summative assessment comparatively as determined using KR_{20} method in secondary schools in Port Harcourt Metropolis?

Table 1; shows KR_{20} Reliability estimates of formative assessment and summative assessment methods

	N	∑X	$\sum X^2$	ð	$ð^2$	I	∑Pq.	KR_{20}	Remark
FAT	94	1389	34317	4.07	16.56	60	6.58	0.52	Low Reliability
SAT	94	1419	37531	4.09	16.73	50	7.31	0.81	High Reliability

From the table, it could be deduced that calculation of KR_{20} was done separately for both FAT and SAT test. The respondent had sum of 1389 and 1419 respectively for FAT and SAT. Sum of squares were 34317 and 37531. Standard deviation value and the variance was 4.07; 16.56 and 4.09; 16.73 respectively. Both tests had 94 items (**i**). The sum of the proportions of students who got the items correctly and wrongly was 6.58 and 7.31 respectively. KR_{20} reliability for both test were 0.52 for the formative assessment (FAT) and 0.81 for the summative (SAT). From these scores, it could be seen that summative assessment had more KR_{20} reliability than the formative assessment.

Research Question Two: What are the reliability index of formative assessment and summative assessment comparatively as determined using split-half method in secondary schools in Port Harcourt Metropolis?

Table 2: shows Split-half Reliability Coefficient estimates of formative assessment and summative assessment methods

Test	Half items	N of Items	N	Rht.	Rft.	Remarks
FAT	30	60	94	0.41	0.58	Low Reliability
SAT	25	50	94	0.77	0.87	High Reliability

The table above shows that N of the half test for FAT was 30 while that of SAT was 25 making it a total N of items to be 60 for FAT and 50 for SAT respectively. Valid total N was 94 without any case exclusion. The table also revealed that split-half reliability of the half test (rht) for FAT was 0.41 and 0.77 for SAT. However, when spearman brown prophecy formula was used to substitute the half test, a Guttman reliability of the full test (rft) was 0.58 for FAT and 0.97 for SAT. Thus when compared it could be seen that split half produces a low reliability estimates for FAT and a high reliability index for SAT.

Research Question Three: What are the reliability index of formative assessment and summative assessment comparatively as determined using test-retest method in secondary schools in Port Harcourt Metropolis?

Table 3: shows test-retest reliability coefficient estimates of formative assessment and summative assessment methods via PPMC.

Test	N of Items	N	R	Remarks
FAT	60	94	0.78	Average Reliability
SAT	50	94	0.55	low Reliability

Table 3 shows that calculated r for formative test (FAT) is 0.78 which shows an average reliability while the summative test had a reliability coefficient of 0.55 which is a low reliability. From these indices, it could be seen that when using test retest method, formative assessment has a higher reliability index than the summative assessment. In all, it is seen that the reliability index of summative assessment in all the reliability estimates methods is higher except for test-retest method that that of formative assessment.

DISCUSSION OF FINDINGS

From findings one, it is revealed that summative assessment had more KR_{20} reliability than the formative assessment. The finding of the study means that test given at the end of the instruction has more dependability that test given during the course of the instructions. The findings also means that test at the end of instruction is more reliable than test given during the course of instruction. Furthermore, the finding also highlights the efficacy of Kuder-Richardson method of internal consistency in identifying the dependability of any measuring instrument. The finding of have also shown that where the assessment format may manipulate the dependability of testing instrument by not capturing its true position, KR_{20} has the ability to unfold such in any given test. This is the reason why there is a significant difference in the reliability indices as shown between the FAT and SAT assessment methods. The findings of the study also signify the important of relying on summative assessment methods in realizing the cognitive abilities of the students. The finding of the study is not surprising to the researcher in any way because he is aware that formative assessment method which often is impromptu and unorganized and which may not give the students the opportunity for adequate preparation may lack credibility compared to standard or summative assessment which may be a bit more organized and expected. The finding of the study is in line with that reported earlier by Khodabakhshzadeh, Kafi and Hosseinnia (2018) who reported significant higher reliability estimated of test as determined by KR_{20}.

Research finding two has also shown that split half produces a low reliability estimates for FAT and a high reliability indices for SAT. once again, the findings shows split-half reliability estimates has the ability to differentiate the reliability estimates of test given at the end of the instruction as well as those given during the course of the instructions. It means that split-half is capable of determining that a test at the end of instruction is more reliable than test given during the course of instruction. The finding of the study also shows that split half method of determining reliability is also very effective and precise in identifying the various reliability level of any measuring instrument. Again, in a situation where any assessment format may affect the reliability of the measuring instruments, the present finding has proved that the split half method of reliability has the ability to unfold such in any given test. The findings of the study also signify the important of summative assessment methods in realizing the cognitive abilities of the students instead of just depending on formative assessment only. The finding of the study is not surprising to the researchers in any way because they are aware that only formative assessment methods may not guarantee adequate or standard items that can guarantee reliability compared to standard or summative assessment which may be a bit more organized and expected. The finding of the study is in line with that reported earlier by Ekolu and Quainoo (2019) who reported that the split-half method is sensitive to similar factors as Cronbach's alpha, unlike the KR 21 coefficient.

From research finding three, it is revealed that using test retest method, formative assessment has a higher reliability index than the summative assessment. This means that giving students test and allowing time before one re-administer the same test to them can help in yielding a better reliability. The reason for the finding could be that the students may have mastered the test and may have developed some level of competences and maturity as well as previous knowledge.

From the analysis, it could also be seen that the relationships between the reliability indices obtained using all three reliability methods shows some level of consistency between KR_{20} a well as the split-hafmethod of reliability as both appeared to be higher in terms of the SAT assessment format. It also indicates that summative assessment seems to have more reliability across the reliability methods indicating that summative assessment is more reliable in assessing student's cognitive achievements. This means that the split-half method is strongly correlated with KR_{20}. Interestingly, all the three methods gave reasonable reliability coefficients though the test retest at some point was very low. The low reliability coefficients obtained for some modules appear to be explained by the very low inter-item relatedness found in the test items.

CONCLUSION

The internal consistency measurement techniques employed comprised the KR_{20}, Split-half methods as well as the test-retest method. The test was designed in formative and summative format. From the review, assessment format has a role to play in the extent of reliability of test instruments. Despite the heterogeneity and small number of test items in both formative and

summative test, majority (KR_{20} and split-half methods exhibited meaningful estimates of reliability coefficient, giving values high enough to guarantee reliability. However, the test-retest method gave low coefficients except for formative assessment. The low values obtained are attributed to poor inter-item relatedness of the test items. The KR_{20} and split-half coefficients are effective in establishing reliability of test in whatever format.

Recommendations

Generally, since these differences exist, it is recommended that;

1. Teachers and test developers should rely more of summative assessment in the course of establishing the cognitive abilities of students.
2. Test developers and teachers should ensure they apply more than one reliability measure when trying to establish the reliability index of the test.
3. Teachers should avoid or at least limit the use of test-retest methods and they are a lot factors that may cofound or interfere with the reliability index of test established using this method of reliability.

REFERENCES

Bijsterbosch, J. Van Der Schee, W. Kuiper, and T. Beneker, (2016).Geography teachers' practices regarding summative assessment: A study of pre- vocational education in the Netherlands," *RIGEO*, 6, (2),118–134.

Iranifard, E., & Roudsari, R. L. (2022).Comparative Research: An Old Yet Unfamiliar Method. *Journal of Midwifery and Reproductive Health*, *10*(3), 3317-3318. doi: 10.22038/jmrh.2022.66873.1954.

Ekolu, S., O. & Quainoo, H. (2019). Reliability of assessments in engineering education using Cronbach's alpha, KR and split-half methods , *Global Journal of Engineering Education*, 21,(1) 24-29.

Khodabakhshzadeh, H., Kafi, Z., & Hosseinnia, M. (2018). Investigating EFL Teachers' Conceptions and Literacy of Formative Assessment: Constructing and Validating an Inventory. *International Journal of Instruction*, 11(1), 139-152. https://doi.org/10.12973/iji.2018.11110a

Smith, W. C. & Kubacka, K (2017)."Education policy analysis archives Appraisal Systems," *Educ. Policy Anal. Arch.*, vol. 25, no. 86, pp. 1–29.

Ahmad, Z. (2020). Summative assessment, test scores and text quality: A study of cohesion as an unspecified descriptor in the assessment scale," *Eur. J. Educ. Res.*, vol. 9, no. 2, pp. 523–535, 2020.

Opara, I,. J. & Uwah, I., V. (2017). Effect of Test Arrangement on Performance in Mathematics among Junior Secondary School Students. *The British Journal of Education. 5*(8) 1-9.

Bhuiyan, A., A., M. (2019). Financing education: A route to the development of a country. J Educ Dev 7: 209–217.

Etzkowitz H (2002) Incubation of incubators: innovation as a triple helix of university-industry-government networks. Sci public policy 29: 115–128.

Skutnabb-Kangas T, Heugh K (2013) Implications for multilingual education: Student achievement in different models of education in Ethiopia. In: Multilingual Education and Sustainable Diversity Work. Routledge, pp 257–280.

Crooks, T. J. (2002). *Educational assessment in New Zealand* Schools. Assessment in Education, 9(2) 237-253.

Ukwuije, R. P. I. & Opara I. M. (2013) School base assessment implicational transformation. *Nigerian journal of Educational Research and Evaluation, 12, (1), 9.*

Ukwuije, R. P. I. (2012). *Educational assessment: A sine qua non for quality education*. Inaugural lecture series No. 83 University of Port Harcourt, River State.

Onunkwo, G. I. N. (2002). *Fundamentals of educational measurement and evaluation*. Owerri Cape Publishers International.

Oviawe, J. & Ojo, K. E. (2008) Perceptions of vocational educators on the relevance of school based assessment in pre-vocational studies in universal basic education. *Nigerian journal of education Research and Evaluation 8(1), 31-39.*

Neukrug, E. , & Fawcett, R. (2019). Essentials of testing and assessment: A practical guide for counselors, social workers, and psychologists (4th ed.). Cengage Learning. Retrieved on 1st October, 2023 from https://www.academia.edu/24201445/Nelson_Jones_Theory_and_Practice_of_Counselling_and_Psychotherapy.