

IMPACT OF NUMBER OF OPTIONS ON PSYCHOMETRIC PROPERTIES OF WASSCE MULTIPLE CHOICE MATHEMATICS EXAMINATION AMONG SENIOR SECONDARY SCHOOL STUDENTS IN KATSINA STATE, NIGERIA

Arinde, Kazeem Ajiboye

Department of Educational Psychology and counseling
School of Education, Federal College of Education, Katsina State

[Email: kazeemarinde01@gmail.com](mailto:kazeemarinde01@gmail.com)

Phone No: +2348059788094

Dr. Adekunle Thomas Olutola & Dr. Habibu Kamilu
Department of Educational Psychology and Counseling
Federal University Dutsin-ma, Katsina State

Abstract

The study investigated the impact of number of options on psychometric properties of the West African Senior School Certificate Examinations (WASCE) multiple choice mathematics paper in Katsina State, Nigeria. The study employed a repeated measures research design. The total population of this study comprised of seventeen thousand, two hundred and fifty-seven (17,257) mathematics students. A sample size of 378 SSSII students was selected through multi-stage sampling technique with proportional sampling technique and Krejcie and Morgan (1970) table for determining the sample size for the study. The instrument for data collection was 2023 WASSCE objectives test item. Validity and reliability were not conducted, this is because, they had already been carried out by the experts in the examination's body. The data collected was analyzed using Cronbach alpha reliability. The findings showed that the reliability indices of the multiple choice WASSCE May/June 2023 Mathematics Examination is 0.81. This implies that the multiple choice WASSCE May/June 2023 Mathematics Examination is reliable. Based on the findings, it was recommended that the Mathematics teachers in Nigerian secondary schools should be committed to the use of multiple choice on a regular basis in order to promptly assess each student's academic achievement.

Keywords: Number of Options, Psychometric Properties, Multiple Choice Items, Mathematics.

Introduction

Examination is designed to meet the requirements of assessing intelligent quotient of an individual learner and was administered to determine whether students could advance from one class to another. Examination is also test of knowledge acquired, or generally a means of assessing Intellectual capacity or ability of an individual (Dunlosky, 2013). The West African Examinations Council (WAEC) is the one of the major examination bodies in Nigeria that have mandate to award Senior School Certificate in various subjects, such as Mathematics, English Language, Biology, and Economics to mention a few.

Assessment in Mathematics is considered as an indispensable tool needed for the transformation of technological development to reality. This is in agreement with Abisoye (2019) who described mathematics as the supporting knowledge of modern sciences. This

means that scientific development depends on a deep understanding of mathematical concepts and procedures; hence a meaningful understanding of mathematics must occur. Mathematics is seen as a fundamental science that is necessary for understanding of most other fields in education. It is glaring that no other subject forms such a strong force among the various branches of science. Ibrahim (2022) affirmed that the knowledge of science remains shallow without Mathematics. It therefore means that, the position of Mathematics in Secondary School Curriculum is very important for scientific development of any nation.

Psychometric analysis is the science of measuring latent traits or constructs in our subjects of interests and it imply analyzing the following psychometric properties. Validity is whether a test measures what it is intended to measure. Reliability is the consistency in measuring what it intends to measure. Difficulty is the easiness or difficulty index in an item and discrimination index relates to how the test items distinguishes between low and high ability level of students (Adonu, 2014). Obilor and Obubere (2020) stated that validity of a test implies that a test measures what it sets out to measure and nothing else. Also reliability means the consistency in measuring what is intends to measure.

One of the primary concerns of testing in the teaching and learning process is that the measurements derived from tests will be helpful in making valid decisions. Examiners and other test users are usually interested only in the results yielded by the test administration. Still, the uses of test results do not realize that the usefulness and appropriateness of test-score interpretation is a direct result of the test's internal characteristics. The internal attributes of a test and its items are technically termed as psychometric properties or characteristics. Psychometric characteristics of a test items refer to certain attributes inherent in the test upon which assessment of conditions is based (Moyinoluwa in Abdullahi & Darazo, 2020). Indicators of a particular test item's difficulty, its ability to discriminate among people having differing amounts of the construct being measured, and the likelihood that the correct answer could be chosen by guessing are each a psychometric property of an item. It is important to investigate and learn about tests' psychometric properties.

Knowing about the psychometric properties of multiple choice option of numbers and test items provides evidence that information obtained using such a test (instrument) can provide a sound basis for decision making. Psychometric properties of multiple choice option of numbers Items include several indicators such as Item Difficulty, Item Discrimination, Item Guessing Parameter and Distracter Index. Item difficulty tries to estimate how easy or difficult is the item, the higher the value, the easier the item or the lower the difficulty. This means that difficulty can be considered as the location index (Ashraf & Jaseem, 2020). This refers to the proportion of examinees that correctly answered an item. This metric takes a value between 0 and 1. High values indicate that the item is easy, while low values indicate that the item is difficult. An ideal item is supposed to have a difficulty index of 0.5, but may be difficult to have items with this index. Hence, an item is acceptable if the difficulty index falls between 0.3 and 0.7. If the difficulty index is less than 0.3, it shows that the item is difficult, while any value greater than 0.7 indicates that the item is very easy.

Moreover, another example of psychometric properties is item discrimination. Item discrimination describes how well an item can differentiate between examinees having the trait below the item location and those having the trait above the item location. According to Okoye (2015), an item is considered good if it is got right by the bright students and failed by the dull ones. Item discrimination refers to the power of the item to differentiate between examinees with high and low levels of knowledge or ability (Thompson, 2016). It is the correlation between item scores and total test scores called the item-total correlation. A good item records passes in the upper one-third than in the lower one-third. The discrimination index ranges from -1 to +1. When the value is negative, it shows that more of the lower group chooses the correct option than the upper group (this is an abnormal situation, and such item is a bad one). On the other hand, items with positive values are most acceptable when the discrimination index is +1.0 and very poor in discrimination if the value is 0. An item is acceptable if the discrimination index falls between +0.3 and +1.0 (Okoye, 2015). In the study of impact of number of options in psychometric properties of WAEC multiple choice mathematics examination, there are two major theories, which are classical test theory and item response theory and their corresponding models.

In addition, using the p values (difficulty indices), discrimination indices (D) can be calculated for each dichotomous item. The higher the D, the more the item discriminates. Items with p-levels in the midrange usually have the best D values and, as will be demonstrated shortly, the opportunity for D to be highest occurs when the p-level for the item is at 0.50. The extreme group method is used to calculate D. There are three simple steps to calculating D. First, those who have the highest and lowest overall test scores are grouped into upper and lower groups (Olutola, 2015).

The upper group is made up of the 25%–33% who are the best performers (have the highest overall test scores), and the lower group is made up of the bottom 25%–33% who are the poorest performers (have the lowest overall test scores). The most appropriate percentage to use in creating these extreme groups is to use the top and bottom 27% of the distribution, as this is the critical ratio that separates the tail from the mean of the standard normal distribution of response error (Cureton, 2009). Step two is to examine each item and determine the p-levels for the upper and lower groups, respectively. Step three is to subtract the p-levels of the two groups; this provides the D. Assume that these data are based on 500 individuals taking a test that is 50 items in length. The highest scoring 135 individuals (500×0.27) for the entire test and lowest scoring 135 individuals for the entire test now make up our upper and lower extreme groups. For Item 1, the upper group has a p level of 0.80 and the lower group has a p level of 0.30. The D, then, is $0.80 - 0.30 = 0.50$. For Item 2, the D is 0.80; for Item 3, it is 0.05; and for Item 4, it is -0.60 .

The psychometric properties (quality of test instrument) are always examined through responses of the examinees that have similar characteristics before being used for the purpose the test is meant for. Thus, items could be analyzed through different models (Adedoyin & Mokobi, 2013). There are two main frameworks by which a test and the items it contains can

be analyzed. It could either be through the Classical Test Theory (CTT) or Item Response Theory (IRT). The central model of CTT is that, observed scores are composed of the true score and error score, denoted as $X = T + E$. Where X = Observed score, T = True score and E = Error score (Juliet, 2018). Classical test theory (CTT) of measurement assumes that each individual has a true score which would be obtained if there are no errors in measurement. It must be noted that each item in a test has three parameters either in CTT which are item difficulty, discrimination and guessing (Jimoh, Daramola, Oladele & Sheu, 2020). The reliability and validity of mathematics multiple choice of type A (five options), type B (four options) and type C (three options). Items discrimination, items difficulty and plausibility of items are detection in multiple choice Mathematics set by WAEC.

Multiple choice items consist of a stem and several alternative answers. The stem is the opening a problem to be solved, a question asked, or an incomplete statement to be completed. The options are the possible answers that the examinee can choose from, with the correct answer called the key and the incorrect answers called distractors. Only one answer may be keyed as correct (Feldet, 1989). Test theories provide a general framework for linking observable variables such as test/examination scores and ability scores of individuals. There are two currently popular statistical frameworks which address measurement problems such as test development and their validity- classical Test Theory (CTT) and Item Response Theory (IRT). Although the models of CTT and IRT were developed for the most part with objective WASSCE mathematics multiple choice test in mind just like the SSCE. They are best suited for this study and apply more broadly to any data collection procedures where the ultimate aim is to arrive at quantification scores of any phenomenon being measured in order to make a decision (Eze, 2016).

In contrast, the models of CTT are generally specified at the level of the test as a whole but several advantages flow the specification of IRT models at item level. This feature of IRT models provides a firm stateside basis for scientific test design. In the case of WASSCE, total test scores are most frequently used to make decisions or relate to other variables of interest. And since the total test score is only as good as the sum of its parts and that means its items, shifting through item-level analysis may seem tedious and beyond the scope and intent of this study (Yu, 2018).

In psychometrics, Items Response Theory also known as latent trait theory, strong true score theory or modern mental test theory, is a paradigm for the design, analysis and scoring of tests, questionnaires and similar instrument measuring abilities, attitudes or other variables. This theory does not assume that each item is equally difficult but it treats the difficulty of each item as information to be incorporated in scaling items (Yu, 2018). It is based on the application of related mathematical models to testing data. Because, it is generally regarded as the superior of classical Test Theory. It is the preferred method for developing scales, especially when optimal decisions are demanded, as in high stake examination. The name Item Response Theory is due to the focus on the theory on the item, as oppose to the test-level focus of classical

test theory. Thus, IRT models are the response of each examinee of a given ability to each item in the test (Yu, 2018).

Statement of the Problem

The performance of students in WASSCE are not comparable. This is based on the Examiner's reports. In the year 2023 NECO Mathematics examination 86.78% pass at credit level while 78.56% of the students who sat for WASSCE Mathematics examinations in the same year passed at credit level. Since the examinations were attempted by same set of students taught by the same teacher, WASSCE and NECO have similar syllabi and award certificates for same purposes, their items are expected have comparable standard with comparable performance. It is probable that some technical weaknesses exist in the multiple choice tests developed and administered by the WASSCE. Therefore, it is necessary to examine the impact of number of options on psychometric properties of WASSCE multiple choice paper Mathematics quality of the items constructed by the examination bodies.

Despite the importance of Mathematics towards achieving scientific and technological developments, it is sad to observe that students' achievement in the subject has remained very low especially in Senior Secondary School Certificate Examination. Studies such as Agashi (2014), Agwagah and Utibe, (2015), Galadima and Yushau (2017), Uloko, and Usman (2017) and Kamilu (2023) separately indicated that there is a poor and fluctuating state of students' academic achievement in Mathematics. Evidence in literature has also shown that most senior secondary school students achieve poorly in many topics of the Mathematics curriculum (Ali, 2016). This is evidenced in the Chief Examiner's Reports (2015 & 2018) of the National Examinations Council (NECO) which consistently reported a poor and fluctuating achievement of students in Mathematics between 2015 and 2018. Frequent failure of Nigerian students in Mathematics subject over the years has been the concerned of all stakeholders in education industry. However, it has been observed that students perform poorly in the Mathematics West African Senior School Certificate Examination (WASSCE) in Nigeria. This became worrisome and doubtful of the extent of quality of WASSCE administered in different years. Therefore, this study investigate the Impact of Number of Options on Psychometric Properties of WASSCE Multiple Choice Mathematics Examination among Senior Secondary School Students in Katsina State, Nigeria.

Objectives of the Study

Specifically, the objectives of the study are to:

1. Find out the reliability index of the multiple choice WASSCE May/June 2023 Mathematics Examination in Katsina State, Nigeria.
2. Determine the difficulty index of the multiple choice WASSCE May/June 2023 Mathematics items in Katsina State, Nigeria.
3. Investigate the discrimination index of the multiple choice WASSCE May/June 2023 Mathematics items in Katsina State, Nigeria.

Research Questions

In order to achieve the objectives of the study, the following research questions were raised to guide the study:

1. What is the reliability index of the multiple choice WASSCE May/June 2023 Mathematics Examination in Katsina State, Nigeria?
2. What are the difficulty indices of the multiple choice WASSCE May/June 2023 Mathematics items in Katsina State, Nigeria?
3. What are the discrimination indices of the multiple choice WASSCE May/June 2023 Mathematics items in Katsina State, Nigeria?

Methodology

The researcher adopted the repeated measures research design. Repeated measure research design is a research design in which subjects are measured two or more times on the dependent variable. Rather than using different participants for each level of treatment, the participants are given more than one treatment and are measured after each treatment (Michael, 2003). Repeated measures design is a design that involves the same subjects that participate in all conditions of the independent variable. The meaning of this is that every circumstance of the experiment contains the exact same group of participants. Repeated measures design is also referred to as within groups, or within-subjects design (Kraska & Marie, 2010).

The population of the study consist of seventeen thousand, two hundred and fifty seven respondents (17,257) which comprises all public senior secondary school two students (SSS Two) in Katsina Zonal Education Quality Assurance, Katsina State, Nigeria. The sample of this study consists of three hundred and seventy-eight (378) SSS II students drawn from a population with 27% of 378(102) in the upper and 27% of 378 (102) in the lower group was used for this study. In selecting the students to represent the samples, multi-stage sampling technique with proportional sampling technique were employed. In selecting the sample size Krejcie and Morgan (1970) table for determining the sample was adopted, which says, that for a population of seventeen thousand, two hundred and fifty-seven (17,257), the appropriate sample size is approximately three hundred and seventy-eight (378), which was distributed across the six randomly selected schools in proportion to their population.

The instrument for the study consists of 2023 Mathematics Multiple Choices items constructed by WASSCE which consist (50 items) for the year. The instrument was administered to 378 students in six (6) public secondary schools two students randomly selected for the study and the researcher did not carried out the validity and reliability of the instrument because it had been validated and the reliability had been carried out by the experts.

Results

Research Question One: What is the reliability index of the multiple choice WASSCE May/June 2023 Mathematics Examination in Katsina State, Nigeria? To answer research question one, the participants’ responses were subjected to Cronbach Alpha reliability.

Table 1: Reliability indices of the multiple choice WASSCE May/June 2023 Mathematics Examination

Variable	N	Reliability
WASSCE May/June 2023 Mathematics Examination	378	0.81

As shown on table 1, the multiple choice WASSCE May/June 2023 Mathematics Examination had reliability index of 0.81. Hence the multiple choice WASSCE May/June 2023 Mathematics Examination in Katsina State, Nigeria is reliable.

Research Question Two: What are the difficulty indices of the multiple choice WASSCE May/June 2023 Mathematics items in Katsina State, Nigeria?

Table 2: Summary of difficulty indices of the multiple choice WASSCE May/June 2023 Mathematics Examination

Variable	Very difficult Items (<0.20)	Moderately Difficult items (0.20-0.80)	Very Easy Items (> 0.80)	Total
WASSCE May/June 2023 Mathematics Examination	Nil	50 (100%)	Nil	50 (100%)

The result in table 2 showed that no item of the multiple choice WASSCE May/June 2023 Mathematics Examination in Katsina State-Nigeria has less than 0.20 indices(very difficult) and none as well has above 0.80 indices (very easy), hence all the items independently has difficulty indices between 0.20 to 0.80. It implies that all 50 (100%) of the multiple choice WASSCE May/June 2023 Mathematics items in Katsina State, Nigeria are moderately difficult.

Research Question Three: What are the discrimination indices of the multiple choice WASSCE May/June 2023 Mathematics items in Katsina State- Nigeria?

Table 3: Summary of discrimination indices of the multiple choice WASSCE May/June 2023 Mathematics Examination

Variable	Items that discriminate properly (D=0.20 and above)	Items that does not discriminate properly (D <0.20)	Total
WASSCE May/June 2023 Mathematics Examination	50 (100%)	Nil	50 (100%)

Table 3 showed that all 50 (100%) the multiple choice WASSCE May/June 2023 Mathematics examination items in Katsina State-Nigeria independently has discrimination indices (D) of 0.20 and above, while no item has discrimination indices that is less than 0.20. This means that all 50 (100%) the multiple choice WASSCE May/June 2023 Mathematics Examination items in Katsina State, Nigeria discriminate properly between the upper and lower group examinees.

Discussion of Findings

Findings of the study revealed that items on May/June WASSCE 2023 multiple choice questions for Mathematics are reliable, moderately difficult and discriminate properly between the upper and lower groups of examinees. This findings agrees with the findings of Jimoh, Daramola, Oladele &Sheu (2020) which shows that items in WASSCE questions are validated, tested and found reliable by the examination body. The findings of this study supports the study findings of Moyinoluwa in Abdullahi & Darazo (2020) which states that WASSCE items are moderately difficult to the examinees. Indicators of a particular test item's difficulty, its ability to discriminate among people having differing amounts of the construct being measured. The findings of this study is also in line with the findings of Olutola (2015) and Thompson (2016) all of which revealed that item discrimination power of the items in WASSCE differentiate between examinees with high and low levels of knowledge or ability.

Conclusion

Based on the above findings, the researcher concluded that, the options supplied by WASSCE May/June 2023 multiple choice items for Mathematics examination in Katsina State, Nigeria are moderately difficult, and discriminate properly between the upper and lower group of examinees.

Recommendations

Based on the findings of this study, the following recommendations were made:

1. Mathematics teachers in Katsina state secondary schools should be committed to the use of multiple choice on a regular basis in order to promptly assess each student's academic and achievement progress as expected for proper and accurate decision making.
2. WASSCE as an examination body should ensure that their items are well constructed and the determination of psychometric properties of the test items should also be considered.
3. Measurement and Evaluation experts should ensure that items are adequately distributed across the cognitive domains of Bloom's Taxonomy of educational objectives to adequately cater for both high achievers and low achiever.

References

- Abisoye, B. O. (2019). Mathematical analysis and modelling. *Journal of applied Mathematics and Physics* 8(2) 33-40.
- Adedoyin, O. O. & Mokobi, T. (2013). Using IRT psychometric analysis in examining the quality of junior certificate mathematics multiple choice examination test items. *International journal of asian social science*, 3(4), 992-1011.
- Adekunle, T. O. (2015). Item difficulty and discrimination indices of multiple choice Biology tests. *Liceo journal of higher education research*, 11 (1):16-30
- Adonu, I. I. (2014). Psychometric analysis of WAEC and NECO practical physics tests using partial credit model. *PhD thesis. University of Nigeria, Nsukka.*
- Agashi, P. P. (2014). Effect of advance organizer and content attainment models on the achievement and retention of Pre-NCE students in geometry. Triangle publishing co. Ltd.
- Agwagah, U. N. V. & Utibe, U. V. (2015). A decade of candidates' performance in NECO-SSCE mathematics in Nigeria. *Journal of education and practice*, 6(25), 25 - 30.
- Ali, A. (2016). Student' ranking based on their abilities on objective type: comparison of CTT and IRT. Edu – COM 2008 international conferences on sustainability in higher education: direction for change Edith Cowan University, Australia, held on 19th-21st November, 2018 research online 592-599 [http:// ro.ecu.edu.au/ceducom/52](http://ro.ecu.edu.au/ceducom/52)
- Ashraf, Z.A. & Jaseem, K. (2020). Classical and modern methods in item analysis of test tools. *International journal of research and review*, 7(5), 397-403.
- Cureton, P. (2009) in Felix, F. (2017). Classical test theory, item analysis within classical test theory: approaches, statistical analysis, and interpretation. Booth publishers.
- Dunlosky, J., Rawson, K. A., Marsh, M. J., Willingham, N., & Willingham, D. T. (2013). *What works, what doesn't?* Scientific American publishers.
- Eze, C.P (2016). Effects of guided-discovery on students' academic achievement and retention of concepts in block/brick laying and concreting. Technical books publishers.
- Federal Republic of Nigeria (2014). National policy on education. Lagos: NERC press.
- Federal Republic of Nigeria (2014). National Policy on Education. Abuja: National Educational Research Council NERC Press.
- Felder, L.S., & Brennan, R.L.(1989). Reliability in educational measurement (3rd ed)., pp.105-146, Macmillan.
- Galadima, I. & Yushua, M. A.(2017). Investigation into performance of senior secondary school students in Sokoto State. *Abacus*, 32(1), 24 - 33.
- Hoover, R.N.(2013). Teachers' instructional use of summative student assessment data. *Applied measurement in education*, Lagos.
- Ibrahim, B. M. (2022). Effects of peer-assessment strategy on students' academic achievement in Mathematics in senior secondary schools of Nasarawa local

government area, Kano state, Nigeria, *International journal of research in education and sustainable development*, 1(1)22-30.

Jimoh, M. I., Daramola, D. S., Oladele, J. I., & Sheu, A. L. (2020). Assessment of items prone to guessing in SSCE Economics multiple-choice tests among students in Kwara state, Nigeria. *A natolian Journal of Education*. 1(2)66-71.

Kamilu, H. (2023). Validation of Attitude towards Mathematics Inventory (ATMI) among senior secondary schools in North-west Geo-Political zone of Nigeria. Ph.d Thesis Bayero University. Kano. Nigeria.

Kraska; Marie (2010), "Repeated measures design", *Encyclopedia of research design, California, USA: SAGE Publications, Inc.*, [doi:10.4135/9781412961288.n378](https://doi.org/10.4135/9781412961288.n378), [ISBN 978-1-4129-6127-1](https://www.sagepub.com/ISBN/978-1-4129-6127-1), [S2CID 149337088](https://www.sagepub.com/S2CID/149337088)

Micheal, Y. L., M. K Green, S. A Farquhar (2003) E- Health & place, 2006 – Elsevier publishers.

Moyinoluwa, T. D. (2015). Implementation of the revised 9-year basic education curriculum (BEC) in the North Central Nigeria: A monitor of Benue state. *Journal of research and method in education*.

Obilor, E. I. & Obubere, M. (2020). Development and validation of social studies cognitive achievement test for junior secondary schools on Rivers state. *International academic journal of educational research*, 5(3), 1-20.

Okoye, R.O. (2015). *Educational and psychological measurement and evaluation* (2nd ed.). Awka: erudition publishers.

Thompson, N.A. (2016). Introduction to classical test theory with CITAS. Minnesota: assessment system corporation.

Uloko, E.S. & Usman, K.O. (2007). Effect of ethno-mathematics teaching approach on students' achievement and interest. *University of agricultural journal of research in science and media* 1(2) 67-72.

Yu CH (2018). True score model and item response theory retrieved from <http://www.creativewistom.com/teaching/WBI/measurement>.