A COMPARATIVE ANALYSIS OF TEST RE-TEST AND EQUIVALENT RELIABILITY METHODS

AJAYI B. K

E-mail: bolanlekike@yahoo.com
+2348037314475, +2347069456122.
INSTITUTE OF EDUCATION, EKITI STATE UNIVERSITY,
ADO-EKITI, EKITI- STATE. NIGERIA

ABSTRACT

The study examined the comparative analysis of test re-test and equivalent reliability methods. Four hundred students were selected as sample from four Secondary Schools in Ado-Ekiti Local Government Area, Ekiti State. Forty items in Agricultural Science objective test was constructed by the researcher. The objective test was administered to four selected schools based on mixed and single Schools. The scores were compared across the sampled schools. The instrument was teacher made constructed items. Means and standard deviations were used as descriptive analysis. The study identified the most appropriate reliability method used for Agricultural Science objective test items by educational researcher. Two hypotheses were formulated while correlation co-efficient was used. The hypothesis was held significant at 0.05 probability level. The findings show that significant differences exist in the two reliability methods. The study also revealed that the educational researchers used appropriate instruments.

The following recommendations were made; adequate awareness should be created among teachers on the relevance of tests items that will give consistent result. The educational researcher should try and use the appropriate reliability method to carry out their research.

Key words: Reliability, Test-Retest, Equivalent or Parallel form, Consistency and Coefficient.

1.1 INTRODUCTION

Reliability is the extent to which individual differences in test scores are attributed to true differences in the characteristics under consideration and the extent to which they are attributed to chance errors. Kolawole (2011) stated that, a test is said to be reliable,if it measures what it is supposed to measure consistently. It also a consistency of scores obtained by the same individuals on different sets of equivalent items. In order to obtain valid results, reliability of measurement is needed. A test cannot be of much value if the score it yields for an examinee today is very different from the score it would have yielded for him under similar conditions yesterday or tomorrow.

According to Kolawole (2011) a test is considered reliable, if the testee performs on the same re-examined with the same test at a different time without additional instruction, if the testee scores the same mark on a different form of the same test or consistent decisions are made using the test results. The use of research instruments appears to be inevitable in empirical researches, hence the need to ensure the reliability.

Ajayi (2007) stated that the tape rule measures the length of a room as it does that of a field in a consistent way such measuring procedure is said to be reliable. There are various methods of estimating reliability, according toKolawole (2011) which are test re-test method or measure of stability, the equivalence form method, split half method and KuderRichadson KR_{20} and KR_{21}. According to Alonge(1989) mentioned some of the methods of estimating reliability as follows; Test re-test, equivalent form method or parallel form, internal consistency method which is made up of; split-half and Kuder Richardson twenty and twenty one. The researcher is concerned onthe comparative analysis of Test re-test and equivalent or parallel form methods.

The study is to present ways for obtaining acceptable procedures for reporting reliability techniques. It will also examine the reliability coefficient, which is more reliable for specific categories of items and tests formats.

Test re-test reliability method as it was explained by Oladunni(1996) as a test instrument administered to a particular group of students at two different occasions. The time for the repetition of the test may be some days, weeks or months, and the scores obtained on two occasions are then correlated. If the students were able to maintain their relative positions in both administrations the results are highly stable, if otherwise the results have two stabilities. In a situation where a test is administered to a group of testee,if in the first group, some scored high in the test,after two weeks of the first administration of the test, the same group of students also scored high, which means the results are highly stable.Suppose thetestees take an intelligence every year in one school district and every two years in another. The teacher in the first district probably assume that intelligence test scores remain stable over-one year period but that retesting is necessary because individuals are likely to respond inconsistently if longer periods of time exist between tests. Also, the teachers in the second district assume that scores marked do not fluctuate over one year period and it is unnecessary to test more frequently than every two years. Which implies that the two school districts are correct but depends on the correlation between intelligence test scores obtained in one year and two years apart.

Alonge (1989) opined that test re-test reliability method refer to a test administered twice with a time lapse between the two tests, this makes it possible for each examinee to have two scores. The two scores are then correlated and a reliability coefficient is got. The Pearson Product Movement Correlations equation is used.

Ajayi (2007) explained that the test is reliable when the test produces identical results. Ajayi (2012) went further that achievement is very paramount and central to the current Nigerian system of education at all levels.

Okpala et al (1993) described equivalent forms method as overcome the problem of appropriate time interval two parallel or alternate forms of an evaluation instrument which are administered concurrently to the same students. The correlation coefficient computed using the two sets of scores is a measure of equivalence of the two tests.

Clinton (1974) in his own contribution refers to equivalent as forms of a test that are often used as pre-test and post-test to assess gains in skill after an instructional period. The first test is given to show where the students are now in the skill area; the instruction then follows, post-test – the equivalent form is given to see how much gain in scores occurred after instruction. When we employ two forms of a test, we are virtually using two equal halves of the same instruction. Equivalent forms are instructed to be a like in degree and range of difficulty. They cover the same areas of knowledge and skill with different items using equivalent forms eliminates or reduces to a minimum, the practices effect that is present when the same test is given twice.

Gilbert (1974) in his own contribution refers to equivalence as determined by constructing two or more forms of an examination and administers them to same persons at about the same time. The correlation between scores on the forms is a measure of their equivalence.

Gronlund (1985) added that only the equivalent form method with an intervening time period between tests takes into account all three types of consistency.

1.2   STATEMENT OF THE PROBLEM

The study presents ways for obtaining acceptable procedures for reporting reliability techniques. It will also examine the reliability coefficient that is more reliable for specific categories of items and tests formats. Specifically this study attempts to answer the following questions.

(i)   What are the available reliability procedures in research?
(ii)   Which of these reliability procedures is more appropriate for agricultural objectives test items?
(iii)   What is the pattern of significance of the Agricultural Science reliability coefficients?

1.3   RESEARCH HYPOTHESES

The test of the hypotheses was presented in turn with respect to the reliability methods.

Ho_1: The test-retest reliability coefficients are not significant for the agricultural achievement test in all the (four schools) that is in school one, school two, school three and school four.

Ho_2: The parallel form reliability coefficients are not significant for the agricultural achievement test in all the four schools that is, in school one, school two, school three and school four.

1.4   RESEARCH METHODS

The essential characteristics of the schools exit as well as the reliability methods that gave the researcher the opportunity of describing this phenomenon, hence, the use of a descriptive method. The population comprised of one hundred (100) students per school, however, made the total

numbers of four hundred (400) students that constitute the sample for the study. Four (4) different Junior Secondary Schools, three students (JSS 3) from Ado-Ekiti Local Government Area of Ekiti State both male and female were selected randomly.

The research instruments used for this study are the forty items, the researcher constructed objective test in Agricultural Science which is a teacher made achievement test. A forty-item test was administered to one hundred (100) students per school making a total of four hundred (400) students. The scripts were collected and graded in other to obtain the scores from these test for the purpose of these reliability study. According to Jimoh and Omorege (cited in Nwadiani2005), view examinations as the best tool for an objective assessment and evaluation of what learners have acquired after a period of schooling, thus any action that undermines examinations poses a great threat to the validity and reliability of examination results and certificate.

The test re-test reliability of the instrument was established by re-administering the test on a group within an interval of two weeks.Mean and standard deviations were used for descriptive analysis while the two hypotheses generated were tested with the use of the Pearson Product Correlation Coefficients.For equivalent, the test was administered to the same students con-currently, the correlation coefficient computed using the two sets of scores.

1.5   FINDINGS AND DISCUSSION

The study examined the status of two reliability methods. This was done through the use of descriptive analysis and hypotheses testing. A comparative analysis was made to test the means and standard deviation of the reliability methods. There was significant difference in the reliability coefficient in each school for the Agricultural Science test items. There were three questions discussed and two hypotheses.

Q1: What are the available reliability procedures in research? The table 1 shows four schools with the two methods, test re-test and parallel or equivalent methods. The test was administered twice to testee in the four schools which are test(1) and test(2) going through the score in the table, the test was reliable in mean and standard deviation in equivalent or parallel form method, the test was administered to the testees concurrently in the four schools. The mean and standard deviation result in School(3) under P1 has the lowest scores while school(1) has the highest mean score and lowest standard deviation. In P2 school(1) also has the highest mean and lowest standard deviation while school(2) has the highest standard deviation.

Q2: Which of these reliability procedures is more appropriately for agricultural objectives test items?In table 2, this question was answered using the scores obtained from the test administered on four hundred(400) students used in this study.In table (2) where we have two methods and four schools, in School (1) the reliability coefficient has the highest score in parallel or equivalent form with 0.2034 followed by Test re-test form with 0.1105.

In school (2) the score for Test re-test was 0.5076 which was a bit higher, followed by equivalent or parallel form with 0.1508.

In school (3) the most appropriate reliability coefficient for Agricultural Science test items was Test re-test of 0.2880 and equivalent or parallel form with 0.2640.

In school (4) the equivalent or parallel form method has scores of 0.5850 followed by Test re-test with the scores of 0.5444.

Test re-test has the highest degree of reliability coefficient for Agricultural Science test items.

Q3: What is the pattern of significance of the agricultural science reliability coefficients?In table 3, the correlation in longitudinal was in 10,000 and schools were on the horizontal level. When looking at the graph, test re-test was low in school (1) but in equivalent or parallel form, it was increased to higher value. In school (2), under test re-test the figure was high but on getting to equivalent or parallel form, the number dropped. In school (3), the test re-test increased to higher value and equivalent or parallel with very lowest value. In school (4) equivalent or parallel form was a bit higher than test re-test. It could be observed that test re-test has the highest correlation scores and equivalent with least correlation.

Ho$_1$: The test-retest reliability coefficients are not significant for the agricultural achievement test in all the (four schools) that is in school one, school two, school three and school four.

In table 4, theHo$_1$: correlation Coefficients was used,the calculated values for the Pearson r were more than r-table at 0.05 level of significance in three different schools. The null hypothesis is therefore rejected. That is, the test-retest reliability coefficient for the agricultural achievement test were significant in most of the schools.

The null hypothesis on test re-test reliability for an agricultural achievement test can be represented thus:Ho: $r_{trr} = 0$

Ho$_2$: The parallel form reliability coefficients are not significant for the agricultural achievement test in all the four schools that is, in school one, school two, school three and school four.In table 5,the r-calculated of 0.2034, 0.2640 and 0.5850 were more than the r-table at

0.05 level of significance except in School (2). Hence the hypothesis was rejected. Therefore the equivalent or parallel form reliability for the agricultural achievement test were significant in some Schools

The results in this analysis both the descriptive analysis and testing of the hypotheses showed that most researchers used test re-test reliability method. The reliability methods can be ranked in order to the common usage. This is in consonance with theory as exemplified by Gronlund (1981), Gilbert (1974) and Kenneth and Stanley (1972) that the reliability methods differ in assumptions and functioning, hence in strength when applied in different environments.

In test re-test, r-table was greater than r-calculated at 0.05 levelof significances, which means there is no significant difference in test re-test reliability coefficient for the agricultural achievement test in all the schools.

For equivalent form, the r-calculated were greater than the r-table with the exception of school two (2). The hypothesis was rejected. Therefore the equivalent form is zero for agricultural achievement test.

1.6 RECOMMENDATIONS

The researcher made the following recommendation based on the findings.

Test and measurement experts should come up with blue prints of reliability procedures and assumptions for their use to assist beginning researchers in selecting appropriate procedures in carrying out research.

The practice of writing projects reports without reporting the reliability coefficients is inimical to true dissemination of research findings, efforts should be made by researchers to report every coefficient obtained in any research endeavour.

Government should embark on the posting of qualified and competent test and measurement experts to each school to further ensure the success of administration of test, such experts will be able to advice on the moderate test or examination as well as enlighten teachers on measurement and evaluation procedures in the school system.

The experts should also enlighten the teachers the importance of these methods, that is the reliability methods, and the importance of instrument used by the researchers.

REFERENCES

Ajayi, B. K. (2007). The Effect of Four Scoring Methods on Multiple Choice Agricultural Science Test Scores, Ekiti State University, Ado-Ekiti. Unpublished Ph.D thesis

Alonge, M. F. (1989). Measurement and Evaluation in Education and Psychology.Adebayo Printing (Nigeria) Ltd, Ado-Ekiti.76-78.

Clinton, I. C. (1974). Measurement for Educational Evaluation Printed in the United States of America.

Kolawole, E. B. (2011). Principles of tests Construction and Administration.Bolabay Publications, Lagos State. Nigeria .38-80

Gilbert, Sax. (1974).Principles of Educational and Psychological Measurement and Evaluation.Printed in the United States of America. 259 – 265.

Gronlund, N. E. (1985).Essentials of Psychological Testing: New York London Harper and Row.

Gronlund, N. E .(1981)., Gilbert, S. (1974)., Kenneth& Stanley.(1972). Educational and Psychological Measurement and Evaluation.127-131.

Oladunni, M. O. (1996). Test Measurement and Evaluation in Education. DemiladeOmotayo Publishers, Ibadan.26-31.

Okpalaet all; (1993).Measurement and Evaluation in Education Published by StirlingHordeu Publishers Nig. Ltd. 43-47.

Jimoh, B. O.,&Omoregie O. E. (2012).Factors that Predispose Secondary School Teachers to Examination Malpractice in Edo State Nigeria. Review of European Studies Vol.4 No 1 March, 2012.245-254.

Q1: What are the available reliability procedures in research?

Table 1: Mean and Standard Deviation

			School 1	School 2	School 3	School 4
TRT	T_1	$\overline{X}$	27.66	23.36	23.17	24.93
		SD	3.99	4.24	4.22	6.92
	T_2	$\overline{X}$	30.54	32.39	30.94	26.99
		SD	3.06	3.02	3.66	4.59
PF or EQUV.	P_1	$\overline{X}$	30.79	31.08	30.94	30.10
		SD	2.56	3.61	3.26	4.64
	P_2	$\overline{X}$	27.81	24.47	24.62	26.90
		SD	3.78	5.07	3.90	3.81

TRT = Test re-test

PF = Parallel or Equivalent Form

Q2: Which of these reliability procedures is more appropriately for agricultural objectives test items? This question was answered using the scores obtained from the test administered on 400 students used in this study.

Table 2: Reliability Coefficients of the two methods:

SCHOOLS	TRT	PF or EQUV.
School 1	0.1105	0.2034
School 2	0.5076	0.1508
School 3	0.2880	0.2640
School 4	0.5444	0.5850

Q3: what is the pattern of significances of the agricultural science reliability of coefficients?

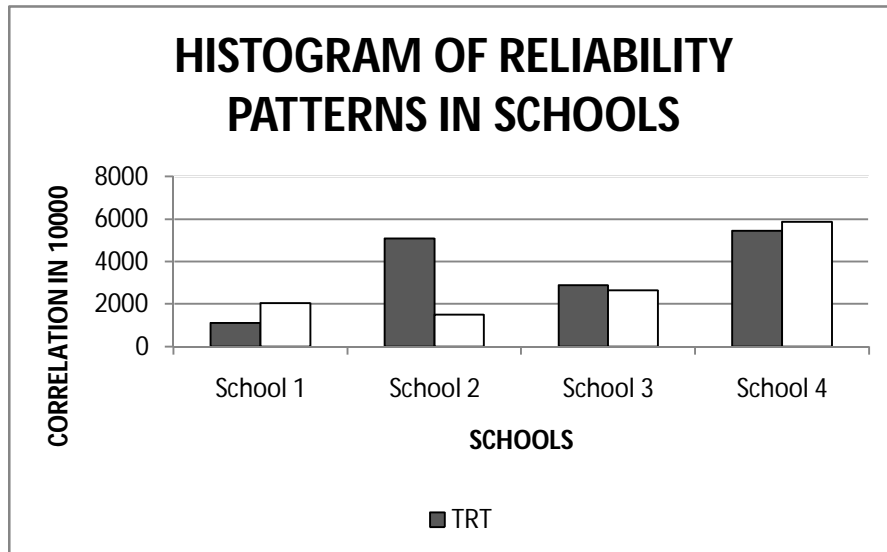**Table 3: PATTERNS OF RELIABILITY COEFFICIENT AS HISTOGRAMS**

HISTOGRAM OF RELIABILITY PATTERNS IN SCHOOLS

(Chart: CORRELATION IN 10000 on y-axis, values 0 to 8000; SCHOOLS on x-axis: School 1, School 2, School 3, School 4; legend: TRT)

Table 4; Ho$_1$: ShowsCorrelation Coefficient for Test re-test Reliability

School	r-calculated	r-table	Significance
School 1	0.1105	0.195	Not significant
School 2	0.5076	0.195	Significant
School 3	0.2880	0.195	Significant
School 4	0.5444	0.195	Significant

r-table = 0.195

* $P > 0.05$

Table 5; Ho$_2$: ShowsCorrelation Coefficient for Equivalent or Parallel Form Reliability

School	r-calculated	r-table	Significance
School 1	0.2034	0.195	Significant
School 2	0.1508	0.195	Not significant
School 3	0.2640	0.195	Significant
School 4	0.5850	0.195	Significant

r-table = 0.195

* $P > 0.05$