

EXTENTS OF RELIABILITY OF EACH OF THREE FORMS OF THE EQUIVALENT OF OBJECTIVE TESTS

JOHNSON SUNDAY, OSIRE (Ph.D.),

*University Of Ilesa,
johnsonsunday446@gmail.com*

VICTOR ABIODUN, LIJOFI,

*University Of Ilesa,
lijofivictor@gmai.com*

&

JOSHUA WORLAY, OYETUNJI,

*University Of Ilesa
oyetunjijoshuall@gmail.com*

Abstract

The study investigated the extent of reliability of three form of equivalence. A correlational research design was used. A Stratified Random Sampling was used to obtain 100 college students. A 24 items multiple choice objective test was constructed. Using Crhonbach Co-efficient Alpha, an index of 65 obtained showed that the instrument had both validity and reliability. Versions of the instrument were constructed to obtain other three multiple choice objective tests. Using simple Correlation, the Pearson Product Moment Correlation showed equivalence of the four instruments. The three null hypotheses of non-equivalence at 0.5 level of significance, $r(80) = 0.85, 0.81, 0.82$ and 0.84 ; $P < 0.05$. The study concluded that scores on parallel tests are always equivalent. The recommendation was made that the equivalent instruments could be used in the same examination to forestall examination misbehavior among other recommendation.

Key Words: Extent of Reliability, Equivalence, Objective Test.

Introduction

The construction and administration of parallels objective test and ensuring that the parallels are reliable are the focus of this study. Reliability is defined by Cohen., L., Manion, L., Morrison, K. (2000), as the synonym for consistency over time of instrument and groups of instruments. As Cohen further stated, Reliability is the degree of consistency at which a test measures the traits it is designed to measure. Cohen et al (2018) Reliability is essentially an umbrella term for dependability, consistency and replicability over time, instrument and group of respondents. Nwogu (2006) opined that Reliability

of a test is the degree to which a test measures accurately and consistently yielding comparable results when administered several times. It was argued by Nwogu (2023) that reliability of an instrument refers to the degree of accuracy with which the instrument measures whatever it is measured.

The concerns of researchers working on the ongoing study about Reliability trials are whether scores obtained from different forms of objective tests through repeated measures are consistent. The reliability coefficient of a test can be estimated using an equivalent form method, in this method, the parallel forms of a test is administered on two groups of the same population under the same conditions and results are correlated. Casmir N. Ebuoh (2015) noted that tests have equivalent forms when they contain similar or different items measuring the same objectives but giving almost similar results. In this case, it could be declared that the scores of a testee will be equivalent on two tests.

A test (informally, exam) is an assessment intended to measure a test taker's knowledge, skills and aptitude. According to Nworgu (2003), is a structured situation comprising a set of questions to which an individual is expected to respond, and each question in the test has a preferred answer. According to Nofisat (2018), The significant purpose of testing in education is that it helps in determining the learning difficulties or weakness, strength and level of mastery of examinees in a given task.

As defined by Faleye (2012), objective forms of a test as tests requiring short answers which could be in the form of a letter grade (such as, A, or B, or C, or D, or E), a word or phrase. These include the Multiple-choice, binary-choice, matching items, short answer, Completion items and close tests. An advantage of using objective test items is its simplicity in scoring as it does not require expertise in scoring. Finding equivalence of objective tests may entail the creation of a parallel form of test, this can be done by administering a test created from the same contents and determining the extent of equivalence of the test. However, Oluwawole, Folorunso, Osire, Olatuni, Aayi, and Oni (2020) reported that; A necessary conditions to solve the problems associated with objective test equivalence is to ensure that the items in each test is similar to another in terms of difficulty, discrimination, validity and reliability.

Despite the efforts of test experts and users in the construction of reliable test items, much has not been done in the areas of discovering the methods that can be used in creating parallel forms of test and

ensuring equivalence regarding reliability. Faleye (2012), noted that it is usually challenging to obtain the same scores from two or more administrations of the same equivalence test.

Statement of the Problem:

The researchers have observed that scores obtained from the repeated administration of equivalence of objective test has not been the same. This may be due to the extent to which different forms of parallel tests show levels of equivalence, hence the need to examine the extent of reliability of three forms of equivalence objective test.

Purpose of the Study:

The purpose of this study is to determine the efficient techniques for the comparison of achievement test measure. The researchers will also determine the extent of equivalence and reliability of scores obtained from the administration of three parallel forms of the objective test designed by the researchers from the same content domain. Furthermore, the study provides information about the forms of the test that yield the highest reliability.

Research Questions:

1. Will there be any relationship between the original test (TYPE A) and its equivalence got through the method of reshuffling of options (TYPE B)?
2. Will there be any relationship between the original test (TYPE A) and its equivalence got through the method of re-framing of stems (TYPE C)?
3. Will there be any relationship between the original test (TYPE A) and its equivalence got through the method of re-shuffle of options and stems (TYPE D)?

Research Hypotheses:

- Ho₁: There will be no significant relationship between the original test (TYPE A) and its equivalence got through the method of reshuffling of options (TYPE B)
- Ho₂: There will be no significant relationship between the original test (TYPE A) and its equivalence got through the method of reshuffle of options (TYPE C)
- Ho₃: There will be no significant relationship between the original test (TYPE A) and its equivalence got through the method of re-shuffle of options and stems.

Methodology

The researchers adopt the Quasi-experimental design type because it allows analysis to be performed where true experiment cannot be used. The four parallel forms of Educational Research Objective Test, (EROT Type; A, B, C and D) is administered on the same sets of students at the intervals of two weeks each and scores were obtained. The test was constructed from the course content of Educational Research using appropriate test blue-print, the students have been taught. Hence, in this study the researchers compared the scores obtained from the four forms of the parallel test. The population for the study comprised 712 of the 2020/2021, 200 level NCE students in four schools; Sciences, Art & Social sciences, Vocational & Technical Education and Early childhood Education who were taught Educational Research in Osun State College of Education, Ilesa, Nigeria. Students numbering 100 were selected across 4 schools using stratified random sampling techniques. A multiple-choice type of objective test titled, Educational Research Objective Test (EROT), which was developed by the researchers was used in the study. The test has 24 items that focus on the Educational Research area of Educational foundations for 200 level students of the College of Education. Three other forms of the test were re-created to make four equivalent forms; (EROT) type A, B, C and D. The type A was the original test, type B was created through re-framing the test stem, type C was created through the reshuffle of options and type D was created through the reshuffle options and stems. Experts of test and measurement edited the instrument. The face and content validity of the instrument was also verified by the expert, with the use of a test blueprint. The concurrent Validity of the instrument was obtained using Pearson's Product Moment Correlation to compare scores from concurrent administration of two forms of the test. This gave a coefficient of 0.69. This coefficient was significant at an Alfa less than 0.05. The Reliability of the instrument was established through Test-retest types of Reliability. This is done by the administration of the instrument twice to the same sample. Cronbach Alfa was used to obtain the correlation coefficient generated from two administrations of the test which was 0.65. This was found at an Alfa less than 0.05. The researcher administered the four parallel forms of tests, Educational Research Objective Test (EROT), Type A, B, C and D to the 40 students at an interval of two weeks each. Four sets of scores were generated for each student's responses obtained from the administered test. All scores were coded into Statistical Package for Social Science (SPSS), statistical software. The coded data were subjected to descriptive and inferential analysis.

Results and Discussion

Research Question 1: Will there be any relationship between the original test (TYPE A) and its equivalent got through the method of reshuffling of options (TYPE B)?

To answer this question, the 24 Items of the original test Type A and its equivalent got through reshuffling of stems Type B were scored 1 mark for each correct response. Thereafter, the responses were subjected to descriptive correlation analysis with the use SPSS statistical software as shown in Table 1.

Table 1: Coefficient of correlation studying extent of relationships between original test TYPE A and its equivalent got through method of reshuffling of options TYPE B

Sources of variation	N	R
Original Type A	40	0.85
Reshuffled Options (TYPE B)	40	
Total	80	

Table 1 represents the relationship between original test (type A) and its equivalent got through reshuffling of options (type B). On the table, the Pearson Product Moment Correlation Coefficient 0.85 indicates a strong and positive relationship. This implies that as the scores obtain from the original test

Type A increases, there will be increase in the students' scores obtained from the equivalent got through reshuffling of options.

Descriptive Statistics Skewness Statistic

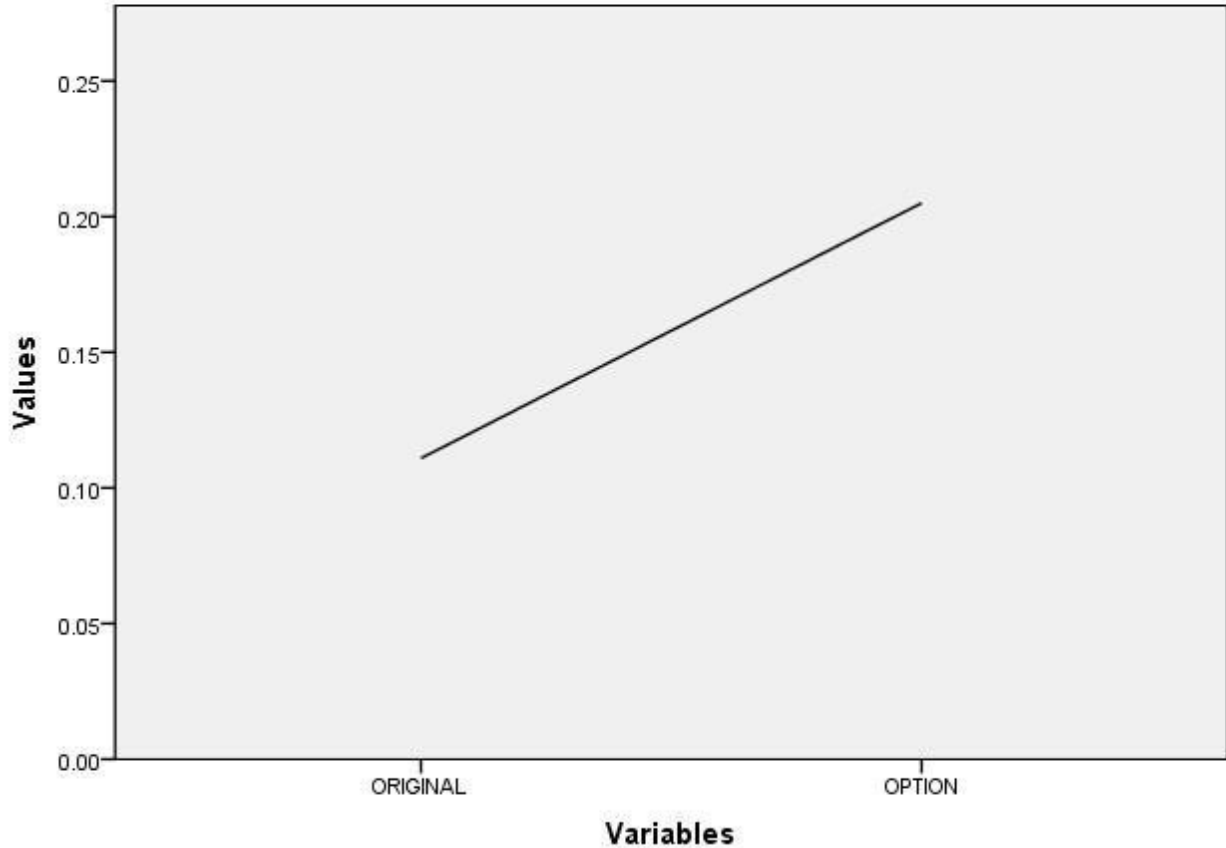


Fig. 1: Graph showing relationship between the original test and its equivalent got through reshuffling of options.

The graph in figure 1 shows a positive relationship between the original test and its equivalent got through reshuffling of options because it is skewed to the right. The graph also shows a strong kurtosis. These indications support the correlation Coefficient of 0.85 which is an index of the relationship between the original test and its equivalent got through reshuffling of options. This also means that the graph in Fig. 1 shows that the relationship is strong.

Research Question 2: Will there be any relationship between the original test (Type A) and its equivalent got through the method of re-framing of stems (Type C)?

To answer this question, the 24 Items of the original test (Type A) and its equivalent got through reframing of stems (Type C) were scored 1 mark for each correct response. Thereafter, the responses were subjected to descriptive correlation analysis with the use SPSS statistical software as shown in Table 2.

Table 2: Coefficient of correlation studying extent of relationships between original test TYPE A and its equivalent got through method re-framing of stems (Type C).

Sources of variation	N	R
Original Type A	40	0.81
Reframing of stems (Type C)	40	
Total	80	

Table 2 represents the relationship between original test (type A) and its equivalent got through Reframing of stems (Type C). On the table, the Pearson's Product Moment Correlation Coefficient of 0.81 indicates a strong and positive relationship. This implies that as the scores obtained from the original test Type A increases, there is increase in the students' scores obtained from the equivalent got through Reframing of stems (Type C).

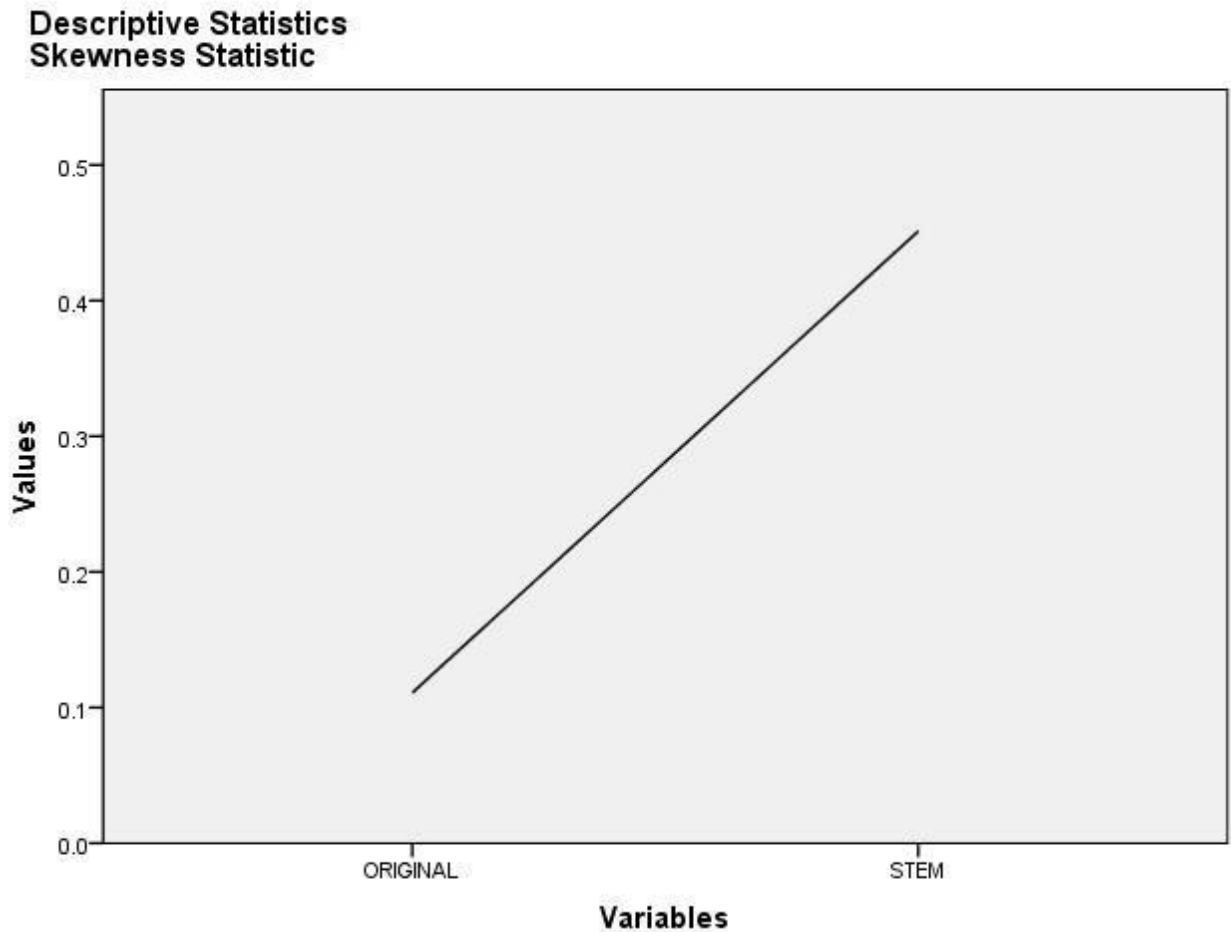


Fig. 2: Graph showing relationship between the original test and its equivalent got through reshuffling of options and stems

The graph in figure 2 shows a positive relationship between the original test and its equivalent got through reshuffling of options and stems because it is skewed to the right. The graph also shows a strong kurtosis. These indications support the correlation Coefficient of 0.84 which is an index of the relationship between the original test and its equivalent got through reshuffling of options and stems. This also means that the graph in Fig. 2 shows that the relationship is strong.

Research Question 3: Will there be any relationship between the original test (TYPE A) and its equivalent got through the method of re-shuffle of options and stems (TYPE D)?

To answer this question, the 24 Items of the original test (type A) and its equivalent got through reshuffling of options & stems (Type D) were scored 1 mark for each correct response. Thereafter, the responses were subjected to descriptive correlation analysis with the use SPSS statistical software as shown in Table 3.

Table 3: Coefficient of correlation studying extent of relationships between original test TYPE A and its equivalent got through method reshuffling of options & stems TYPE D.

Sources of variation	N	R
Original Type A	40	0.84
reshuffling of options & stems TYPE D	40	
Total	80	

Table 2 represents the relationship between original test (type A) and its equivalent got through reshuffling of options & stems TYPE D. On the table, the Pearson Product Moment Correlation Coefficient 0.84 indicates a strong and positive relationship. This implies that as the scores obtain from the original test Type A increases, there will be increase in the students' scores obtained from the equivalent got through reshuffling of options & stems TYPE D. and vice versa.

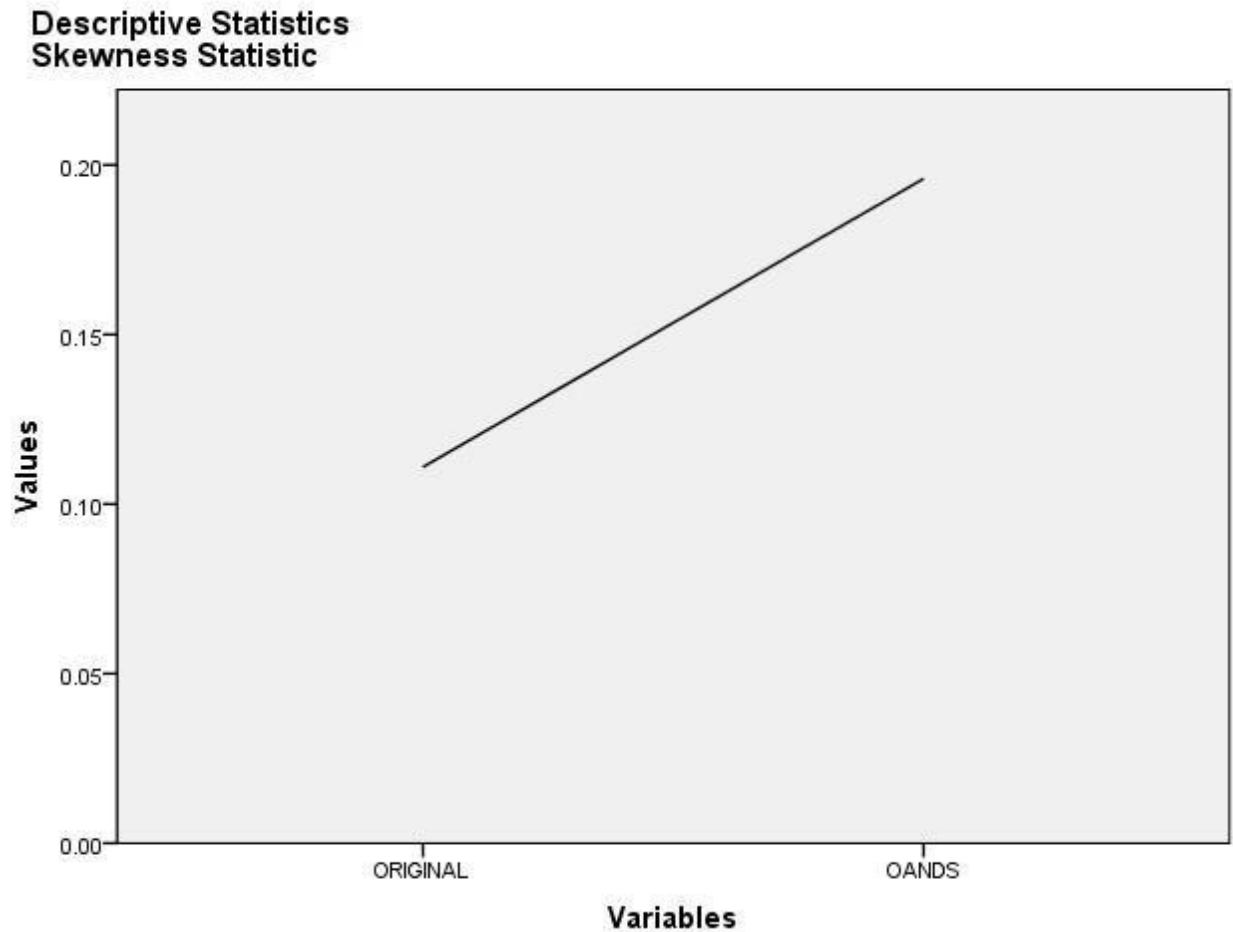


Fig. 3: Graph showing relationship between the original test and its equivalent got through reshuffling of options and stems

The graph in figure 3 shows a positive relationship between the original test and its equivalent got through reshuffling of options and stems because it is skewed to the right. The graph also shows a strong kurtosis. These indications support the correlation Coefficient of 0.84 which is an index of the relationship between the original test and its equivalent got through reshuffling of options and stems. This also means that the graph in Fig. 3 shows that the relationship is strong.

Hypotheses Testing:

Hypothesis 1: There will be no significant relationship between the original test (TYPE A) and its equivalence got through the method of reshuffling of options (TYPE B).

To test this hypothesis, responses to the original test (TYPE A) and the equivalent got through reshuffling of options (TYPE B) were subjected to Pearson Correlation analysis and the result is presented in Table 4.

Table 4: Results of correlation analysis studying extent of relationships between original test TYPE A and its equivalent got through method of reshuffling of options TYPE B

Source of variation	N	r	Results
ORIGINAL (TYPE A)	40	0.87	The relationship between the original test and its equivalent got through reshuffling of options is high and positive
RESHUFFLED OPTIONS (TYPE B)	40		

*. Correlation is Significant at the 0.5 (2-tailed).

Table 4 shows that the relationship between original test (type A) and its equivalent got through method of reshuffling of options TYPE B was significant ($r = 0.86$; $p < 0.05$). The table further shows that the magnitude of the relationship (0.86) was high and positive. This implies that as the scores obtain from the original test (Type A) increases, there exists an increase in the students' scores obtained from the reshuffling of option (Type B).

Hypothesis 2: There will be no significance relationship between the original test (type A) and its equivalent got through the method of reframing of stems (Type C)?

To test this hypothesis, responses to the original test (type A) and the equivalent got through reframing of stems (type C) were subjected to Pearson Correlation analysis and the results is presented in Table

5.

Table 5: Results of correlation analysis studying extent of relationships between original test TYPE**A and its equivalent got through method of reframing of stems TYPE C**

Source of variation	N	r	Results
ORIGINAL (TYPE A)	40	0.82	The relationship between the original test and its equivalent got through reframing of stems is high and positive.
REFRAMING OF STEMS (TYPE C)	40		

*. Correlation is Significant at the 0.5 (2-tailed).

Table 5 represents the relationship between original test (Type A) and the reframing of stems (Type C) was significant ($r = 0.82$; $p < 0.05$). The table furthers shows that the magnitude of the relationship (0.82) was high and positive. This implies that as the scores obtain from the original test (Type A) increases, there exist increase in the students' scores obtained from the reshuffling of option test (Type B)

Hypothesis 3: There will be no significance relationship between the original test (TYPE A) and its equivalent got through the method of reshuffle of options& stems (TYPE D)

To test this hypothesis, responses to the original test (Type A) and the equivalent got through reshuffle of options& stems (Type D) were subjected to Pearson Correlation analysis and the results is presented in Table 6.

Source of variation	N	r	Results
ORIGINAL (TYPE A)	40	0.84	The relationship between the original test and its equivalent got through reshuffling of options is high and positive.
RESHUFFLED OPTIONS & STEMS (TYPE D)	40		

*. Correlation is Significant at the 0.5 (2-tailed).

Table 6 represents the relationship between original test (Type A) and the reshuffle of options & stems (Type D) was significant ($r = 0.84$; $p < 0.05$). The table furthers shows that the magnitude of the relationship (0.84) was high and positive. This implies that as the scores obtain from the original test (type A) increases, there exists an increase in the students' scores obtained from the reshuffle of options/ stems (Type D).

Discussion of Findings:

The Descriptive Statistics results, which answers the research questions, are consistent with the results obtained from tests of hypotheses of the study. Also, the results of hypothesis tests confirm and generalize the answers to the research questions over the formation of equivalence of objective tests.

In line with this, the test of Ho1 gives the results that there is a significant relationship between the original test (type A) and its (type B) equivalent got through the method of reshuffling of options. This means that the original form of the educational research objective test (type A) is equivalent to the reshuffling of options (type B) just as the answer to research question 1 says.

The result of the Ho2 which confirms and generalizes the answer to research question 2 is that there is a significant relationship between the original test (Type A) and its equivalent got through the method of reframing of stems. This means that the original form of educational research objective test (type A) is equivalent to the reframing of stems (type C) just as the answer to question 2 says.

The result of Ho3 which confirms and generalizes the answer to research question 3 is that there is a significant relationship between the original test (type A) and its (type D) equivalent obtained through the method of re-shuffle of options & stems. This means that the original form of educational research objective test (type A) is equivalent to the reshuffle of stems & options (type D) just as the answer to question 3 says.

Conclusion:

The conclusion is based on generalize-able nature of the findings of this study. Therefore, the original form of educational research objective test is equivalent to its forms obtained through reframing of stems, reshuffle of options and reshuffle of stems /options. Hence, this makes the scores obtained from one form of the test comparable to the others.

Recommendations:

The following are recommendation in supportive of the findings which is proffered as follows:

1. Topics on reliability measures should be included in the course content of teachers in tertiary institutions.
2. Scores on parallel tests which their equivalence have been confirmed to be reliable could be used to substitute a student's performance where the students answer script is omitted.
3. Parallel method of finding equivalence of objective tests should be used in standardizing tests of college of education students.

4. The different forms of equivalence of the test can be used together in examinations to forestall malpractices

References:

- Casmir N. E. (2015). Measures of equivalence as a predictor of student's reliability in biology objective test. *International Journal of Scientific and Engineering Research*. (6) 24.
- Cohen, L., Manion L. & Morrison, K. (2018). *Research methods in education*. (8th ed). London, Britain: Routledge.
- Cohen, L., Manion, L. & Morrison, K. (2000). *Research methods in education*. (5th ed). London, Britain. Routledge.
- Faleye, B.A. (2012). *Educational test and measurement* (1st ed). Ile-Ife, Nigeria. Obafemi Awolowo University Press.
- Nofisat, A.O. (2018). *Linear and separate calibration method of equating continuous assessment of secondary school students*. Dissertation of Educational Foundations and Counselling. Ile-Ife, Nigeria. Obafemi Awolowo University.
- Nwogu, B. G. (2006). *Educational research: basic issues and methodology*. Ibadan Nigeria: Wisdom Publishers Limited.