OPEN ACCESS

Journal of Advances in Education and Philosophy
Abbreviated Key Title: J Adv Educ Philos
ISSN 2523-2665 (Print) | ISSN 2523-2223 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: https://saudijournals.com

Original Research Article

Effect of Item Order on the Reliability of Mathematics Test among Secondary School Students in Rivers State

Dr. Opara, I. M[1*], Ogbanu, G. I[2]

[1]Department of Educational Psychology Guidance and Counselling, University of Port Harcourt, Rivers State, Nigeria
[2]Department of Engineering Technology, Homik Oil and Gas Facility Engineering Schools, Rivers State, Nigeria

DOI: 10.36348/jaep.2023.v07i11.003 | **Received:** 17.08.2023 | **Accepted:** 23.09.2023 | **Published:** 15.11.2023

***Corresponding author:** Dr. Opara, I. M
Department of Educational Psychology Guidance and Counselling, University of Port Harcourt, Rivers State, Nigeria

Abstract

The study investigated the effect of item order on the reliability of mathematics test among secondary school students in Rivers State. One research question and one null hypothesis guided the study. Quasi – experimental research design was used and the study was carried out in nine public senior secondary schools in three local government areas in River State which was drawn using Simple Random Sampling Technique. A sample size of 507 students were drawn from the population of 904 through Stratified Random Sampling Technique and the application of Taro Yemen formula to determine the minimum sample size. The instrument used was Multiple Choice Format Mathematics Performance Test (MCFMPT) arranged in ascending, descending and inconsistent order of difficulty and students' scores were used for data collection. The reliability of the instrument was 0.75 using split-half method of establishing reliability and Spearman Brown prophecy formula for internal consistency of the instrument. Mean and Standard deviation were used to answer research question while t-test and Analysis of variance (ANOVA) were used to test the null hypotheses at 0.05 alpha level. The results revealed that item order based of ascending, descending and inconsistent order do not have significant effect on the reliability of mathematics test among secondary school students. Based on the findings, it was recommended that all pattern of item order (ascending, descending and inconsistent order of difficulty) should be encouraged among secondary school students as it has no significant effect on reliability coefficient. Students should be motivated to improve their real abilities in mathematics related courses and not judging their underperformance in mathematics based on the arrangement of items.
Keywords: Item order, Reliability, Mathematics Test.

INTRODUCTION

Reliability is a major consideration in evaluating any instrument, it is the consistency to which an instrument measures what it was designed to measure. If a student's level of achievement is to be measured with test scores, the scores supposed to be similar under repeated testing, two halves of a single test, alternate forms of the same test and individual item measurement of the test (Opara, 2021). Reliability is one of the psychometric properties of any good test. According to Denny and Dylan (2018), the term "psychometric" refers to the measurement of psychological traits like personality, IQ, and academic achievement and for any measuring instrument to be accepted as meeting a need, it must possess the two key psychometric qualities of validity and reliability. However, this study is focused on only reliability of mathematics test when items are arranged in ascending, descending and inconsistent order of difficulty.

Numerous individuals, organizations, and employers of labour, among others, highly value mathematics because of its widespread application in many aspects of human life, including academic disciplines like introductory technology, biology, chemistry, and physics, as well as agricultural science, mathematics is seen as being indispensable (Johnson, 2004). The reason the subject is given priority as a school subject is due to its special significance. In reality, the Global Relationship for the Assessment of Instructive Achievement (IEA) (1974) has also linked studying mathematics to crucial foundations for adult life. In order to handle a variety of practical tasks and real difficulties, science is also used to dissect and communicate data and ideas (Ann & Bill, 2006). Again, Employers in the engineering, construction, pharmaceutical, finance, and retail sectors have all stated that they continue to need individuals with the right mathematics skills (Smith, 2005). At the secondary school level of education, this circumstance requires that all students be integrated in mathematics instruction right within the classrooms (Sydney, 2019; Hill, 2001). Numerous accounts from individuals and groups of individuals provide substantial evidence that secondary school students generally perform poorly in mathematics around the world. For instance, the National Research Council reported in the late 1980s that the global state of mathematics education was deteriorating, particularly with regard to the enrolment and performance of minority groups in mathematics and science courses (Ezeife, 2002). Locally, similar reports on students' poor performance on mathematics were noted (Chief Examiners' report, 1993- 2000; Raimi, 2001; Igbo, 2004; Aguele, 2004). It is unfortunate that the general performance of students in mathematics has been observed to be poor and which will have either positive or negative effect on the reliability (Agwagah, 2000; Ekele, 2002; Kurume, 2004).

Reliability is the extent to which measures yield consistent result when it is administered to the same person(s) at different times and/or locations (Ary, Jacobs, Razavieh, 2002). When a test produces different scores when it is re-administered to the same group of persons, it implies that scores are different when the trait (or item) is not; such a test cannot be validly measuring what it was designed to measure. Furthermore, the consistency or stability of a measurement is referred to as reliability. The respondent will receive the same score on subsequent administrations of the test or instrument if no other unrelated variables have an impact on the result. Since random errors are likely to alter repeated assessments of any occurrence, a respondent rarely receives the exact same score over testing. Thus, a totally reliable measure is uncommon to achieve, the aim of testing is to eliminate chance errors and increase measurement dependability (Daniel &Frederick 2008).

The exact test result and generalizability are shown in reliability. Consistent results also increase trust that the outcomes are connected to the measurable attribute. On the other side, inconsistent scores hinder students' progress since they give the idea that a student is strong or poor in a particular topic when that may not be the truth, creating pressure for the student to do more or less than they are capable of (Naibi & Louisa, 2013).Furthermore, reliability is obtained through statistical correlation, and is expressed mathematically as a numerical value called a coefficient, represented by the symbol r, and with values ranging between -1.00 to +1.00 (Joseph, 2005).

A value nearer to 0.00 indicates inconsistent scores, while values nearer to + 1.00 shows greater score similarities/consistency (Joseph, 2005). Accuracy of test results is a concern of reliability.

The Classical Test Theory (CTT), whose central tenet is that an individual's achieved score(X) is the sum of a true score and an error score that is seen, is one of the main theories utilized to describe the concept of reliability in this research endeavour (Thissen, 2017). It is represented by the simple equation: $X = T + E$, where X represent the actual obtained score, T represent the true score and E represent the error component. The true score comprises the candidate's stable characteristics on the trait of interest, and reflects the exact value of his ability [or achievement] on that trait (Thissen, 2017), the qualities of the candidate, the test, and the testing environment that have nothing to do with the trait of interest but nonetheless affect the score and cause it to vary are what's known as the error score, officially called measurement error andthese include student error, guessing, fatigue, stress, illness, motivation, excitement; poor examination environment like lighting or seating; or ambiguous test instruction and/or item (McCormick and Pressley, 2019).

The CTT essentially implies that the true score is constant and that score variance is caused by random or measurement errors originating from these causes. Consequently, the error score results in inconsistent scores whereas the real score helps to ensure consistency (Guy, 2000). Consequently, true score has been described as the score that a candidate would have gotten when the measurement is error-free (Onunkwo, 2002). Finally, he CTT believes that error-free measurements are theoretically impossible because all measuring devices (and circumstances) are susceptible to some degree of mistake (Mehrens and Lehman, 1992). This is why, despite being theoretically desired, perfect score consistency, in form of +1.00 is often impracticable in test application (Salvia & Ysseldyke, 2001).

Furthermore, there are various indices or techniques for measuring reliability such as test-retest method, split half method, Equivalent form method, Cronbach Alpha and Kuder Richardson 20 (Traub, 2018). The equivalent method proposes creating two equivalent versions of the same test, which should yield the same results, administering these tests to the same individuals, and correlating the results to assess how closely the two results coincide. Repeated measures (test-retest) include giving the same test—or parallel versions of the same test—to the same individuals on two separate times, the results are then correlated to produce a coefficient. The test-retest coefficient is known as the temporal stability index because it reveals the stability and the extent to which the test results may be generalized across various occasions and times (Ashworth, 2010). Unfortunately, due to two administrations occasions, the use of test-retest are not achievable in so many testing situations (Burton, 2001).

This research work will adopt single administration type of reliability which include: Split-half and Kuder- Richardson, because Internal consistency is achievable, however Cronbach Alpha won't be added or used in this work because, they are only applied on items that are multiple scored while the items on the research work are dichotomously scores. These techniques check how consistently test takers respond correctly or incorrectly across test items sorted according to their degree of difficulty in order to determine how consistently a test item measures the same feature. They are utilized for test questions that measure uniform or comparable subject, like a mathematics test. The reason behind the use of these methods is that since the items in such tests are the same, and designed to measure the same factor or construct, the candidate should respond the same way, or consistently, to all the items, or at least a large number of them. Consequently, most of the responses of a good candidate, on a homogenous test should correlate or agree with each other (Cronbach, 1984).

The obtained co-efficient is called an internal consistency index, by far the most reported indexes of internal consistency reliability are the Cronbach Alpha, for scales, and the Kuder Richardson 20 (K-R 20) formula to tests internal consistency (Cortina, 1993). The K-R 20 is used for dichotomously-scored tests, such as multiple-choice tests and is more suitable for power rather than speeded tests (Onunkwo, 2002). In a nutshell, score inconsistency is as a result of random measurement errors, and the reliability status of a test is the extent in which scores are free from these random errors, for a particular group of candidates. Based on this, many reasons of measurement error came from the procedures involved in developing, standardizing on the part of test developers and guessing the answer to the test items by the student, another contributory factor could be the order in which the items are arranged (Osadebe, 2015).

Item order is the process of arranging test items in a systematic pattern, in which they may appear in ascending, descending and inconsistent order of difficulty (Dong and Cory, 2022). Test item is usually transformed before it is administered to the representative group in order to obtain reliability. As such, item ordering may have the potential of affecting the reliability coefficient, either positive or negative. Items are typically placed in accordance with their degree of difficulty; however, this study will focus on item order in an objective test (multiple choice test). Items are typically arranged according to the type of item utilized (ascending, descending and inconsistent level of difficulty). The degree of difficulty or easiness of an item is expressed by a numeral, called its difficulty index (or p-value), computed during item analyses. p-values which ranges between 0.00 to +1.00; easier items have higher values nearer +1.00, such as 0.75, 0.86, and test is more difficult when one has lower values, nearer 0.00, such as 0.22 or 0.35 (Osabede 2015). The canonical strategy for positioning educational (achievement) test items according to their difficulty is in ascending order (Anastasi & Urbina, 2017), starting with simple items and ending with the most difficult ones. The idea behind this is that if candidates answer the easier questions first and are successful, it will build their confidence and give them a mental boost, which will stimulate them, lower their exam anxiety, and promote more successful answers to the following difficult questions (Mehrens and Lehman, 1992). However, a candidates who encounter the more difficult items first (descending order), especially in a timed test, may spend a lot of time on one specific question and not finish the test before the simpler items. Items may also be placed in an inconsistent order (mixed order), this method involves placing difficult items throughout the test at specified intervals, then followed by subsequently easier ones. The ideas behind this method is that an ascending order technique disappoint the candidate when they encounter and attempt too many difficult items in a row. Consequently, they end up not answering these items at all, guessing, and cheating on them, and this can't showcase the candidates' true ability on that trait (Ekele, 2002).

The inconsistent order is particularly eminent because it's this technique that is been used by the Stanford Achievement Test, one of the oldest and foremost standardized achievement tests in the United States (Davidson, 2019). The preceding review has shown results that have sometimes been conflicting, depending on the situations or variables involved. All of the variables have been linked to test results, and numerous studies have demonstrated that item arrangement affects test scores but has no significant impact on reliability of test scores. Since test scores determine the psychometric criteria, which are essential for any measurement strategy, any factor that impinges on the reliability of test scores is a threat to precision, accuracy, objectivity, external validity and overall validity (Pettyetal, 2009).

Several studies have showed that the specified reliability coefficients have been inconclusive. According to Brenner (2010) found that item arrangement significantly influenced test performance, but this relationship occurs when the examinations were administered as speed test rather than power tests. A recent researcher found out that significant relationship between item order of difficulty and reliability existed only when the students were allowed 35minutes of testing time on an 80-item test.

The aim of this study is to investigate effect of item order on reliability of mathematics test among senior secondary school students in Rivers State. In order to achieve the specific objectives of this study, the following research question was answered:

- To what extent do ascending order of difficulty, descending order of difficulty and inconsistent order of difficulty have effect on the reliability of mathematics test?

The following null hypothesis was formulated and tested at 0.05 alpha level.

- Ascending, descending and inconsistent order of difficulty do not have significant effect on reliability of mathematics test among secondary school students in Rivers State.

METHODS

The study adopted a quasi – experimental research design. The design is ideal and relevant for the study because it investigated on the causal impact of item order on reliability of mathematics test without random assignment. It involves the manipulation of the independent variable without random assignment of participant to conditions or orders of condition. Quasi experimental research does not rely on random assignment, instead subjects are assigned to groups based on non-random criteria. The target population was public senior secondary school students in nine schools in Rivers State. The nine schools were drawn from three local governments using simple random sampling technique and each local government produces three senior secondary school each.

As at the time of study, the population of the study was 904 senior students in all the nine schools (RVME, 2019). Three secondary schools in Port Harcourt Local Government produces 401 SS2 students, Obio/Akpor Local Government produces SS2 290 students and Emohua Local Government produces 213 SS2 students, which sum up the population of 904. The sample size of 507, students was drawn from the population through Stratified Random Sampling Technique with the application of Taro Yemen Formula which was used to determine the minimum sample size. The instrument for data collection was Multiple Choice Performance Mathematics Test (MCPMT). This instrument was arranged in ascending, descending and inconsistent order of difficulty. The reliability coefficient of the MCPMT was 0.75 which was established through split- halfto get half reliability coefficient which was later converted to the reliability of full test using Spearman Brown Formula. Mean and Standard Deviation were used to answer the research question while ANOVA and t-test were used to test the null hypothesis at 0.05 alpha level.

RESULTS

Research Question 1: To what extent do ascending order of difficulty, descending order of difficulty and inconsistent order of difficulty jointly have effect on the reliability of mathematics test?

Table 1: Mean and Standard Deviation of item order on the reliability of mathematics test

Item Order	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
Ascending Order	169	19.9408	15.84174	1.21860	17.5351	22.3466
Descending Order	169	18.4379	14.35510	1.10424	16.2579	20.6178
Inconsistent Order	169	19.5207	15.22924	1.17148	17.2080	21.8334
Total	507	19.2998	15.13761	.67228	17.9790	20.6206

Table 2: Test of Homogeneity of Variance

		Levene Statistics	df 1	df 2	Sig
Reliability	Based on Mean	3.621	2	504	.027
	Based on Median	.545	2	504	.580
	Based on Median and with adjusted df	.545	2	500.440	.580
	Based on trimmed mean	3.300	2	504	.038

Table 1 outlined the N (number of students), mean, Std. Deviation, std. Error, lower Bound and upper Bound of 95% confidence interval for mean, the minimum score and maximum score, respectively for each of the three-item order (ascending, descending and inconsistent order). The same statistical description is also given for the Total of 507 students that were observed and measured. Each of the item order were administered to 169 students, the mean and std. Deviation respectively for ascending order are 19.9408 and 15.84174 for descending are 18.4379 and 14.35510 for inconsistent order are 19. 5207 and 15.22924 and for the total are 19.2998 and 15.13761. It is seen that the mean score and standard deviation of ascending order of difficulty are nearly the same with inconsistent order of difficulty but higher than descending order of difficulty. However, it can be expressed that item arranged in ascending order of difficulty has the highest mean and standard deviation followed by inconsistent order of difficulty, while the descending order of difficulty has the lowest mean and standard deviation.

Table 2 is a test of homogeneity of variance. The table contain Levene Statistics, df1, df2 and sig that are significant, indicating that the variance of Reliability of the three-item order are not homogeneous and are therefore not assumed to be equal across the three groups.

Hypothesis 1: Item Order (Ascending, Descending and Inconsistent Order of Difficulty) do not have significant effect on the reliability of mathematics test among secondary school students in Rivers State.

Table 3: Analysis of Variance (ANOVA) of Item Order on Reliability of Mathematics Test

Reliability					
	Sum of Squares	df	Mean Square	F	Sig
Between Groups	203.247	2	101.623	.443	.643
Within Groups	115745.183	504	229.653		
Total	115948.430	506			

Table 4: Homogeneous Subsets

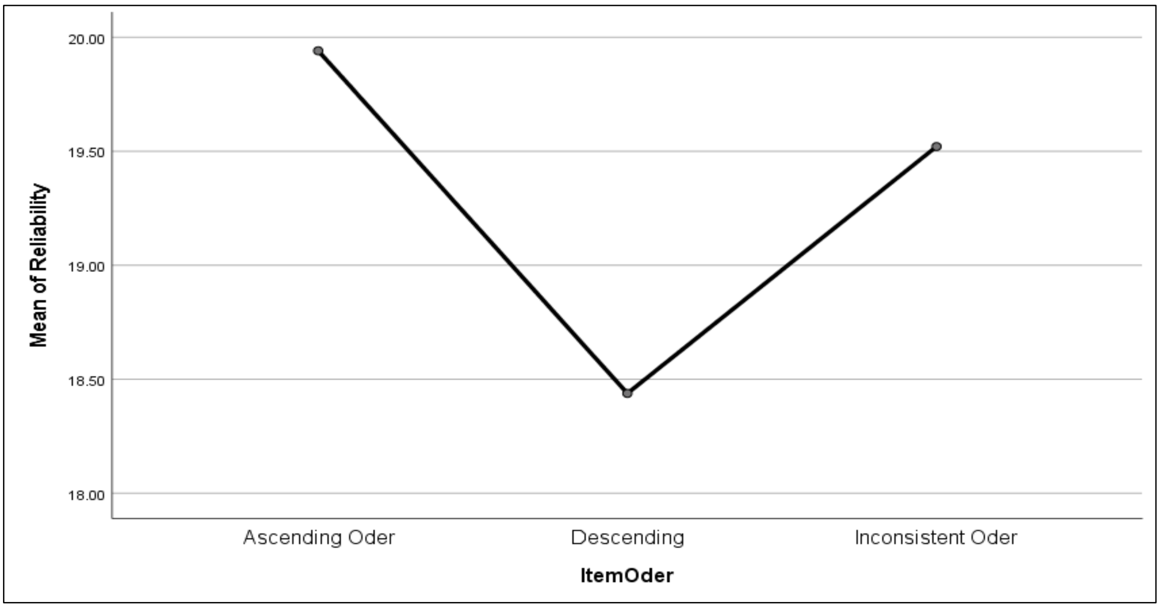
	Item Order	N	Subst for alpha = 0.05
			1
Scheffe[a]	Descending Order	169	18.4379
	Inconsistent Order	169	19.5207
	Ascending Order	169	19.9408
	Sig		.660
Ryan-Einot-Gabriel –Welsch Range	Descending Order	169	18.4379
	Inconsistent Order	169	19.5207
	Ascending Order	169	19.9408
	Sig		.633

Table 3 is the main table on which testing of the postulated null hypothesis depend. It shows the ANOVA F, mean square, degree of freedom (df) and sum of square the two sources of variation (i.e. Between Groups and the Within Groups) as well as the total. For the Between Group, the mean square is 101.623 which is function of its sum of square (203.247) divided by the df (2). For the Within Groups, the mean square is 229.653 that is the function of its sum of square 11574.183 divided by its df 504. The total sum of square is 115948.430 (an addition of the sum of square for Between group and Within Groups) and the Total df of 506 is derived from adding the df for Between Group and Within Groups. The F is 0.443, got as a function of Between Groups Mean Square divided by the within Group mean square. The computed F of .443 has .643 (read as higher than .001 or greater than .0005) sig. P (value).

Since the P-Value (sig) of .643 is greater than the chosen alpha of .05, the null hypothesis that "the three-item order (ascending, descending and inconsistent order of difficulty) do not have significant effect on reliability of mathematics test" is accepted. Therefore, the alternate effect of the three-item order (ascending, descending and inconsistent order of difficulty) on reliability of mathematics test "is rejected.

Table 4 shows that each of the three-item order of difficulty has a reliability mean that is significantly different from the other. Each of the mean difference is statistically significant as no two-mean fall under the same homogenous subset.

Means Plots



The mean plot has graphically expressed the relative position of the types of item order of difficulty (ascending, descending and inconsistent order) along the horizontal axis against mean of reliability at the vertical axis. This is the pictorial description of the significant effect of the item order on reliability to show that the descending order of difficulty is incomparably lesser reliability co-efficient than the ascending and inconsistent order of difficult.

DISCUSSION OF FINDINGS

The findings revealed that ascending, descending and inconsistent item order of difficulty shows no significant effects on reliability of mathematics test. This implies that consistency in an examinee scores do not depend on the arrangement of test items from simple to complex, complex to simple or in an inconsistent manner but on the construction of good test items during classroom assessment/testing. Thus, it implies that the study will enable student to come to the consciousness that item arrangement doesn't matter but ability to prepare well is the ultimate and ability of a test constructor to construct a good test. The quality of individual test items in particular and the total test in general must be guaranteed by the body responsible for the development and administration of the test, if this is done it will take far reaching decision on whatever may be responsible for any short coming in the examination especially with regards to reliability.

The findings agreed with the study of Bodas and Ollendick (2005) who argued that test item arrangement of any format does not have effect on reliability. This result is however in line with that reported by Ollemu and Essey (2008) who noted that item order does not affect reliability coefficient but students tend to perform significantly better in their mean score when questions are arranged in a mixed order format. Similarly, Barbara, Ansorge, Parker and Lowry (2005) reported finding in support of the present one.

The findings disagreed with the study of MacFarland, Ryan and Ellis (2002) who found that item arrangement has effect on psychometric properties. The difference in findings may be attributed to testing environment, test duration, etc. Overall, this study found results similar to Carlson and Ostrosky (1992), Gohman and Spector (1989) and Brenner (1964), Monks and Stalkings as noted by Tei-Firstman (2011) while it is in direct contradiction with that reported by Bodas and Ollendick (2005) who argued that test item arrangement of any format does not have effect on reliability.

CONCLUSION

From the result of the findings, the conclusion is as follow:

Item order based on ascending, descending and inconsistent order do not have significant effect on the reliability of mathematic test among secondary school students in River State.

RECOMMENDATIONS

1. All the three pattern of test arrangement (ascending, descending and inconsistent) as stipulated in this research work should be encourage among students as it has no significant effect on the consistency of test scores (reliability).
2. Student should be motivated to improve their real abilities in mathematics related courses and not judging their underperformance based on the arrangement of test items.
3. Seminars and workshop should be organized on a regular basis for classroom teachers and examination bodies to update their knowledge on the use of professional processes in test construction such as trial testing of test items, use of test and measurement expert in drawing test items etc. which likely not strictly followed.

REFERENCE

- Anastasi, A., & Urbina, S. (2017). *Psychological testing* (7th Ed.), Engel wood Cliffs, NJ: Prentice Hall.
- Ashworth, A. E. (2010). *Testing for continuous assessment*. Ibadan: Evans Brothers.
- Aguele, L. I. (2004). Remediation of Process Errors Committed by Senior Secondary School Students in Sequences and Series. Unpublished Ph.D. Theris, UNN.
- Ann, B. V., Wilbur, D. M., & Dutton, L. (2016) *Mathematics Children use and Understanding.* Napa, California: Rattle OK Publishers.
- Ary, D., Jacobs, L. C., & Razavieh, A. Z. (2002) *Introduction to Research in Evaluation (6th Ed.).*Australia. Thomson Learning Wadsworth.
- Agwagah, U. N. V. (2000). *Teaching Number Bases in Junior Secondary school, Board, Abacus*, Journal of Mathematics Association of Nigeria.
- Bodas, J., & Ollendick, T. H. (2005). Test Anxiety: A cross-cultural perspective of clinical child and family psychology review, 8(6), 65-88.
- Babara, S. P., Ansorge, C. J., Parker, C. S., & Lowry, S. R. (2005). Effect of item arrangement, knowledge of arrangement, test anxiety and sex on students' performance. *Journal of Educational Measurement, 19*(1) 49-57.
- Brenner, M. H. (2010) *Test difficulty, reliability, and discrimination as a function of item difficulty order*. Journal of Applied Psychology, 48, 98-100.
- Burton, R. F. (2001) *Do item-discrimination indices really help us to improve our tests?* Assessment and Evaluation in Higher Education, 26(3), 2\3- 220.
- Brenner, M. H. (1964). Test difficulty, reliability, and discrimination as functions of item difficulty order. *Journal of Applied Psychology*, 48, 98–100.
- Cronbach, L. J. (1984). Essentials of psychological testing (4th ed.). New York*: Harpers and Row.*

- Cortina, J. M. (1993). What is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*, *78*(1), 98-104.
- Carlson, J. L., & Ostrosky, A. L. (1992). Item sequence and student performance in multiple-choice examinations. *Journal of Economics Education*, 23, 232-235.
- Denny, B., & Dylan M. (2018). Psychometrics. University of Amsterdan, Amsterde, Netherlands.
- Daniel, L. S., & Frederick, L. C. (2008). The SAGE Encyclopedia of Lifespan Hu man Development. SAGE Publications, Inc, Thousand Oak.
- Dong, X., & Cory, C. (2022). Item Analysis. https://doi.org/10.1002/9781118970843.Ch97
- Davidson, M. I. (2019). Review of the Stanford Achievement Test. In J. V. Mitchell, (Ed.) 9th Mental Measurement Yearbook. Lincoln: *Buros Institute of Mental Measurement.* Retrieved November 13 2013 from http:/www.unl.edu/buros.html.
- Ekele, J. (2002). Development and standardization of quadratic aptitude test for upper primary School Pupils. Unpublished PhD Thesis. University of Nigeria Nsukka.
- Ezeife, A. N. (2002). Interactions of Culture and Mathematics in an Aboriginal classroom. http://ehlt.flinders.edu.au/education/iej/articles/v3n 3/ Ezeife/paper.pdf.
- Guy, F. Y. (2000). Measurement and evaluation. Bunton, *Illinois: Apton.*
- Gohman, S. F., & Spector, L. C. (1989). Test scrambling and student performance. *Journal of Economics Education*, *20*(3), 235-238. doi: 10.2307/1182298.
- Hill, B. (2001). The Importance of Mathematics in Early Childhood Education. Saint Martin's College Lindsey Petersen, USA. Saint Martin's College press.
- IEA. (1978). IEA = International Association for the evaluation of Educational Achievement.
- Igbo, J. N. (2004). Effect of Peer Tutoring II on the Mathematics Achievement of Learning-Disabled children. Unpublished Ph.D. Thesis, University of Nigeria, Nsukka.
- Johnson, O. C. (2004). Test and measurement in child development: handbook II, San Francis co; *Jersy. Bass. Inc. 1*(2).
- Kurume, M. S. C. (2004). Effects of Ethnomathematics Approach on Students Achievement and Interest in Geometry and Mensuration, Ph.D. Thesis, University of Nigeria, Nsukka.
- MacFarland, L., Ryan, A., & Ellis, A. (2002). Effects of item placement on faking behavior and test measurement properties in personality testing. *Journal of Personality Assessment*, *78*(2), 348-369.
- Mehrens, W. A., & Lehman, I. J. (1992). Measurement and evaluation in education and psychology (4th ed.). Forth Worth: *Harcourt Brace College Publishers.*
- McCormick, C. B., & Pressley, M. (2019). Educational psychology: Learning, instruction, assessment. New York: *Addison Wesley Longman.*
- Naibi., & Louisa. (2013). Effect of Item arrangement on test reliability co-efficient: implication for testing. *Journal of research in education and society*, *9*(5) 55-56.
- Onunkwo, G. I. N. (2002). Fundamentals of educational measurement and evaluation. *Owerri: Cape Publishers.*
- Opara, I. M. (2021). Test construction and measurement, concepys and applications Owerri: Reliable Publishers.
- Ollemu, S., & Etsey, Y. K. A. (2008). The impact of test item position on multiple-choice test on students' performance at the Basic Education certificate Examination (BECE) level in Ghana. Retrieved 17th July 2015 from http://www.edu.ghana.WAEC/outline.cog.2hthl.
- Osadebe, P. U. (2015). Construction of valid and reliable test for assessment of student. *Journal of Education and Practice.*
- Raimi, A. (2001). Excerpts from Poor Performance Review. University of Rochester, Washington. http://www.Mathematically Correct. Com/mspap.tm
- Sydney, S. L. (2019). En-chanting, Fascinating, Useful number. Teaching Children Mathematics 1, 486-491.
- Smith, M. R. (2005). Making mathematics count: The Report of professor Advain Smiths Inquiry into Post 14 Mathematics Education: *Mathematics for citizen. UK.*
- Salvia, J., & Ysseldyke, J. (2001). Assessment (8th ed.). Boston: *Houghton Mifflin.*
- Thissen, D. (2017). Reliability and measurement precision. In H. Wainer (Ed.). Computerized and adaptive testing: A primer. Hillsdale, NJ.:*Lawrence Erlbaum* Pp. 161-185.
- Traub, R. E. (2018). Measurement methods for the Social Sciences - Reliability for the Social Sciences: Theory and applications, *Thousand Oaks, CA: Sage. 3*(6).
- Tei-Firstman, R. I. (2011). Test item arrangement on student test scores. *M.Ed Unpublished thesis,* University of Port Harcourt.
- Petty, R. E., Brimol, P., Loersch, C., & McCaslin, M. J. (2009). The need for cognition in M.R. Leary & R.H Hoyle (Eds), Handbook of Individual Differences in Social behavior (pp.318-329). *New York. NY Guilford Press.*