

A Comparative Analysis of Reliability Methods

B. K. Ajayi

Institute of Education, kiti State University, Ado-Ekiti, Ekiti- State. Nigeria

Abstract

The study is concerned with the inappropriate usage of the reliability methods. It also identified the most commonly used of the reliability methods and the most appropriate methods for agricultural science objective test items. Reliability of a test is the extent that repeated measurements give consistent results for individual. A test is said to be reliable if it measures what it is supposed to measure consistently. There are many methods of measuring reliability which are: Test-retest, parallel form, Split-half, other methods of measuring internal consistency include Kuder Richardson KR20 and Kuder Richardson KR21. The study made a comparative analysis of reliability in four selected secondary schools in Ado-Ekiti Local Government Area of Ekiti state. Forty (40) items in agricultural science objectives test were constructed by the researcher. The objectives test was administered to the students in the four selected schools. The scores were compared across the sample schools. The descriptive analysis such as mean, standard deviation and comparative analysis of reliability methods were used. The hypothesis was held significant at 0.05 probability level. The result of the analysis shows that there was no significant differences between test-retest, parallel form, split half and Kuder-Richardson methods in agricultural science objectives test. The findings revealed that significant different exist in reliability methods employed in the sampled schools as a result of the findings, the instrument is recommended to be used in analyzing reliability method.

Keywords: Comparative, Reliability, Split-half, Test-retest, parallel form and Kuder -Richardson

1. Introduction

The use of research instruments appear to be inevitable in empirical researches. The indiscriminate way the researchers used instruments in form of questionnaire and the like without its reliability coefficient leaves much to be desired. The researcher of this paper is concerned with the inappropriate usage of the reliability methods. Ogunniyi (1990), defined reliability as the suitability or accuracy of the data to be collected from a given test even when test repeated two or more times under similar situations. Ajayi (2007) went further that the test is reliable when the test produces identical results.

Kolawole (2001), defined reliability as consistency, that is, whether the test produces the same results on repeated administrations. He went further that a test is considered reliable if the testee performs about the same, when re-examined with the same test at a different time (without additional instruction). Ogunyemi (2003) in his own view, defined reliability of a test as the degree of precision or consistency of measurement obtained for using instrument which means the measuring instrument should be consistent by giving the same result or identical results, he went further that an instrument is said to be reliable if it consistently yield close or identical results when used over and over again. Alonge (1989) describes reliability in terms of the degree of confidence that can be placed on measurement. That is, reliability is a matter of degree. Therefore, Reliability refers to how well two measurements of a test agreed with each other. Kolawole (2011) went further that there are three methods of estimating reliability which are; test-retest (measure of stability), parallel form (measure of equivalence), split-half (measure of internal consistency) Kuder Richardson 20 and 21. Ajayi (2013) mentioned that test re-test reliability of the instrument was established by re-administering the test on a group within an interval of two weeks. Test re-test reliability can also be called measure of stability this is the repeatability of test overtime to get the same result with the same person. Test re-test is the degree in which test scores are consistent from one test administration to the next to estimate reliability by means of the test re-test method, it is an index of stability. A test is being administered to a fairly large number of testees the scores were collated after some weeks, let's say two weeks interval the test was repeated to the same testees with the same questions without adding any instruction to the same set of testees This is known as test re-test reliability. The time lag may be some days, weeks or months according to Oladunni (1996), Alonge (1989) opined that if the time lag between the two administrations of the test is short, the reliability coefficient could be inflated because some of the examinees would still remember the answer or might have found an easier technique of answering the items. The scores on the two administrations are not independently obtained and the correlation between them will be spuriously high. Also, if the students were able to maintain their relative positions in both administrations, the results are highly stable, if otherwise the results have low stability. There are two major ways of conducting test re-test reliability measurements according to Ajayi (2007) the first major way is by immediate re-test while the other one is by testing after some interval. The parallel is also known as alternative or equivalence form method, Oladunni (1996) defined parallel form reliability as two parallel forms of a test that are administered to the same group of testees with a given time interval, usually not long. Two tests are parallel or equivalent, if the sample

tests the same subject matter. That is tests of the same difficulty level and discriminating power to group of testee. In Split half method of estimating reliability, there is general agreement that split half method is split into odd and even numbered questions and scores. According to Kolawole (2011) went further to explain that the sum of scores of odd numbered questions are 1,3,5,7 and of individual testees were correlated with their corresponding scores in even numbered questions which are 2,4,6,8 and 10. The split half method is done by administering a test on the sample and split the score of an individual testee. There are many factors that influencing reliability of a test according to Kolawole (2011) length of test, variability of the group, ability level of testees, objectivity of scoring, scoring technique, method of construction, testing conditions, instability of scores, error in the individual, ambiguity, giving option, range of measuring instrument, difficulty of the test and testing condition information showed that, test re-test method is commonly used more than the other methods such as split-half, equivalent or parallel form and kuder-Richardson. Ajayi (1998), opined that the commonly used reliability procedure in education researches are; in Bachelor degrees in Education, out of twenty (20) past projects examined between 1990 to 1998, sixteen (16) employed the test re-test method. Test re-test (measure of stability), the test is given and reported on the same group, and then co-efficient of correlation is computed between the first and second set of scores. Kolawole (2011) mentioned that the key concept for test re-test procedure is that of stability. It answers the question concerning how stable or dependable are these measurements over a period of time. Higher reliability indicates that individuals remain rather uniform, or maintain their rank positions in spite of changes in whatever psychological functions this test measures. A low correlation coefficient of reliability indicates test functions measure is not stable. No one made used of parallel form, three (3) used split-half, only one researcher used the Kuder -Richardson method.

Twenty (20) projects were examined under Masters in Education project, seventeen (17) test re-test method, none was used under parallel form, three (3) used split-half and none under Kuder- Richardson. Under Journal, fifteen (15) projects were examined, eight (8) used test-retest method, none under parallel form, two under split-half while five (5) used Kuder-Richardson.

He concluded that, most researchers used test-retest reliability method. Gronlund (1985), the two reliability forms of the test are administered to the same group of pupils in close succession and the resulting test score are correlated.

Gulhiken (1936), speaks of parallel tests which he defines statistically, parallel tests have equal means, equal variances and equal inter correlation with one another. If two or more parallel forms are combined to make a longer form, spearman brown would be applied to estimate the reliability of the longer form.

For a split-half, a score for each half can be obtained from a single test administered; one might put the odd numbered items and the even numbered items into half. The scores of odd numbered questions are correlated with their corresponding scores in even numbered questions.

Oladunni (1996), refers to the result from the two forms of the test that are correlated in which the correlation co-efficient gives a measures the same sample of the degree to which both forms measure of equivalence is the degree to which both forms measure the same samples of behaviour. Kolawole (2001), mentioned that if the purposes of giving a test is not for long range prediction but for the purpose of drawing inference about the knowledge, one has in a subject matter, then emphasis should be on coefficient of stability.

Kuder-Richardson method is a way to avoid splitting a test into halves as used in the split-half method. Kuder and Richardson (1937), in their paper the theory of the estimating the test reliability' published in psychometrical (1937), proposed a method for estimating test reliability in the method scores of a single test are necessary to determine the test reliability. Alonge (1989), explained that Kuder and Richardson KR20 and KR21 are fast and easy method of estimating the reliability coefficient. In KR20 and KR21 proposed a method for estimating test reliability in the method scores of a single test are necessary to determine the test reliability.

2. Statement of the Problem

The indiscriminate methods of reporting the reliability of instrument leaves much to be desired. It appears that many researchers do not use the appropriate reliability techniques for a given problem of investigation. In this study, the following questions were raised.

- a. What are the reliability procedures commonly used by educational researchers?
- b. Which of these reliability procedures is more appropriately for agricultural objective test items?

3. METHODOLOGY

Descriptive research was used that employed the correlation design using stratified random sampling techniques; different secondary schools were selected randomly. A total of 400 students consisted male and female from Ado Local Government Area of Ekiti State were selected. The sampling techniques incorporated across the schools, the nature of the schools whether old or new, male or female and the grade of the school. The instruments used are forty items objectives test in Agricultural science. One of the instruments was an alternate form of the other tests from the purpose of the parallel form reliability method.

The correlation design allows the researchers to investigate relationships that exist among the reliability methods and across the scores obtained in schools.

4. Research Hypotheses

Based on the statement of the problem, the following research hypotheses were raised:

H₀₁: The test re-test, parallel form, split-half and Kuder-Richardson reliability co-efficient are not significant for the agricultural achievement test in every school.

H₀₂: The split-half, Kuder –Richardson, test re-test and parallel form reliabilities were not significant for the patterns of co-efficiency in line graph of agricultural achievement test in all schools.

4.1 H₀₁: Reliability Co-efficient for Test-Retest, Parallel Form, Split-Half and KUDER -RICHARDSON Methods in Four Schools.

	T.R	P.F	S.H	K.R
School 1	0.1105	0.2034	-0.3925	0.465
School 2	0.0576	0.1508	-0.0567	0.4611
School 3	0.2880	0.2640	-0.389	0.4523
School 4	0.5444	0.5850	0.1516	0.8041

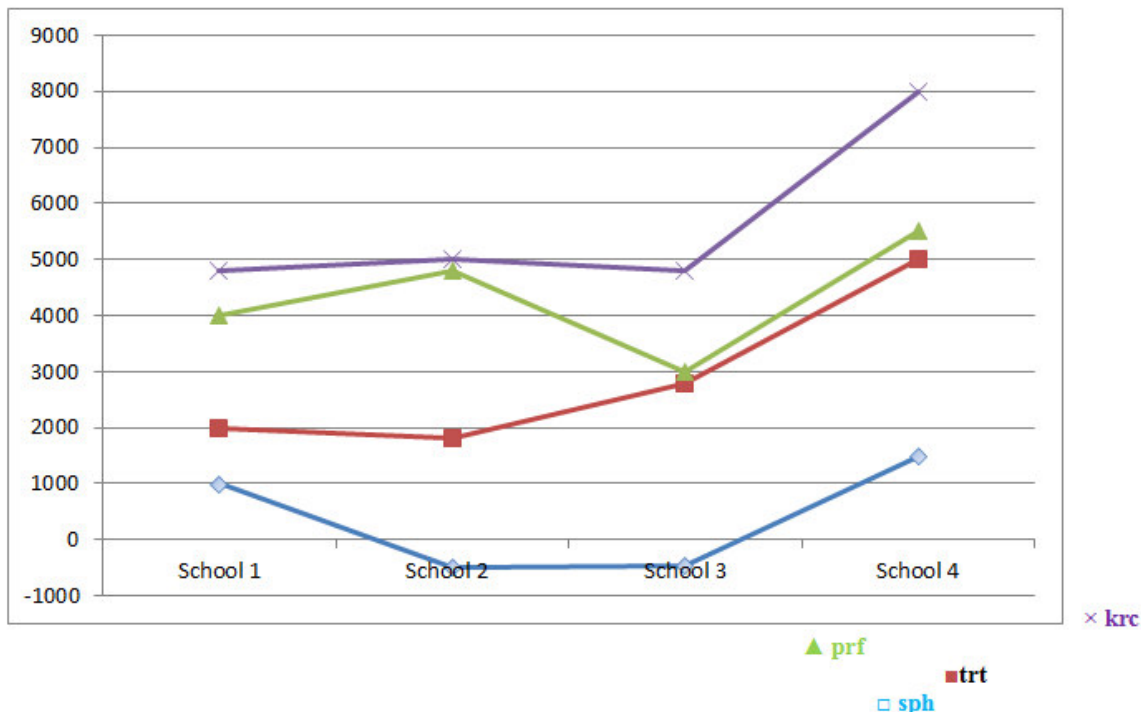
In the table above, we have four (4) schools, in school one (1) Test-retest has 0.1105 and 0.2034 in parallel form while Split-half was -0.3925. The reliability co-efficient has the highest score under Kuder -Richardson with 0.465.

In school two (2), the scores for Test-retest was 0.0576 and Parallel form of 0.1508 followed by Split- half of -0.0567, and Kuder -Richardson of 0.4611 which is a bit higher.

In school three (3), Test-retest of 0.2880 and Parallel form of 0.2640 with Split-half of -0.389 which show that the most appropriate reliability co-efficient for agricultural science test items was Kuder-Richardson with the score of 0.4523

Then, in school four (4), the Test-retest has 0.5444 with Parallel form of 0.5850 while the Split-half has 0.1516. Therefore, Kuder-Richardson had the co-efficient of 0.8041 as the highest scores.

4.2 H₀₂: Line Graph of reliability patterns in four schools



LINE GRAPH

LINE GRAPH OF RELIABILITY PATTERNS IN SCHOOLS

In the line graph showing the reliability patterns in schools the correlation was indicated in 10,000

Longitudinally the interval was 1,000 between the numbers. Looking at the graph, it is evident that Kuder-Richardson moves in a straight line across the three schools before moving sharp up to give the highest correlation coefficient recorded. In case of split-half, the line started from the first school with a relative low

correlation coefficient. It moves diagonally downward up to the negative vertical school. In parallel form method, it is low in school 1 and increase in school 2 while it drops in school 3 and rise sharply along with the graph line of test re-test. The last reliability pattern which was test re-test ,the graph moves in zigzag form with correlation dropping sharply at an instant and increasing at other time. It could be observed that

Split-half had low and negative correlation coefficients and Kuder-Richardson has the highest correlation.

5. RESEARCH FINDINGS.

The study revealed that the educational researchers used instruments in appropriate ways. It revealed that significant difference exist in reliability methods in the sampled schools. It was discovered that of all the methods, Kuder -Richardson was the most appropriate method for objective test, because in all the four schools in which the tests were administered, Kuder-Richardson has the highest scores, followed by test-retest, parallel form and the split-half was the least.

6. CONCLUSION

Based on the findings of this study, it revealed that Kuder-Richardson was the most appropriate method for objective test because in all the four schools in which the test was administered, kuder-Richardson has the highest scores, followed by test-retest, parallel form and the split-half was the least. Also most educational researchers used test-retest reliability method for their research.

The correlation in longitudinal was in 10,000 and schools were on the horizontal. When looking at the graph, test-retest reliability was low but in parallel form it was increased to higher value and in case of split-half, the figure increased but could not reach the high co-efficient in Kuder-Richardson.

In school (two(2), under test-retest, the figure was high but on getting to parallel form the number dropped in split-half, it decreased to negative values and it also increased to the highest value in Kuder-Richardson, method has the highest correlation scores in all the schools and split-half methods has the least correlation scores with negative value.

7. RECOMMENDATIONS

The following recommendations are made, based on the findings of the study. The educational researchers should try and use the appropriate reliability method in carrying out their research.

Tests 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. Efforts should be made by researchers to report every coefficient obtained in any research. The experts should also enlighten the teachers the importance of administering a reliable test in the school system as well as the importance of instrument used by the researchers.

REFERENCES

- Ajayi, B.K. (1998),A Comparative Analysis of Reliability Methods M.ed Thesis, 1-4,Ondo-State University, Ado-Ekiti. Now Ekiti State University Ado Ekiti
- Ajayi, B.K. (2007),The Effect of Four Scoring Methods on Multiple Choice Agricultural Science Test Scores.Unpublished Ph.D thesis. University of Ado-Ekiti. Now Ekiti State University Ado- Ekiti
- Ajayi, B.K. (2013),A Comparative Analysis of Test Re-Test and Equivalent Reliability Methods, International Journal of Education and Research vol.1 No. 6 June, 2013
- Alonge, M.F (1989),Measurement and Evaluation and Psychology, Adebayo Printing Nigeria Limited. 71,
- Gronlund, N.E. (1985),Measurement and Evaluation in Testing (5th Edition) New York; Macmillian Publishing Company, 91-98.
- Gulliksen, H. (1936), The content reliability of a test, Psychometrika , 1, 189-194,
- Kolawole, E.B. (2001), Test and Measurement, Yemi Prints and Publishing Services, Ado-Ekiti. 19,
- Kolawole, E.B. (2011), Principles of Test Construction and Administration, Lagos, Bolabay Publications Nig. Ltd.
- Kuder,G.F., Richardson M. (1937);The Theory of the Estimation of Test Reliability,Psychometrika 2, 151-160
- Ogunjemilua, M.F. A Comparative Analysis of Psychometric Properties of Two Senior School Certificate Mathematics Examinations Council M,ED. Unpublished.
- Ogunniyi, M.B.(1990), Educational Measurement and Evolution Edition. Longman Publishers. Lagos, Nigeria.
- Ogunyemi,M.F. (2003), A Comparative Analysis of Psychometric Properties of Two Senior Schools Certificate Mathematics Examinations Council, M.Ed thesis. Unpublished.
- Oladunni, M.O.(1996),Test and measurement and Evaluation in Education, Published by Demilade Omotayo Publishers. Ibadan, 26-31.