ASSESSMENT OF RANDOM AND SYSTEMATIC ERRORS IN 2024 NECO SSCE MATHEAMATICS MULTIPLE CHOICE TEST ITEMS IN ETHIOPE EAST LOCAL GOVERNMENT AREA, DELTA STATE

Emamuyovwi ESUME

UNIVERSITY OF BENIN
BENIN CITY

OCTOBER 2025

ASSESSMENT OF RANDOM AND SYSTEMATIC ERRORS IN 2024 NECO SSCE MATHEMATICS MULTIPLE CHOICE TEST ITEMS IN ETHIOPE EAST LOCAL GOVERNMENT AREA, DELTA STATE

Emamuyovwi ESUME
PG// EDU1313211
B.Sc (Ed) (Abraka)

A THESIS WRITTEN IN THE DEPARTMENT OF EDUCATIONAL EVALUATION AND COUNSELLING PYSCHOLOGY, FACULTY OF EDUCATION AND SUBMITTED TO THE COLLEGE OF POSTGRADUATE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTERS IN MEASUREMENT AND EVALUATION OF THE UNIVERSITY OF BENIN, BENIN CITY, EDO STATE, NIGERIA

OCTOBER 2025

CERTIFICATION

We, the undersigned, certify that this research work was carried out by **Emamuyovwi ESUME.** in the Department of Educational Evaluation and Counselling Psychology, Faculty of Education, University of Benin, Benin City, Edo State, Nigeria.

_____ _____

Dr. Praise K. Adeosun **Date**
Supervisor

_____ _____

Rev. Fr. A.A Adubale (PhD) **Date**
Head of Department

DEDICATION

This work is dedicated to the Almighty God for His grace that brought about its successful completion.

ACKNOWLEDGEMENTS

The researcher is eternally grateful to God Almighty for the successful completion of this work. The researcher also deeply appreciates her supervisor, Dr. Praise K. Adeosun for her assistance , guidance and invaluable contributions, which greatly aided the completion of this research.

The researcher is also deeply grateful to the Head of the Department of Educational Evaluation and Counselling Psychology, Rev. Fr. Dr. A.A. Adubale, for his intellectual contributions and support. The researcher‘s profound gratitude and appreciation goes to her esteemed Professors in the department - A.U. Osunde, (Mrs.) E.O. Egbochuku, . O.K. Omorogiuwa,, I Alika, (Mrs.) O.N. Aihie, for their contributions, enormous support and academic guidance. Doctors- Archbishop U.C. Ataha, (Mrs.) H.J. Osarumwense, (Mrs.) M.U. Orheruata, P.K. Adeosun, A.V. Uyigue, M.N. Igbineweka, (Mrs.) Y. O. Osunde, (Mrs.) V.O. Idusogie, W.O. Jesurobo, P.C. Ojiyi, P.E. Ima-Osagie, E.O. Omogbai, M.E. Denedo, G.I. Nwokwule, O.O. Kingsley, A.N. Idugboe, E.N. Ohiozua, E.G. Aliu, and C.M. Anyiam, and Mrs U.Atuma for their immense support, guidance and contributions.

Special thanks to Mr. Edward Agbonze and the baby of the house Eseoghene Esume for their support all through the course of the program without them this this program would not have been completed. Thanks to my mum, Mrs. Oghenerobu rioe Esume, Mrs Rose Ubogu , Curtis Apkotowho , uncle Eghworo Ovie, Mrs. Tracy Isaac and Eng. Adewale Adeyinka, for their unwavering support and prayers throughout the duration of this program. Your faith in me made this possible. May the good Lord bless each and every one of you.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

LIST OF APPENDICES

ABSTRACT

This study investigated random and systematic errors in 2024 NECO SSCE Mathematics Multiple Choice Test Items among senior secondary school three students in Ethiope East Local Government Area, Delta State. Nigeria. The study specifically assessed extends' of random and systematic errors and also the difference between observed random and systematic errors in 2024 NECO SSCE mathematics multiple choice test items administered to examinees in 2024. Three research questions guided the study.

The study adopted a survey research design on a population of 1,270 SSS III students in Ethiope East Local Government Area of Delta State.The sample size was 320 selected through random sampling techniques in two stages. The instrument used for data collection was adopted from NECO. The instrument was the official NECO standardized achievement mathematics multiple choice 60 test items with four options A-D administered and used for the in-school Senior School Certificate Examination (SSCE) in 2024. The instrument was not validated because it was standardized. For the reliability of the instrument, it was administered on twenty students twice with three weeks' time gap and subjected to Pearson Product Moment Correlation statistics to establish its dependability. The reliability coefficient of .84 was obtained. The data used for the study were collected independently by the researcher with the assistance of classroom teachers. Data were analysed using Crobach Alpha, Paired t-test and variance statistics respectively in R software.

Findings revealed that the 2024 NECO SSCE Mathematics multiple-choice items exhibited 15.6% random error, 6.09% systematic error, suggesting a small but significant level of consistent bias in measurement with random error of 90.48% substantially higher than systematic error of 9.52%. It was concluded that the 2024 NECO SSCE Mathematics multiple-choice test items were more affected by unpredictable inconsistencies than by systematic bias. It was therefore recommended that item development procedures need to be enhanced by NECO in order to improve in the aspect of item-writing, moderation, and pilot-testing processes to ensure all multiple-choice items are clear, discriminating, and free from

ambiguity that could increase random error. Also, that teachers; should adequately exposure students to standardized testing conditions and develop classroom assessments aligned with NECO standards to minimize performance fluctuations in summative assessment.

.

CHAPTER ONE

INTRODUCTION

Background to the Study

Education often entails processes to gain knowledge, skills and understanding through study, experience or learning from teachers/instructors. In its broad sense, education seems to be the process of socialization by which citizens of a nation in all geographical location learn to adapt in an effort to conquer their environment. It often encompasses the process of developing the cognitive, affective and psychomotor domains of learning of individuals and groups to equip them with skills and knowledge for survival and progress daily. Education could be recognized both as ends in its-scope and as means to achieve many other ends. Education by its nature is in the position as the real economy-sector that basically produce human resources from cradle to adulthood for national development.

Education in Nigeria has undergone changes and challenges from the pre-colonial, independence era, to the present. As one of the most populous countries in Africa, Nigeria's education system potentially highly shapes the future of her citizens through schooling and certification which often cumulated from educational assessment data. The history of education in Nigeria dates back to the colonial era, where the British introduced formal education primarily for administrative purposes which has now metamorphose beyond administrative aim to other purposes such as placement, promotion, evaluation of educational outcomes among others. The National Policy on Education (NPE) stipulates that "Education in Nigeria is an instrument "par excellence' for affective national development" spelt out in unequivocal terms the philosophy and objectives that underlie its investment in education (NPE, 2016)

The National Policy on Education emphasized free and compulsory education at the basic levels while education at the senior secondary level is free however, not compulsory. It

could be described as government investment into education which is the government and other educational stakeholders' asses from time to time using various assessment techniques such as test, questionnaire, checklist, interviews, observations among others to ascertained educational outcomes and services delivery. One of the potent assessment instrument used to assess senior secondary students is test. Test is one of the measurement instruments in the educational system used to measure students' traits or behaviour, especially achievement test. Tests can be described as structured assessments designed to evaluate knowledge, skills, and competencies in different subjects and grade levels. Test items construction often employed systematic method to write final items to measure student's understanding, abilities, or performance in particular areas or subject.

Test, as an educational measurement tool, plays a vital role in assessing students' learning, guiding instructional practices, and informing educational policies. It seems to typically involve a set of questions or tasks that students must complete under controlled conditions. The responses supplied by students on test items presented to them are scored and often quantified numerically to ascertain the total score of test takers wile students are further judge based on their total score; a process that can be described as measurement and evaluation. However, measurement often contains errors maybe due to the imprecision of educational measurement instrument, inappropriate test administration, or other factors related to test takers such as guessing, examination mal-practice or aberrant responses even when the student seems to be intelligent. For example, a student may guess an option from a multiple-choice item correctly or incorrectly which further impact total test score. Also, an intelligent student could be observed to response to test item indiscriminately or due to examination misconduct or mal-practice.

Examination mal-practice may thrive when test is taken in an un-conducive learning environment such as poor: lighting, ventilation, seating arrangement among others.

Behavioural triggers could also be responsible for random errors such as a student seeing his/her peers finishing and submitting earlier than them, thinking that everybody knows the answer except he/she. Environmental stressors such as unreliable power supply during computer based multiple choice tests could result in random errors. Hence, test takers total scores could be affected by chance errors from measurement. Measurement due to chance factors is often term random errors while errors due to technical factors can be correctly described as systematic errors.

Systematic errors usually result from technical related to item quality analysed by item parameters such as item difficulty and item discrimination indices. Also, in subjects like Mathematics, symbols could be confusing to students (such as + and * as well as \ and √). Students may overlook a symbol completely such as a minus sign skipped a question and not to come back to it due to lack of item clarity. However, it should be noted that it is difficult to completely eliminate random errors. The errors emanating from test scores are explained differently within measurement framework, often termed as measurement theories. Measurement theories in the practice of assessing of human behaviour in the educational system, is often guided by two theories known as Classical Test Theory (CTT) and item Response Theory (IRT). A common assumption to both CTT and IRT theories is the belief that each learner has and carries around some amount or level of behaviour which is often called ability or trait (Warm, as cited by Nenty, 2015). It means that, given ability or trait is a psychological (cognitive, affective, or psychomotor) property of an individual and every person possesses each of psychological traits to various degrees.

The CTT and IRT theories explain measurement errors differently. Classical Test Theory (CTT) explains measurement error as random fluctuations around a true score, assuming a single underlying construct and a single test. The implication of In the Classical Test Theory (CTT) is that, for every observed score (X) a true score (T) and an error score (E)

exist .Symbolically it is denoted mathematically as X = T+E. where X is the total (observed score), T = true score and E = error score. It means that the observed score is a combination of a true score (that is the result of an individual's ability level seen at a glance) and an error score that could be due to random or systematic factors.

Although, it seems difficult to completely eliminate random errors since random error is the error that may be attributed to unknown and uncontrollable external factors that randomly distort some observations (Ayanwale et al., 2022). For example when students guess answers on a test, their test scores may not accurately reflect their true knowledge or abilities. This guessing introduces variability that does not correspond to actual learning or understanding. Yashim, et al., (2021) explained systematic errors as consistent and repeatable inaccuracies that arise from flaws in an assessment design-tools, such as poorly constructed test items or biased scoring methods, biased questions favouring certain groups, environmental distractions, and inconsistent instructions among test-takers; errors which can lead to a consistent overestimation or underestimation of a student's true ability. Identifying and mitigating random and systematic errors is essential for ensuring fair and accurate assessments (Yashim et al., 2021). In a nutshell, errors in measurement, whether random or systematic, may be associated with measurement from test items employed in the assessment of students. Test as an instrument for measurement serve as form of data for assessment.

Assessment refers to a systematic process used to evaluate and measure students' knowledge, skills, abilities, and performance because is a critical tool for measuring students' academic performance, providing valuable insights into their understanding and skills. Assessment data often identify strengths and weaknesses, guiding instructional strategies and interventions. Assessment helps determine how well students have understood learning material and whether they have achieved the learning objectives. It provides valuable feedback to students and educators about strengths and areas for improvement, guiding

further learning and instruction. Assessments could also provide constructive feedback that informs and guides future learning of students. Assessment feedbacks may be culminated to formative or summative assessments depending on it use.

Formative is usually a term used to relate the assessment that is of a continuous process during learning to monitors student progress and provides feedback to enhance teaching strategies. It seems to be low-stake assessment, as it makes use of class quizzes and class discussions and does may not affect final grades of students, focusing rather on identifying areas needing support. In contrast, summative assessment occurs at the end of an instructional period to evaluate overall learning outcomes and effectiveness. Summative Assessment seems to be high-stakes assessment that contributes to final grades and provides a comprehensive overview of what students have learned. Assessment at the end of a programme may lead to evaluation where valued-judgement are made using some form of test items.

Evaluation tends to encompass a broader process of using the data gathered through tests in measurements of traits to make informed decisions. Educational evaluation may use the results from tests and measurements (along with knowledge of measurement error) to make critical or informed decisions or judgments regarding teaching, learning, placement, whether it will be used for diagnosing students' needs, guiding on-going instruction, or assessing final outcomes. Evaluation may be used for placement, promotion, and certification among others. Evaluation and assessment are often related because the data obtained from measurement using tests can provide vital information for decision making. Hence, summative assessment done at the end of a term or programme is important and should be free from bias due to random and systematic errors. Various examining bodies conduct examination in Nigeria using test items that could provide data for measurement, assessment

and evaluation programme such as the West African Examination Council (WAEC) and National Examination Council (NECO). However, the researcher focus is on NECO.

The National Examinations Council (NECO) is a Nigerian examination body established in 1999 to conduct credible and reliable national examinations for secondary school students. One of its major examinations is the Senior School Certificate Examination (SSCE) usually administered to graduating class at the senior secondary school level, which is used for certification and placement such as entry into higher institutions of learning. NECO has become one of the popular examination bodies since its inception in 1999 conducting examinations in various school subjects. Mathematics is a core subjects for all students, which is also one of the school subject NECO administered test items on annually. The Mathematics test items contain both essay and Multiple-Choice (MC) test items, designed to assess students' comprehension and application of mathematical concepts. The Mathematics Multiple-Choice (MC) test items are scored objective using keys for correct options while essay test items are scored using marking schemes.

Mathematics as a school subject plays a crucial role in the educational system of Nigeria, influencing various fields such as science, technology, engineering, and economics among others. The teaching and learning of mathematics are essential for developing critical thinking and problem-solving skills among students. Mathematics is the abstract science of numbers, quantity, structure, space, and change. It encompasses various fields, including arithmetic, algebra, geometry, calculus, and statistics. Mathematics is not only a body of knowledge but also a language used to describe patterns, relationships, and phenomena in the natural world. Its teaching and performance is therefore crucial to the Nigerian education system. Critical observation showed that most candidates sitting for the subject fail during the Senior School Certification Examination across the nation. Could random and systematic

errors be the cause of this failure in the summative assessment of NECO Mathematics assessment? There is need for investigation.

The number of students participating in SSCE continues to grow probably due to population surge. Hence, maintaining consistent quality in education across all regions is crucial to ensure better students' academic performance because Mathematics test scores are used for assessment, especially summative assessment.

Statement of the Problem

Summative assessment plays critical role in decision taking about students using total test score and shaping educational outcomes. NECO as an accredited examination body in Nigeria conduct summative assessment that is used to take high-stake decision on examinees. The precision of NECO test items related to random and systematic errors is vital for policy and practice in ensuring quality assurance in assessment data. However, there seems to be lack of comprehensive research focusing on random and systematic errors combined in NECO Mathematics Multiple-Choice Items (NMCTI) used for high/large scale assessments from the literature and personal experience. It appears that no comprehensive research specifically focused on random and systematic errors in NECO Mathematics Multiple-Choice Items; rather studies often focus on item psychometric properties like difficulty and discrimination. While some research identifies issues like poor item discrimination in past tests, and others use Item Response Theory (IRT) to analyze item statistics, a dedicated study analyzing the types and impact of random and systematic errors on NECO Mathematics items is currently lacking (Aborisade & Fajobi 2020).

NECO Mathematics Multiple-Choice Items (NMCTI) used for assessments has gained popularity over the years since its inception in 1999. However, Mathematics test-taking in assessment recorded less than 70% credit pass rates in national senior school certificate examination leaving more to be desired. NECO chief executive Wushishi, stated in

2025; that only 60. 60% of the total 1,805,216 candidates, who wrote the SSCE in 2024, obtained credit passed in Mathematics. This implies that 711,255 failed mathematics in 2024.. In 2022, of the 1,601,047 candidates who sat for the examination, 368, 240 students failed mathematics I,694, 637 wrote the examination in 2023 . 38.4% failed. This indicated that 650,740 failed mathematics in the examination. Wushishi further stated that, the South-western region of Nigeria led in five credits passes including. Mathematics and English with 15.48% in the standard educational performance ranking while the South region came least with 5.666%. Ethiope East Local Government Area, Delta State is located in the Southern region of Nigeria. This reveals a gradual increase in failure rate. A situation that sees about half a million people failing examination annually raises an undying question; such as what could be the extent of random and systematic errors in affecting this unacceptable failure rate in mathematics

The statistics implies that of every three students passing mathematics examination, two students are equally failing it. It is worrisome and calls for attention, or else mathematics which empowers student with critical, analytical skills as a fundamental to numerous field of study may lose popularity as a core subject in national education curriculum someday. Priority must be given to research in mathematics; causes of failure by Measurement and evaluation experts, especially in multiple choice examination test items which appears an assessment that enables students improve in their performance quicker. However, this appears not the case in mathematics. Regular exposure to multiple choice assessments ought to help students become comfortable with mathematics structure and build confidence for future assessments. The assessment familiarizes students with the format of standardized tests. This should have ensured increase in knowledge of the student. However, as revealed by NECO statistics, students' performances are dwindling by the year. Could errors be an issue in NECO summative assessment related to random and systematic errors? This study aims to

address assessment measurement errors in 2024 NECO SSCE mathematics multiple choice test items with a focus on random and systematic errors detection among students in Ethiope East Local Government Area, Delta State.

Research Questions

Three research questions were raised to guide this study.

1. To what extent does the 2024 NECO SSCE Mathematics multiple choice test items exhibit random errors among Senior Secondary School students in Ethiope East Local Government Area?
2. To what extent does the 2024 NECO SSCE Mathematics multiple choice test items exhibit systematic errors among senior secondary school students in Ethiope East Local Government Area?
3. What is the difference in the random and systematic errors of the summative assessment of 2024 NECO SSCE Mathematics multiple choice test items administered towards secondary school students examination performance in Ethiope East Local Government Area?

Purpose of the Study

The purpose of the study is to determine the assessment measurement errors In 2024 NECO SSCE Mathematics Multiple Choice Test Items in Ethiope East Local Government Area, Delta State. Specifically, the study sought to:

- examine the random errors of the summative assessment of 2024 NECO SSCE in Mathematics multiple choice test items administered to secondary school students in Ethiope East Local Government Area.
- determine the systematic errors of the summative assessment of 2024 NECO SSCE in Mathematics multiple choice test items administered to secondary school students in Ethiope East Local Government Area.

- evaluate difference between the proportion of random errors and systematic errors found in the 2024 NECO SSCE Mathematics multiple-choice test items administered to secondary school students in Ethiope East Local Government Area.

Significance of the Study

Measurement errors in the 2024 NECO SSCE Mathematics multiple-choice test items have significant consequences across various stakeholders when published. Hence, the findings of this study benefits is associated with perceived usefulness to students, teachers, future researchers, examination bodies, psychometricians, Policymakers and eventually contribute immensely to knowledge.

The result of this study will enable students to avoid anxiety and ensure motivation. Inaccurate scores can cause unnecessary stress for students. If they feel the assessment is not a fair reflection of their knowledge, their trust in the educational system and motivation to learn can decline. An example is a highly motivated student who prepares diligently for the examination but receives a poor score because of multiple flawed items. The student may become demotivated, losing trust in the education system. They might believe their hard work was for nothing, causing psychological distress and potentially discouraging future academic efforts.

The study result will be useful to teachers because it enables to use examination results devoid of errors to gauge the effectiveness of their teaching methods and identify areas where students are struggling. Errors can provide misleading feedback, causing teachers to adjust their instruction based on inaccurate data. For example the result of this research will save teachers, from viewing students' results incorrectly and conclude that the students have a fundamental weakness in that particular mathematical concept. This could have caused them to waste time and resources on unnecessary remediation, rather than addressing actual learning gaps.

The study result will be significance for future researchers from compromised validity due to the measurement errors. Errors in high-stakes test data threaten the validity of educational research that relies on it. For example a future doctoral student is planning a study on the effect of a new teaching method on student achievement, using NECO scores as their primary metric. If the NECO data is compromised by measurement errors, the student's research, no matter how rigorously designed, will be built on a faulty foundation. Their conclusions would be unreliable, and their study would likely be invalid.

Significance of the test to developers and examination bodies is ensuring quality control. Significant measurement errors signal a flaw in the test development process. It indicates that test items may be poorly constructed, ambiguous, or biased. For example, a test item may require students to find the "area of the triangle below," but the accompanying diagram is either missing or has incorrect dimensions. Examination bodies like NECO must ensure items are free from writing errors and measure what they intend to. Developers and examination bodies will benefit from the result of this study by ensuring test items consistently measure students' abilities. Errors in measurement can reduce the test's reliability, meaning a student's score might not accurately reflect their true mathematical proficiency. This can call the examination body's credibility into question.

The result of the study will be of significance to psychometricians because it will enable them to prevent significant measurement errors present as a challenge. Psychometricians are responsible for ensuring tests are fair, valid, and reliable. It highlights the need for advanced statistical models, such as Item Response Theory (IRT), to better analyze and improve test items. The study will also enable psychometricians to identification of bias. They must investigate test bias, such as Differential Item Functioning (DIF), which can occur when test items unfairly favor or disadvantage specific subgroups of students based on gender, socioeconomic status, or other factors. For example an item intended to test

algebraic skills is worded in a way that is culturally biased or uses a complex linguistic structure. The result of this study will help Psychometricians analyses to reveals that one subgroup of students consistently performs worse on this item than another, even when both groups have similar overall mathematical abilities.

Policymakers can also benefit from the result of this study because they rely on NECO SSCE results to assess the quality of mathematics education nationwide. The result of the study will help uphold trust in the system as it will ensure examination results are not compromised by measurement errors. Public trust in the country's educational standards and assessment bodies are paramount. The study will ensure significant measurement errors which can be misleading picture of students' overall mathematical achievement, potentially leading to flawed policy decisions about curriculum, teacher training, and educational funding are avoided. For example the low scores lead policymakers to conclude that specific regions or schools are underperforming. This will definitely affect educational policy decision towards such region. Finally, the importance of this study lies in its contribution to knowledge. The findings of this research will spur other researchers to carry out more study on the influence of errors NECO SSCE mathematics.

Scope and Delimitation of Study

This study cantered on random and systematic errors of 2024 NECO SSCE 60-items Multiple Choice Test. The study was delimitated to Government-owned senior secondary school three (SSS3) mathematics students in Ethiope East Local Government Area, Delta State potentially with high population. The SS3 were targeted because they are the cohort for whom the NECO SSCE examination was designed and they have completed the full senior secondary mathematics curriculum required for meaningful engagement with the test items. Thus students in SS1 and SS2 were excluded as they have not covered the

complete syllabus and may not possess the instructional readiness needed for unbiased assessment.

Definition of Terms

The following terms have been defined operationally to mean how they have been used in the study.

Assessment: this is a term that covers any activity in which evidence of learning is collected in a planned and systematic way, and is used to make a decision about performance resulted learning.

Measurement error: Measurement error "refers to the difference between a student's actual knowledge or ability and the score they obtained on a test, essentially meaning the inaccuracy in a test result as a representation of a student's true abilities

Multiple-Choice Item (Questions): It is the objective test item format which contains a problem part (*stem) and a* list of suggested solutions (alternatives*)*. The alternative also consists of one correct or best alternative (*answer*) and a number of incorrect or inferior alternatives called the.

Random error: This refers to chance difference between the observed and true values of scores obtained by a candidate.

Systematic error: this is a proportional difference consistent between an observed and true values of scores obtained in a test by students due to item bias or item quality.

Test: It is a series of standardized items to be answered by students.

Test Item: This is a single unit of the test such as that based on serial number allocated to each test item.

CHAPTER TWO

REVIEW OF RELATED LITERATURE

In this chapter, related literature was reviewed adopting the following under-listed sub-headings:

- Theoretical Framework of the Study
- Concept of Test
- Concept of Measurement Errors
- Concept of Summative Assessment
- Mathematics as a School Subject
- National Examination Council (NECO).
- Random Error in Assessment
- Systematic Error in Assessment
- Summary of the Reviewed Literature

Theoretical Framework of the Study

The current study anchors on the Classical Test Theory (CTT) propounded by Charles Spearman in 1905. The CTT also named the True Score Theory states that, the differences between the responses of examines observed scores, are made up of a true score and an error term. Classical Test Theory (CTT) holds that an examinee's observed score (X) is the sum of a true score (T) and an error term (E), expressed as $X = T + E$. The true score reflects the examinee's real ability if measurement were perfect, which never occurs because tests are susceptible to various influences. These influences break down into random error (unsystematic, chance factors) and systematic error (consistent bias from test conditions, item quality, among others.).. CTT emphasizes reducing random error so the observed score mirrors the true score more closely, and it introduces the standard error of measurement to gauge this accuracy Anikweze, 2010; Bichi, 2016; Vincent & Shanmugam, 2020).

The CTT affirmed that, the true score of an examiner captures the exact-value of the examinee's ability or attitude and written mathematically as X = T+ E in an expression to indicate association among X. T and E.

Where X = observed score

T = true score

E = error term

According to Yashim et al., (2021), collaborating the CTT, indicated that an observed score (X) is summation of a true score (T) and an error score (E) which reflects extraneous influences of a measurement process in measurement. Also, the general concern of CTT is to cope effectively with the random error portion (*E*) of the raw score. The lower the random error in a measure, the more the raw score potentially reflects the true score exemplified in CTT (AERA et.al, 2014; Shavelson, et.al., 2019) . Tests developed and improved over the years have adhered to one or another of the classical theory approaches. By and large, the tests are well developed and quite worthy at the time and effort that have gone into constructing the items that made up such test.

The model central to the Classical Test Theory, are the three concepts: observed test scores (X), which is the result of true score (T) and error score (E). True scores are the examinees' real score if there were no errors in measurement instruments; however, that is highly improbable as instruments are rarely perfect, thus the observed test scores for each individual is the outcome of the examinee's true ability influenced by error, either on higher or lower score. CTT also introduces the concept of standard error of measurement to account for how much the error has affected the reading on true scores; the larger the standard error of

measurement, the less accurate the measurement of the intended attribute, and vice versa (Magno as cited by Bichi, 2016).

The CTT is very relevant to the current study because it relates to participants observed total mathematics summative assessment score, true score and the error term in the 2024 NECO MCTI. In other words employing the CTT, summative assessment of 2024 NECO Mathematics Multiple Choice Items provides useful tools like item difficulty, discrimination, and distractor analysis, all of which help improve test quality (Bichi, 2016; Vincent & Shanmugam, 2020). However, CTT has limitations: its indices are sample and test dependent, so results may not generalize across different examinee groups or test forms. This study leverages CTT to investigate how much of the variance in the 2024 NECO Mathematics Multiple-Choice Test Items is due to random error (chance factors) and systematic error (technical flaws). By dissecting the error term, the researcher is able to estimate true ability and assess the reliability of the summative assessment.

The Concept of Test

Test is often used to connote items/tasks or series of items/tasks prepared by classroom teachers, experts or examination bodies for examinees to responds to, under pre-stated criteria's within a time bound for the purpose of measuring attributes, traits or skills, usually to generate assessment and/or evaluation data for decision taking. Ukwuije and Opara (2012) defined test as series of tasks or items used systematic to collect observations, judges to be exemplary of the curriculum of the subject being estimated.

The word "test" usually in its narrow perspective refers to a procedure in which a standard series of questions are presented, while the students supply written or oral answers (Oyetunde, 2022) defined tests as questions or task designed to or assess a representative sample of pupil or students (learners) behaviour from which the totality of that behaviour can be inferred. Test is used to observe a person's behaviour in a numerical scale or in a category-

system. A numerical scale is used when a person is described as having 60/100 performance; a category–system is used when a person is said to have passed or failed. Adom et al. (2020) explains a test to mean administering a given tool or undertaking a procedure to solicit students' responses as information, which provides the basis to make judgement or evaluation regarding some characteristics such as skills, knowledge, and values.

Tests are one of the commonly used assessment tools in educational assessment, probably because of its dependability nature. A test instrument is a standard set of items (questions) to be answered (Nworgu as cited in Omoruan, 2017). Beyond being considered as an instrument, tests can also be seen as standard procedures used to systematically measure a sample of behaviour by posing a set of questions (Linn, 2008; Oyetunde, 2022). Tests are designed to measure the quality, ability, skill or knowledge of a sample against a given standard, which usually could be deemed as acceptable or not. In educational practice, tests are methods used to determine the students' ability to complete certain tasks or demonstrate mastery of a skill or knowledge of content. Tests can take the form of multiple choices or a weekly spelling (Adom et al., 2020; Braun et al 2006; Imasuen, & Adeosun,. 2023). Tests can take the form of essay, objectives such as multiple choices questions or a weekly spelling (Adom et al., 2020). Similarly, Braun, et al., (2006) stated that, testing is the process of measuring single or multiple concepts, under a set of pre-determined conditions.

Testing is used to measure the level of students' learning. In other words a test words, a test is a set of questions which examinees are expected to respond to. It is based on the response provided by the examinees from the series of questions that teacher obtains data for measurement and evaluation of examinees. Tests assist the teacher to know the extent of the learners' mastery of the objectives in the subject or programme. Asuru (2015) defined test as a set of tasks, questions, situations, intended to elicit particular types of behaviour; it is an evaluation instrument used to measure skills, knowledge, intelligence, ability, aptitude,

attitudes, interests, attention span, motivation, competence, and the like, hence it could take any form. Achievement test as an ability test that is designed to appraise what the individual has learned to do as a result of planned previous experience or training often provided in school (Obilor & Obubere, 2020). It is specifically used to find out the extent to which a testee has achieved, gained or mastered certain information or skill after he has been exposed to some training. The teacher is able to know the students' progress in a subject and whether or not the stated objectives have been realized through achievement tes

The word "test" usually in its narrow perspective refers to a procedure in which a standard series of questions are presented, while the students supply written or oral answers (Oyetunde, 2022) defined tests as questions or task designed to or assess a representative sample of pupil or students (learners) behaviour from which the totality of that behaviour can be inferred. Test is used to observe a person's behaviour in a numerical scale or in a category-system. A numerical scale is used when a person is described as having 60/100 performance; a category–system is used when a person is said to have passed or failed. Adom et al. (2020) explains a test to mean administering a given tool or undertaking a procedure to solicit students' responses as information, which provides the basis to make judgement or evaluation regarding some characteristics such as skills, knowledge, and values.

Tests are one of the commonly used assessment tools in educational assessment, probably because of its dependability nature. A test instrument is a standard set of items (questions) to be answered (Nworgu as cited in Omoruan, 2017). Beyond being considered as an instrument, tests can also be seen as standard procedures used to systematically measure a sample of behaviour by posing a set of questions (Linn, 2008; Oyetunde, 2022). Tests are designed to measure the quality, ability, skill or knowledge of a sample against a given standard, which usually could be deemed as acceptable or not. In educational practice, tests are methods used to determine the students' ability to complete certain tasks or demonstrate

mastery of a skill or knowledge of content. Tests can take the form of multiple choices or a weekly spelling (Adom et al., 2020; Braun et al 2006; Imasuen, & Adeosun,. 2023). Tests can take the form of essay, objectives such as multiple choices questions or a weekly spelling (Adom et al., 2020). Similarly, Braun, et al., (2006) stated that, testing is the process of measuring single or multiple concepts, under a set of pre-determined conditions.

Testing is used to measure the level of students' learning. In other words a test words, a test is a set of questions which examinees are expected to respond to. It is based on the response provided by the examinees from the series of questions that teacher obtains data for measurement and evaluation of examinees. Tests assist the teacher to know the extent of the learners' mastery of the objectives in the subject or programme. Asuru (2015) defined test as a set of tasks, questions, situations, intended to elicit particular types of behaviour; it is an evaluation instrument used to measure skills, knowledge, intelligence, ability, aptitude, attitudes, interests, attention span, motivation, competence, and the like, hence it could take any form. Achievement test as an ability test that is designed to appraise what the individual has learned to do as a result of planned previous experience or training often provided in school (Obilor & Obubere, 2020). It is specifically used to find out the extent to which a testee has achieved, gained or mastered certain information or skill after he has been exposed to some training. The teacher is able to know the students' progress in a subject and whether or not the stated objectives have been realized through achievement test.

Achievement which can be essay or objective formats is an instrument given at the end of teaching/learning programme (Iwuji, 2007). The author explained that an achievement test is a test used to find out how much testees are able to achieve in a course he/she has been taught. Achievement test according to are designed to evaluate past and present learning by measuring the progress which students have made as a result of instruction or training (Ipaye, as cited by Osadebe, 2016). Thus, the teacher has the responsibility of assessing how much

learning or change in behaviour that has occurred as a result of a programme of instruction as summative assessment. One of the important type of test used for summative assessment of students which can capture most of the subject matter-content taught is the objective test or multiple choice test.

Multiple Choice Tests (MCTs) are tests that require examinees to answer a question or complete an incomplete statement by selecting the correct option from a list of suggested answers or alternatives. A multiple-choice test consists of a stem and two or more options, with the examinee expected to pick the single correct option; items are scored dichotomously as correct or incorrect (Hohensinn & Kubinger, 2009; Paniagua & Swygert, 2016). Because MCTs can be scored objectively—by humans or machines—they are suitable for investigating random and systematic errors in summative assessments for research purposes, eliminating scorer bias that may occur with essay tests. Amedahe and Asamoah-Gyimah (2016) noted that multiple-choice tests are widely used for objective measurement of examinees' traits.

Multiple-choice tests are popular due to several advantages. They can serve as diagnostic tools, support formative assessment, and evaluate a broad range of knowledge. They are also versatile, measuring not only recall of facts but also higher-order learning outcomes (Díaz & Verdejo-Carrión, 2020; Hicks, 2014; Koepf, 2018). However, issues of reliability and validity are linked to instrument precision.

Test Reliability

Reliability refers to the degree to which results from a measurement procedure can be replicated. Although reliability contributes to validity, it is not sufficient on its own. Lack of reliability may arise from divergence between observers or instruments, or from instability of the attribute being measured, all of which affect validity. Reliability indicates consistency of scores across time, raters, and items (Bolarinwa, 2015; Wong et al., 2012). It is often

described in terms of equivalence, stability, and internal consistency (homogeneity). Reliability is the consistency of scores or information obtained when an instrument is applied at different times. It relates to the precision of scores, minimizing errors caused by changes in administration conditions, scorer subjectivity, item ambiguity, or lack of motivation. Increasing the number of items in an objective test usually raises the reliability coefficient (Majid et al., 2020). A reliable instrument is free from bias and ensures stable measurement across time and items. Some qualitative researchers use "dependability" instead of reliability to convey stability and freedom from error.

Haradhan (2017) defines reliability as a measurement that yields consistent results with equal values. It measures consistency, precision, repeatability, and trustworthiness (Chakrabartty, 2013). Thus, an observed score should reflect the true score, making reliability a necessary but not sufficient component of validity. In quantitative research, reliability means consistency, stability, and repeatability of results. In qualitative research, it refers to the consistency of an approach across different researchers and projects. Reliability is the extent to which a questionnaire, test, observation, or any measurement procedure produces the same results on repeated trials. In other words, the stability or consistency of scores over time or across ratters and reliability applies to scores, not people.

Reliability can be divided into two main types: stability and internal consistency. Stability refers to the ability of a measure to remain unchanged over time despite uncontrolled testing conditions or changes in respondents. It assesses how much a person's score can be expected to stay the same from one administration to the next. Methods to test stability include test-retest reliability and parallel-form reliability. There are three common reliability coefficients: Stability (test-retest); Consistency of scores over time; Equivalence: Whether two or more parallel forms of an instrument produce similar results and Internal consistency: The cohesion of responses to items meant to measure the same construct. If an

instrument produces reliable scores, they should be similar for the same group of students across administrations. The correlation between scores is the stability coefficient (Bolarinwa, 2022; Miller et al., 2015). Since double administration is often impractical, internal consistency (computed from a single administration) is commonly used especially in Objective tests having a scoring key, eliminating subjectivity. For large groups, computer software can calculate a reliability coefficient (often internal consistency). The standard error of measurement can also be estimated, indicating the accuracy of individual scores; higher reliability reduces measurement error.

Test-Retest Reliability (Stability)

Test-retest correlation provides an indication of stability over time (Pedisic et al., 2014). It occurs when the same or similar scores are obtained on repeated testing with the same respondents (Wong et al., 2012; Deniz & Alsaffar, 2013). Scores are thus consistent from one time point to another. This method assumes that the measured characteristic does not change between tests (no testing effect) and that the time interval is short enough to avoid memory effects but long enough to avoid recall bias (Engel, 2013). Correlation coefficients $(r) \geq 0.70$ are generally considered good (Singh et al., 2011). If observations are recorded, intra-observer reliability can be computed by having the same observer make two measurements.

Alternate-Form Reliability (Equivalence)

Alternate-form reliability refers to the degree of agreement between two or more instruments (e.g., different questionnaires) designed to measure the same construct and administered at nearly the same time. It is tested through a parallel-forms procedure, where alternative versions of the same measure are given to either the same group or different groups of respondents. The items are reworded or reordered to create two similar but not identical forms. These forms can be administered simultaneously or with a brief delay. A higher correlation between the forms indicates greater equivalence. However, because it is

difficult to confirm that two tests are truly parallel, this method is rarely used in practice (Bolarinwa, 2022).

Internal Consistency Reliability (Homogeneity)

Internal consistency evaluates the extent to which all items on a test measure the same construct. An advantage of internal consistency is that it can be estimated after a single test administration, avoiding problems associated with repeated testing (Wong et al., 2012). Common indices include split-half reliability, coefficient alpha, and Kuder-Richardson formula 20 (KR-20). Split-half reliability divides the test into two halves (e.g., odd vs. even items), administers both halves to the same group, and correlates the scores. Coefficient alpha (Cronbach's alpha) and KR-20 represent the average of all possible split-half estimates. Coefficient alpha is used for items with multiple response options (e.g., Likert scales), whereas KR-20 is appropriate for dichotomous items (yes/no, true/false) (Bolarinwa, 2022).

Concept of Measurement Errors

Measurement errors in educational assessment refers to the discrepancy between a student's observed score and their true score on a given construct or trait being measured. It represents the difference between the actual (potentially inaccurate or biased) data yielded by the assessment and the ideal, error-free information that accurately reflects the learner's true ability, knowledge, or characteristic.

In simplified terms, measurement error can be expressed as:

x = xo – xt

Where:

- **x** is the measurement error,
- **xo** is the observed (possibly distorted) score obtained from the assessment, and
- **xt** is the true score, which represents the student's actual level of performance or ability, free from error.

This framework reflects the principle that assessments in education do not perfectly measure the target trait (e.g., intelligence, reading ability, or test anxiety), and that every score contains some level of error, whether from the test-taker, the instrument, or the testing environment (Bassey, et al., 2021; Clapper, et al., 2005 Esomonu & Okeaba, 2021). It calls for systematic error in measurement for ethical implications in learners' assessment in the educational system. In this context, the most accurate estimate of a test-taker's true score is the average of their observed scores obtained across multiple testing occasions. The variation around this average reflects the theoretical concept known as error variance.

Measurement error may arise in two main forms: systematic error, which is associated with issues of construct validity, and random error, which pertains to the reliability of the assessment. While efforts can be made to reduce it, random error can never be fully eliminated (Yu, 2008). Yashim et al., (2021) characterized measurement errors in educational assessment sources of both systematic and unsystematic errors that are prevalent in educational assessment related to students test misconduct of any kind and error associated with teachers such as items development, test administration and scoring process among others. Yashim et al., further stated that, errors can be controlled to some extent by standardizing the measurement procedures and increasing the number of observations that are sampled for each person hence; the score will be formed by averaging over a sample of observations,

Random Error of Measurement

Random error is the discrepancy that results from unknown and uncontrollable external factors that affect some observations differently from others. For example, during data collection, some participants may be in a more positive mood, causing them to give more favourable responses to constructs such as self-esteem, satisfaction, or happiness, while those in a poorer mood may answer less favourably. Because it is impossible to predict or

control each person's mood, this variability is treated as random noise. Similarly, at an organizational level, sudden regulatory changes or environmental fluctuations might impact the performance of some firms but not others Yashim et al., (2021. Consequently, random error is usually regarded as "noise" and commonly disregarded in measurement.

Factors Responsible for Random Errors

Test-Taker Behaviour can introduce random error. Factors such as fatigue, anxiety, stress, illness, or simple carelessness may cause a student to misread a question or make a mistake, even when the content is understood (Alruwais, 2020). Guessing also contributes to random error. In high-stakes multiple-choice exams like NECO, students may guess, especially when pressed for time. Such guessing adds variance to scores that does not reflect true ability (Bai & Wang, 2018). Environmental Conditions often create random fluctuations. Distractions such as noise, lighting, seating arrangement, or room temperature can affect performance unpredictably (Ogunlade & Adebayo, 2019). Administration Inconsistency may further produce random variation. Differences in how invigilators enforce rules or deliver instructions can impact students unevenly (Idowu et al., 2020).*Management of Random Errors in Summative Assessment*

Management of Random Errors in Summative Assessment can be done by key roles such as ensuring standardize testing environment, that provide test centers uniform, conducive environments to help reduce the influence of external disruptions (Arowolo & Abiodun, 2018). Clear Instructions should be provided. Misunderstanding instructions can lead to random errors. Clearly communicated and uniformly administered instructions help to reduce this issue (Ajayi & Fakorede, 2020). Item Analysis should be made use of. Identifying items with unusual response patterns (low discrimination or unusual difficulty) helps flag questions prone to guessing or misinterpretation (Daramola & Akinkuotu, 2022). Pre-test or Pilot Items should be employed. Testing items on a sample group prior to final administration

can help detect erratic performance and minimize unforeseen issues (Odekunle & Obilor, 2019).

Systematic Error of Measurement

Systematic errors in educational measurement refer to consistent and repeatable inaccuracies that arise from flaws in the assessment process, test design, administration, scoring, or interpretation. These errors introduce bias that can lead to unfair or invalid conclusions about student performance. Unlike random errors, which are unpredictable and tend to cancel out over a large sample, systematic errors can result in significant consequences (Twycross & Shields, 2004 cited in Haradhan (2017). Systematic errors may occur as a result of test bias, cultural bias, poorly worded questions, inadequate of test construction procedures and selection. To reduce systematic error cultural background and experience should be consider, test items should be well validated and free from bias. Assessment scores are influence by systematic errors related to test items distortion, administered by examination bodies. Moreover, measurement from test items cumulate into data that are often used for formative or summative assessment, hence test items are to be evaluated for error(s) evidence for the purpose of quality assurance in assessment data. .

Factors Responsible for Systematic Errors

Factors related to item quality, bias or Differential Item Functioning (DIF) could be responsible for system errors in measurement, and by extension influencing summative assessment data (Atuma, 2014; Omole & Onuka, 2021). Some test items may unfairly be of advantage or disadvantage to certain groups, such as urban verses. rural students, due to differences in language, context, or prior exposure (Omole & Onuka, 2021). Curriculum Misalignment also causes systematic error. The NECO Mathematics items may not fully reflect what is taught in schools (or overemphasize certain topics) if that should be the case,

then her test items creates systematic disadvantage for certain students (Okonkwo & Ilogu, 2020).

Poorly constructed test items such as questions that are ambiguous, poorly worded, or culturally unfamiliar can consistently confuse or mislead students, especially those from disadvantaged backgrounds (Abah, 2019). Location-based educational inequities such as differences in infrastructure, teacher quality, and access to resources between urban and rural schools can result in systematic performance differences, even when ability is equal (Ugwoke et al., 2022). When teachers in urban areas are consistently of higher quality than those in rural areas due to rural urban migration, there will be a consistent inequality in student learning outcomes. Also scoring and keying errors also brings about consistent error during assessment, even in objective tests like multiple-choice items; systematic scoring errors can occur if the answer key is flawed or incorrectly uploaded into an automated system (Ojo & Adewale, 2018).

Management of Systematic Errors

The management of systematic errors can be through item review and experts validation. Items should be reviewed by subject matter experts to ensure fairness, clarity, and cultural neutrality (Omole & Onuka, 2021). Use of a table of specification (TOS) should be enforced. Aligning test items with the curriculum ensures proportional representation of topics and cognitive levels (Okonkwo & Ilogu, 2020) stated that Differential Item Functioning (DIF) Analysis helps to identify items that behave differently for subgroups with similar overall ability (Adekunle & Fasasi, 2023). There should also be pilot testing across school types. Test items should be piloted in both urban and rural settings to detect potential location-based biases (Abah, 2019). Application of Item Response Theory (IRT) is also recommended. IRT helps to identify biased or mishitting items and provides more precise measurement compared to Classical Test Theory (Bai & Wang, 2018) in assessment.

Assessment in the world of education connotes systematic process of gathering, evaluating, and interpreting data to inform stakeholders' a learner's knowledge, abilities, attitudes, and performance. Simply put, assessment enables educators to understand what pupils / students have learned, how effectively they have learned it, and what needs to be improved in both teaching and learning. In education, Assessment is a vital component that helps shape students' learning experiences. Teachers use a range of evaluation methods to assess students' knowledge, skills, and overall development. Among the most common approaches are formative and summative assessments, each with distinct roles in the educational process. Although both formative and summative assessments are important for measuring student learning, they differ notably in their purpose, timing, and the results they aim to achieve**.** According to Ismail et al, (2022) summative assessment impact academic through motivation, attitude toward learning, test anxiety and self-regulation skill in learning because of the role assessment plays in teaching and learning.

Concept of Summative Assessment

Summative assessment in measurement refers to a form of evaluation conducted at the conclusion of educational programme or an instructional timeline, such as : the end of a term, test administration to measure academic progress Its main goal is to assess the extent to which students have achieved the intended learning outcomes. These assessments are typically formal, graded, and carry substantial weight in determining a student's final academic standing. Summative assessment could be in form of standardized exams like WAEC, NECO, and JAMB among others. An effective summative assessment brings about mastery of key learning objectives on students feedback and showcase their ability to integrate course concepts into a cohesive understanding.

Summative assessments should align with course goals, build on prior formative assessments, and provide rich feedbacks. Summative assessment evaluates student learning at

the end of a course, focusing on the final product and determining grades and progression, unlike formative assessment which monitors progress throughout the learning process. Overall learning goals are established according to the relevant curriculum standards. According to Lane (2018), summative assessments are typically designed with standardized criteria, to ensure they can be administered consistently across large groups of students, multiple cohorts, at different time frame. The data gathered from these summative assessments whether at the individual student, cohort, school, or system level provides educators and administrators with valuable insights to evaluate student knowledge in relation to intended learning objectives. Additionally, this data enables comparisons between current and previous cohorts, as well as across different schools.

Blankman (2024) highlighted the value of mathematics summative assessment as a beneficial tool. Summative assessment enables teachers to gauge learning and assign grades at the classroom level. When applied rigorously and thoughtfully, it can also measure the overall performance of a district, state, or even an entire country. Similarly, Ngunjiri (2022) explained that assessment in Mathematics classrooms reviews the role of assessment in bringing changes in students' mathematics performance. Also, that item quality assessment is a key factor in improving the learning of mathematics as the relationship between learning and summative assessment is strong and robust.

Mathematics as a School Subject

Mathematics as a school subject is a fundamental subject that deals with numbers, quantities, and shapes. It involves logical reasoning, problem-solving, and critical thinking to understand and analyze various mathematical concepts. Applications of knowledge expand beyond the science-based subject to other fields of study. According to Bekeking (2003), Mathematics has numerous applications in various fields. It is applied in science and technology where it is used to describe laws of physics, model population growth, and

understand complex systems. In economics and finance, mathematical models are used to predict market trends, manage risks, and make informed investment decisions. The world of Engineering, mathematics is used to design, develop, and optimize systems, structures, and processes.

The benefits of studying mathematics equipped individuals to develops problem-solving skill**s** such as critical thinking, analytical skills, and problem-solving strategies. It also Improves logical reasoning which potentially enhances logical reasoning and analytical skills. Mathematics is a fundamental to other fields of study putting it career opportunity enhancer ranging in sciences, social sciences, humanities among others. Mathematics as a subject is examined for certification for senior secondary graduating class generally described as summative assessment by WACE, JAMB and NECO and other examination bodies in a large scale assessment. NECO as an examination introduced amidst the existence of WACE have gained popularity over the years. According to Wushishi, (2025) it NECO is widely accepted as entry qualification into tertiary institutions.

National Examination Council

The National Examination Council (NECO) is a Nigerian examination body established in 1999 to conduct senior secondary certificate examinations (SSCE) and general certificate in education (GCE) exams. NECO's primary functions include conducts Senior School Certificate Examination (SSCE) and General Certificate Examination (GCE) examination for senior secondary school students (Esomonu & Okeaba, 2021) . She certifies candidates: who pass the exams providing them with certificates. Online services are provided by NECO through online registration and result checking for candidates making NECO viable in a technology driven world.

Achievements recorded by NECO has achieved include: increased accessibility**,**

Improved examination process as implementation measures to improve her examination process, technology to enhance efficiency and accuracy and percentage increased pass rate (Wushishi, 2025). In 2024, 60.60% of candidates obtained credits and above in mathematics. NECO plays a crucial role in promoting mathematics education, develops examination questions that test candidates**.** NECO develops examination questions that test candidates' understanding on mathematical concepts and computations. The provision of feedback to candidates, teachers, schools and the general public on performance in mathematics examination is a milestone gain because it supplies data for research.

Random Error in Assessment

Random errors in educational assessment refer to unpredictable fluctuations that occur during the assessment process, impacting the reliability of test scores. Random errors are caused by temporary factors like student fatigue, distractions, misinterpretation of questions, or random guessing, particularly in multiple-choice test items(Adeosun &Ukana, 2024; Blankman, 2024). In summative assessments, where tests are used to evaluate cumulative learning outcomes, random errors pose a significant threat to reliability; because assessments inform high-stakes decisions, such as graduation or certification. It implies that, score fluctuations due to random error can misrepresent a learner's true competence.

Multiple-choice items, commonly used in mathematics assessments, are prone to random errors due to guessing or misinterpretation. Environmental inconsistencies, such as noise or poor lighting, can also introduce errors. In mathematics multiple-choice items, random errors can arise from misreading symbols, overlooking units, or encountering test environmental distraction such as noise. In such a case, random errors interfered with total score whereby, actual mathematical ability will not be reflected due to the conditions under which students respond to test items. Understanding random errors is crucial, because chance errors can mask students' true proficiency, leading to misclassification and poor instructional

decisions (Thorndike & Thorndike-Christ, 2010). This highlights the importance of minimizing errors to ensure accurate summative assessments data. Ajayi and Fakorede (2020) stated that the impact of test administration practices on students' performance in standardised assessments inconsistencies in administration related to timing, supervision, and noise introduced random fluctuations in students' performance in WAEC and NECO examinations which in turn impact the reliability of the examination outcomes. Flaws like guessing contribute to random errors (Alruwais. 2020; Daramola & Akinkuotu, 2022).

Achor et al. (2024) investigated the impact of teachers' assessment practices on students' acquisition of process skills and performance in Basic Science. The study employed a cross-sectional survey design, with a population of 20,885 participants. A sample of 393 was selected through a multi-stage sampling procedure. Data were collected using the Basic Science Teachers' Assessment Practice Observation Scale (BSTAPOS), the Basic Science Process Skill Acquisition Test (BSPSAT), and the Basic Science Performance Test (BSPT). ANOVA and an independent-samples t-test were used for analysis. The findings revealed that teachers' assessment methods significantly influenced students' acquisition of process skills and their performance in Basic Science.

Imasuen and Adeosun (2023) investigated the application of Generalizability Theory on measurement errors in the 2019 WAEC Mathematics objective examination for senior secondary students.. The study utilized a survey design. The study population was 5,697 while the sample size was 570, selected using multi stage sampling procedure. The instrument used for data collection was the 2019 WAEC Mathematics multiple-choice test items, which had already been validated by WAEC. Data were analyzed with EduG software (version 6.0-e) using ANOVA and generalizability analysis. The results showed that the student-teacher interaction accounted for the largest proportion of measurement error (68.9%), followed by the student effect (27.5%), and the residual interaction among all facets (3.6%).

A high generalizability coefficient of 0.97 was achieved when the number of teachers was increased to 78, indicating strong reliability for ranking students based on performance.

Adeosun and Ukanah (2024) estimated the standard error of measurement for the 2022 Basic Science Multiple-Choice Test Items of the BECE using bootstrap methods. The study employed a descriptive research design. The population comprised 3,504 Junior Secondary School 3 (JSS3) students from 14 public and 20 private schools in Edo State, while the sample consisted of 350 students selected through a multistage sampling procedure that combined stratified sampling, simple random sampling, and proportional allocation. Data were collected with a test instrument containing sixty (60) multiple-choice items taken from the 2022 BECE Basic Science and Technology examination. Bootstrap statistics and the bootstrap t-test for independent samples were used for analysis. The results revealed a moderate level of variation in test scores between students from public and private schools on the 2022 BECE Basic

Ohiozua and Omorogiuwa (2025) conducted a study on the determination of aberrant examinee's responses of 2018 and 2019 May/June Biology objective questions of National Business and Technical Certificate Examination in Nigeria. The study utilized survey research design of expost facto on a population of 57,357 (28121 for 2018 and 29254 for 2019) students scored responses. The sample size for the study was 13,533 comprising of 7086 and 6447 for 2018 and 2019 students scored responses respectively, selected through stratified, random, cluster sampling techniques step-wisely in a multistage procedure. The instrument for data collection were the 2018 and 2019 May/June NABTCE 50 items Biology multiple choice test questions. Data analysis was done using Perfit software, a non - parametric Item Response Theory (IRT) statistics to determine aberrant examinees' responses. Findings showed that, the number of abnormal or unusual responses from test-takers in the 2018 and 2019 NABTCE Biology multiple-choice exams was at a moderate level.

Gabriel Bazimaziki et al. (2019) conducted a corpus-based study titled "Analysis of Students' Errors in Summative Evaluation," focusing on undergraduate students' errors in a literary subject taught in English. The research employed a qualitative design. Three research questions guided the study: (1) what errors do students make in their final evaluation? (2) What effects do these errors have? (3) How can students be helped to reduce these errors? The study uncovered three key implications: it highlighted for language teachers the areas that need greater emphasis; it showed students the importance of addressing these challenges to improve language use and achieve higher proficiency; and it suggested that consistent effort and commitment can help learners correct such errors in assessments.

Denedo (2023) estimated item information function of computer science achievement test among primary school pupils in two Universities staff schools in Benin and Port-Harcourt Cities. The study adopted the descriptive research design with a sample size was 443 primary six school pupils. The instrument for data collection was computer science achievement test consisting of 52 multiple choice items. The data was analyzed using the Rasch model with Winstep 4.8.2. The finding revealed that CSAT was able to provide satisfactory information on measurement errors.

Ismail et al. (2022) examined the effects of formative versus summative assessment on academic motivation, attitude toward learning, test anxiety, and self-regulation skills in Iran. The study employed an experimental research design with a sample of 72 Iranian EFL learners selected through convenience sampling. Participants were assigned to two experimental groups (a summative group and a formative group) and a control group. All three groups completed pre-tests to measure test anxiety, motivation, and self-regulation skills. Afterward, one experimental group received instruction based on formative assessment principles, while the other experimental group was taught using summative assessment methods.

The control group received instruction without any pre-planned assessment. Following a 15-session intervention, post-tests for test anxiety, motivation, and self-regulation skills were administered to all groups to evaluate the effects of the instruction on language achievement. Finally, an attitude questionnaire was given to both experimental groups to examine their perceptions of the impact of formative and summative assessment on English learning improvement. Data were analyzed using Bonferroni tests, ANOVA, and a one-sample t-test. The results showed that formative assessment was more effective in enhancing academic motivation, reducing test anxiety, and improving self-regulation skills. Participants also expressed positive attitudes toward formative assessment.

Olela et al. (2021) examined how summative evaluation affects classroom practices in Rarieda Sub-County, Kenya, grounding the study in extrinsic and intrinsic motivation theories. A descriptive design with a correlational approach was employed. Stratified and simple random sampling was used to select 250 teachers for the sample. Data collection involved a questionnaire and focus-group discussions. The data were analyzed descriptively, thematically, and with Pearson Product-Moment Correlation. Results indicated that teachers had favorable perceptions of both classroom summative evaluation and national examinations. Moreover, significant correlations emerged between teachers' perceptions of classroom summative evaluation and classroom management, syllabus completion, and between perceptions of national examinations and lesson preparation.

Systematic Error in Assessment

Systematic errors in educational assessment are consistent, repeatable inaccuracies that occur due to flaws in test design, administration procedures, or content bias, which affect test scores in a predictable manner. Unlike random errors, systematic errors have a directional influence consistently increasing or lowering examinees scores based on non-ability factors such as language, socio-economic status, or cultural background (Adekunle &Fasasi, 2023;

Adeosun and Imasuen, 2023 Okonkwo and Ilogu 2020). In mathematics multiple-choice assessments, systematic errors can arise from encountering of poorly worded items or item with inappropriate difficulty index. Systematic errors either have an advantage or disadvantage for certain group of students not due to difference in ability but due to error from the test instrument or test items.

According to Nitko and Brookhart (2014), systematic errors arise from identifiable sources such as ambiguous items, misaligned content, or culturally biased test language. bias, inequitable access to instruction, or misalignment between curriculum and assessment. For instance, questions that require context unfamiliar to rural student, such as banking or technology may be of disadvantage to the concerned students. Systematic errors pose a serious threat to the validity of summative assessments, consistently favoring or hindering performance based on student unique demographics characteristics and socio-economic status, school location, or language-proficiency. Systematic errors are analyzed to assess how predictable and recurring flaws may have impacted student performance. As Thorndike and Thorndike-Christ (2010) explained; systematic errors compromised the validity of summative assessments by ensuring test scores to reflect something other than the intended construct.

Thus, understanding of measurement errors sets the stage for an integrated analysis of how both random and systematic errors interact to affect the fairness, reliability, and interpretability of results in large-scale assessments like that of NECO Mathematics. By actively addressing sources of measurement error, educational institutions can strengthen the credibility, equity, and effectiveness of summative assessments, ensuring that all students are evaluated based on their true abilities related to observe scores in summative assessments.

Batubo, and Ambrose (2022), conducted a study on analyses of reliability and the content validity of basic education certificate examination mathematics objective test items of 2018 to 2020. The study was carried out using descriptive survey design. The population for

the study was 19,056 students' scripts while the sample size was 1,500 students' scripts randomly selected using multistage sampling procedure. Past questions, and curriculum of Mathematics of JSS1 to JSS3 were used for data collection. Analysis of data was done using R Software to calculate the reliability coefficient and content validity. The findings revealed that objective test items content area were not adequately covered.

Essien, et al (2024), assessed mathematics multiple choice items using Item Response Theory Model on upper basic 9 students in Akwa Ibom State. The research design used for this study was the survey design on a population of 69,100 answered scripts of BECE 2021 mathematics. The sample size was 3,456 answered scripts selected through multistage sampling procedure. The instrument used for data collection was the BECE mathematics objective test items. The data was analyzed using Bilog-MG3 statistical software. The findings from the studied showed that the test items failed to differentiate abilities of test-takers accurately.

Ebuoh and Okafor (2017) investigated the effects of systematic, random and mistake (control) measurement error on SSS 2 on students' achievement in Biology tests in Enugu State. The study employed a non-equivalent quasi-experimental design. The sample size was 378 Senior Secondary School 2 (SSS 11) randomly selected. Each group was assessed based on a different type of measurement errors: systematic error for Experimental Group I, random error for Group II, and mistake error for controlled Group III. Data were generated using a Biology Achievement Test (BAT). The Analysis of Covariance (ANCOVA) was used for data analysis. The findings revealed that students evaluated with systematic error testing in Biology performed significantly better than those assessed using random error tests. The control group, associated with mistakes, recorded the lowest achievement levels in Biology.

Summary of Reviewed Literature

The Classical Test Theory (CTT), introduced by Charles Spearman in 1905 is the theoretical ground on which the study hinged. The CTT is based on the explanation that a respondent true test score comprised errors and true score assumptions exemplifies the differences between the responses of examines error term could be random or systematic. This framework highlights the foundation for measurement theory for over 80 years and best suited for traditional testing situations either in group or individual settings in which all members of the target population is administered the same sets of test and concepts of: test, measurement errors related to validity and reliability, summative assessment and mathematics as a school subject and NECO. NECO is a significant national examination body involved in the conduction of examination for senior secondary schools in over 12 subjects including mathematics.

The literature reviewed indicated that, the influence of random and systematic errors on summative assessment in mathematics multiple choice items must be given credence in students learning and its outcome, because it has bearing on actual examination procedure and impact observed score of students. The increasing adoption of CTT in judging total test scores highlights the need for educational stakeholders to ensure that random and systematic errors are reduced to the barest minimum, particularly in mathematics multiple choice items. The literature indicated that, treat to test item validity and reliability are embedded in random and systematic errors from measurement.

Studies have investigated measurement errors in aspect deviated from random and summative assessment of 2024 NECO multiple choice test items. Ayuba et al., (2021) viewed random and systematic errors from student misconduct and teachers test scoring point of view. Imasuen and Adeosun (2023) also measurement errors from teacher - student view point using generalizability theory. Ohiozua and Omorogiuwa (2025) examined measurement errors from aberrant examinees response using National Business and

Technical Certificate Examination biology multiple choice items. However, this study is investigating random and systematic errors on summative assessment of NECO 2024 mathematics multiple choice assessment in Delta State from the view point of external factors (random errors) and test quality (systematic errors). This constitutes a gap which this study is intended to fill.

CHAPTER THREE

METHODOLOGY

In this chapter, the method and procedure that will be used in this study is discussed under the following subheadings:

- Design of the Study
- Population of the Study
- Sample and Sampling Techniques
- Research Instrument
- Validity of the Instrument
- Reliability of the Instrument
- Method of Data Collection
- Method of Data Analysis.

Design of the Study

The study employed the Survey research design. Survey research design is preferred because only a representative sampled was studied to make generalization about the entire population in this study. Also, the Survey research design enables the collection of standardized data from a large group of students under uniform conditions, thereby ensuring comparability of responses across all test items. This uniformity is essential to accurately examine how students respond to each multiple-choice item and for identifying consistent pattern in their response.

Population of the Study

The population for this study was 1,270 SSS III students in Ethiope East Local Government Area of Delta State. See Appendix 1 for the population of students and number of schools according to the local government areas in Delta State.

Sample and Sampling Technique

The sample size was 320 students selected through random sampling techniques in two stages. In the first stage eight (08**)** schools were selected through balloting by replacement. In the second stage, 40 students were selected at random from the eight schools previously selected in stage one from senior secondary schools three student in Ethiope East Local Government Area, Delta State using multistage sampling procedure in three stages. The sample size dependability was 279 based on Krecial and Morgan (1970) Table for determining reliable size.

Research Instrument

The instrument that was used for this study was an adopted official NECO 2024 Mathematics multiple choice questions for the Senior School Certificate Examination (SSCE). The 2024 Senior School Certificate Examination contains 60 items with four options A-D.

The instrument measured the mathematics scheme of work from the first year through the third year of senior secondary school. The test items are presented in Appendix II.

Validity of the Research Instrument

The multiple choice items constructed were validated and standardized by the National Examination Council before it was administered to students. Hence, the instrument was not subjected to further validation by the researcher.

Reliability of the Research Instrument

The reliability of the instrument was established by administer the test items to 20 senior secondary school student using test-retest method. It yielded a reliability coefficient of .84

Method of Data Collection

The instrument was administered by the researcher and with the assistance of two trained research assistant in each of the sampled schools. The test was administered within the stipulated time limit of 90 minutes original allotted by NECO. Thereafter, answered scripts were retrieved for scoring and collation by the researcher.

Method of Data Analysis

The responses of examinees were scored 1 for correct option and 0 for incorrect option using a scoring key. Thereafter, research question one, two and three were answered using descriptive statistics, paired t-test statistics and variance respectively in R software.

CHAPTER FOUR

PRESENTATION OF RESULTS AND DISCUSSION OF FINDINGS

In this chapter, presentations of results of data collected from fieldwork and discussion of findings were logically done. The results are presented using tables in line with research questions raised in chapter one to achieve the aims of the study.

Presentation of Results

Research Question One: To what extent does the 2024 NECO SSCE Mathematics multiple choice test items exhibit random errors among senior secondary school students in Ethiope East Local Government Area?

Table 1: Descriptive Statistics of Random Error of 2024 NECO SSCE Mathematics Multiple Choice Items

N	R	Random Error (%)
320	.844	0.156 (15.60%)

Table 1, show a p-value of .844 and random error of 0.156 which is 15.60% . This implies that 2024 NECO SSCE Mathematics Multiple Choice Items exhibited random error. The random error accounted for 15.60%.

Research Question Two: To what extent does the 2024 NECO SSCE Mathematics multiple choice test items exhibit systematic errors among senior secondary school students in Ethiope East Local Government Area?

Table 2: Paired t-test Statistics of Systematics error of 2024 SSCE NECO Mathematics Multiple Choice Items

Test	N	Mean	Std Deviation	Mean Difference	T	df	Sig.
1	320	19.91	8.49	-1.28	-3.276	319	.000
2		21.19	7.27				

Table 2: shows a calculated t of -3.296, a p- value of .000 and mean difference of -1.28. It is significant $p < .05$. This implies that 2024 NECO SSCE Mathematics Multiple Choice Item exhibited systematics error. The systematics error accounted for 6.09%. The

relatively small systematic error (6.09%) indicates that the test maintained reasonable fairness, while the random error (15.6%) reflects some limitations in precision and reliability. Consequently, the NECO SSCE Mathematics test can be described as a fair but slightly imprecise assessment tool.

Research Question Three: What is the difference in the random and systematic errors of the summative assessment of 2024 NECO SSCE Mathematics multiple choice test items administered towards secondary school students examination performance in Ethiope East Local Government Area?

Table 3: Descriptive of Difference between Random and Systematics Errors of 2024 NECO SSCE Mathematics Multiple Choice Items

Error Type	Variance	% of Total Error
Random Error	15.60	90.48
Systematic Error	1.64	9.52

Table 3 shows variance of 15.60 and 1.64 for random and systematics errors respectively. It also shows the value of 90.48% and 9.52% for random and systematic errors respectively. Since random error has higher percentage than the systematic error, this implies that the 2024 NECO SSCE Mathematics Multiple Choice Item is mostly affected by random inconsistencies rather than the systematic error. To strengthen the credibility of national examinations, NECO and other testing agencies should prioritize error minimization strategies, including better item writing, standardized administration, and psychometric evaluation of test items before final deployment. Reducing both random and systematic errors will ensure greater accuracy, fairness, and reliability in measuring students' mathematical ability.

Discussion of Findings

The findings from research question one revealed 15.60% of random errors. This implies that, the 2024 NECO SSCE Mathematics multiple-choice test items accounted for 15.6% random error. This could be as a result of unpredictable or chance factors. These

factors could be due to student fatigue, illness, student misread question, student guessing in multiple choice item. Students may have misread some items, response to other items aberrantly even when they are highly intelligent and guessing of options to perceived difficult items since errors usually fluctuate across students and items. This result suggests that while the 2024 NECO Mathematics test demonstrates acceptable reliability related to true-score variance it still contains a moderate level of random fluctuations. The findings collaborated the findings of Esomonu and Okeaba (2021) who found that NECO Mathematics examinations are prone to hidden sources of random error, affecting score dependability. Also, Imasuen and Adeosun (2023) reported that random error arising from student–item interactions constituted the largest share of error variance in WAEC Mathematics tests. The implication is that the random error found in this study is within the normal range observed in large-scale assessments. It suggests that there is need for students to avoid guessing in test–taking, control their emotions to prevent test anxiety often resulting in random error. The need for, clearer instructions on test items used in examinations by item development reduce random fluctuations is not rule-out.

Findings from research question two accounted for 6.09% systematic error. This suggest that the 2024 NECO SSCE Mathematics multiple-choice test item displayed 6.09% systematic error. This could be as result of bias across test items, some items may contain words that are not clear to examinees. It may also be due to poorly calibrated difficulty ranges, or resulted leading to uniform bias inherent in the test items. It can be that items favours student have more in-depth exposure to the content of items used for the study. This finding agrees with Bassey et al. (2021), who reported that systematic error remains a challenge in Nigerian assessments due to inadequate test-construction skills and lack of standardization. Similarly, Oyetunde (2022) found residual systematic variance in NECO Mathematics due to rater effects and item-bias patterns. Internationally, Brookhart (2015) and

the AERA, APA & NCME (2014) standards emphasis that systematic errors undermine test validity by introducing consistent distortions in measurement. The implication arising from this finding to measurement and evaluation experts, is that it is necessary to conduct item analysis calibration and differential item detection periodically to provide trend evidence needed for imperilment items development associated with quality assurance.

The findings from research question three shows the value of 90.48% and 9.52% for random and systematic errors respectively. This implies that the 2024 NECO SSCE Mathematics Multiple Choice Items are affected more by random inconsistencies rather than the systematic error. In other words, the 2024 NECO SSCE Mathematics Multiple Choice Items are. basically, affected by random inconsistencies rather than systematic bias. Psychometrically, this reflects an assessment that is fairly unbiased but not optimally precise because of an observed slight technical error across all examinees who took the test in this study. This finding potentially reflects that examinees variations in test-taking strategies, anxiety and stress level that introduces variability that are basically random. It is also possible that some items are easier for others while some examinees find similar items to be more difficult potentially accounting for more random errors than systematic errors due to predominate random variations.

The predominance of random error mirrors findings by Imasuen and Adeosun (2023), who observed that random fluctuations were the dominant source of score variability in WAEC Mathematics. Similarly, Ohiozua and Omorogiuwa (2025) findings affirms that random errors commonly occur moderately in 2018 and 2019 May/June Biology objective questions of National Business and Technical Certificate Examination in Nigeria. Oyetunde (2022) noted that random components (student × item interaction) contributed more to total error than systematic facets in NECO Mathematics. This finding aligns with international research suggesting that large-scale multiple-choice tests often have 80–95% of total error

attributable to random variance (Shavelson et al., 2019). Hence, while the 2024 NECO Mathematics test can be regarded as relatively fair, its precision could be enhanced by increasing item numbers, improving discrimination indices, and ensuring consistent test administration conditions.

CHAPTER FIVE

SUMMARY CONCLUSION AND RECOMMENDATIONS

In this chapter, the summary, conclusions and recommendation of this study based on the findings of analyzed data and interpretations are presented.

Summary

This study evaluated measurement inconsistency associated with random and systematic errors In 2024 NECO SSCE Mathematics Multiple Choice Test Items in Ethiope East Local Government Area, Delta State. Nigeria. The study specifically assessed extend and of

random and systematic errors and also, the difference between observed random and systematic errors in of 2024 NECO SSCE Mathematics multiple choice test items administered to senior secondary school three students in Ethiope East Local Government Area of Delta State. To achieve the purpose of the study, three research questions were answered in line with the purpose of the study.

The study adopted a survey research design. The population of the study was 1,270 SSS III students in Ethiope East Local Government Area of Delta State.The sample size was 320 selected through random sampling techniques in two stages. The instrument used for data collection in this study was a `standardized test adopted from NECO. The instrument was the official NECO 2024 Mathematics multiple choice 60-items with four options A-D administered and used for the in-school Senior School Certificate Examination (SSCE) in 2024.. The instrument was not validated because it was a standardized. For the reliability of the instrument, it was administered on twenty student twice with three weeks' time gap and subjected to Pearson Product Moment Correlation statistics to establish its dependability. The coefficient of .84 was obtained which was considered high enough for used in the study. .

The data used for the study were collected independently by the researcher and with the assistance of classroom teachers. Permission and informed consent were obtained from the

school management and participants of the study through a letter and physically presence. Research question one, two and three were answered using Crobach Alpha statistics, Paired t-test statistics and variance respectively in R software. Findings showed that:

1. The 2024 NECO SSCE Mathematics multiple-choice items exhibited 15.6% random error, indicating moderate but manageable score variability due to chance factors.
2. The test also exhibited 6.09% systematic error, suggesting a small but significant level of consistent bias in measurement.
3. The random error (90.48%) was substantially higher than systematic error (9.52%), implying that the test was more affected by unpredictable inconsistencies than by systematic bias.

Conclusion

Based on the findings the researcher concluded that the 2024 NECO SSCE Mathematics multiple-choice test items were more influenced by random inconsistencies than by systematic biases. In other words the test scores were fairly free from directional bias but moderately affected by unpredictable variations in students' performance and testing conditions.

Recommendations

Based on the findings and conclusions draw from this study, the following recommendations are made:

1. Item Development Procedures need to be enhanced by NECO in order to improve in the aspect of item-writing, moderation, and pilot-testing processes to ensure all multiple-choice items are clear, discriminating, and free from ambiguity that could increase random error.

2. Adoption of innovation whereby reliability coefficients and standard errors of measurement along with official results are reported to the public to promote transparency and public confidence in national assessments should be practice by examination bodies.
3. The 2024 NECO SSCE Mathematics test, though generally reliable and fair, can still be improved by addressing random inconsistencies and residual biases. Through the regular psychometric analyses to estimate the reliability and error structure of test items before and after administration as new concept related to implementation of regular psychometric audits in assessment as a policy.
4. Teachers should provide students with adequate exposure to standardized testing conditions and develop classroom assessments aligned with NECO standards to minimize performance fluctuations in summative assessment.

Contribution to Knowledge

This study contributes to the field of educational measurement and evaluation by:

1. This study has provided empirical evidence on the extent of random and systematic errors in the 2024 NECO SSCE Mathematics examination.
2. The study demonstrating that random errors contribute more to total measurement error than systematic errors in Nigerian large-scale assessments and profane a crucial implication that this study reaffirm that measurement precision and fairness remain vital to the credibility of high-stakes examinations in Nigeria.

Suggestions for Further Studies

To build on the findings of this research, future studies should:

1. Conduct comparative studies across multiple LGAs or states to determine whether error patterns are consistent nationwide.
2. Investigate **i**tem bias across gender, school type, and location using Differential Item Functioning analysis to detect equity issues.

3. Compare random and systematic error patterns in other subjects (such as, English, Chemistry, Biology among others) to understand discipline-specific test reliability.

REFERENCES

Abedalaziz, N. (2020). A gender-related differential item functioning of mathematics test items. *The International Journal of Educational and Psychological Assessment, 5*(1) 34-45

Abedalaziz, N. (2021). Detecting DIF using item characteristic curve approaches. The *International Journal of Educational and Psychological Assessment, 8*(2), 1-15.

Achor, E. E., Agogo, P., & Musa, J. (2024). The Impact of Teachers' Assessment Practices on Students' Acquisition of Process Skills and Performance in Basic Science. *Available at SSRN 5052615*.

Adeosun, P. K., & Ukanah, I. A. (2024). Estimating Standard Error of Measurement in BECE 2022 Basic Science Multiple Choice Test Items "in Etsako West Local Government Area of Edo State" Using Bootstrap Methods. *NIU Journal of Social Sciences*, *10*(2), 175-180.

Adeyemi, B. E. (2011). A comparative study of students' academic performance in public examinations in secondary schools in Ondo and Ekiti states, Nigeria. Current Research *Journal of economic theory, 3*(2), 36-42.

Adom, D., Mensah, J. A. & Dak, D. A. (2020). Test, measurement, and evaluation: Understanding and use of the concepts in education. *International Journal of Evaluation and Research in Education (IJERE). 9*(1): 109~119. DOI: 10.11591/ijere.v9i1.20457

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing.*

American Educational Research Association.

Ajayi, A. O., & Fakorede, A. O. (2020). Impact of test administration practices on students' performance in standardized assessments. *Journal of Educational Practice and Policy, 5*(2)*, 42–50.*

Alordiah, C. O. (2010). Relative effectiveness of two assessment procedures on junior secondary school students' achievement in mathematics in Agbor, Delta state. An unpublished MEd Dissertation, Delta state university, Abraka, Delta state, Nigeria.

Alordiah, C. O. (2015). Comparison of index of differential item functioning under the methods of item response theory and classical test theory in mathematics. An Unpublished PhD Thesis, Delta State University, Abraka, Delta State, Nigeria

Alruwais, N. M. (2020). Factors affecting student performance in e-assessment: An exploratory study. Education and Information Technologies, *25*(2), 1137–1155. https://doi.org/10.1007/s10639-019-10050-9

Amedahe, F. K., & Asamoah-Gyimah, K. (2016). *Introduction to measurement and evaluation* (6th Ed.). Hampton press, Cape Coast

Ayanwale, M.A., Chere1-Masopha, J. & Morena, M. C. (2022). The Classical Test or Item Response Measurement Theory: The Status of the Framework at the Examination Council of Lesotho. *International Journal of Learning, Teaching and Educational Research. 21*(8), 384-406, https://doi.org/10.26803/ijlter.21.8.22

Ayodele, O. J. (2021). Gender difference and performance of secondary school students in mathematics. *European Journal of Educational Studies, 3(1),* 173-179

Bassey, A. E., Ovat, A. O., & Ofem, E. E. (2021). Systematic error in measurement: Ethical implications in learners' assessment in the Nigerian educational system. *Pan-African Journal of Education,* 2(1), 45–56. https://openaccessglobal.com

Bekeking. (2023). NECO Mathematics Questions and Answers 2023/2024.

Bevington, P. R., & Robinson, D. K. (2003). *Data reduction and error analysis for the physical sciences*. McGraw-Hill.

Bichi, A. A. (2016). Classical test theory: An introduction to linear modeling approach to test and item analysis. *International Journal for Social Studies,2*(9), 27 – 33 doi: https://edupediapublications.org/journals

Braun, H., Kanjee, A., Bettinger, E., & Kremer, M. (2006). Improving education through assessment. innovation, and evaluation. *The American Academy of Arts and Sciences*,

Chakrabartty, S. N. (2013). Best split-half and maximum reliability. *IOSR Journal of Research & Method in Education,* 3(1), 1-8.

Daramola, C. A., & Akinkuotu, J. A. (2022). Enhancing item validity through item analysis: A study of West African mathematics assessments. *Journal of Educational Development, 11*(3), 85–98.

Denedo, M. E. (2023). *Estimating item information function of computer science achievement test for primary school pupils in South-South States, Nigeria*. PhD Thesis written and summited to the postgraduate school of studies , University of Port-Harcourt , River State, Nigeria.

Ebuoh, C.N. & Ojkafor, G (2017). Effects of systematic and random error test on student's achievement in Biology in Enugu educational zone. *ASSEREN Journal of Evaluation, 2(1)*, 135-144. doi: https://eprints.gouni.edu.ng

Essien, V. E., lban Amanso, E. O., Ogba, U. F., Williams, I. B., & Solomon, R. U. (2024). Assessment of Mathematics multiple-choice items using item response theory model of junior secondary school three students in Akwa Ibom State, Nigeria. *Journal of Educational Research in Developing Areas*, *5*(1), 99-107.

Esomonu, N. P.-M., & Okeaba, J. U. (2021). Analysis of hidden sources of error in NECO Mathematics examination. *European Journal of Education Studies,* 8(10), 145–157. https://www.ej-edu.org

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, *15*(3), 309-333.

Hohensinn, C. & Kubinger, K. D. (2009). On varying item difficulty by changing the response format for a mathematical competence test. *Austrian Journal of Statistics, 38*, 231-239.

Idusogie, O.V & Osunde, A. U. (2021). Assessment of multiple group confirmatory factor analysis in determining measurement invariance in 2018 NECO BECE Agriculture multiple choice items. *Journal of Educational Evaluation and Counselling Psychology (JEECP) 5,10-20.*

Imasuen, N. P., & Ogbomwan, S. M. (2014).Boostrap estimate for non—linear regression analysis. Nigerian annual journal for natural sciences, 5. 101-118

Imasuen, K., & Adeosun, P. K. (2023). Estimating sources of measurement error in WAEC Mathematics objective test using generalizability theory. *International Journal of Psychological and Behavioral Sciences,* 8(2), 1–12. https://www.sciencepg.com

Ismail et al, (2022) in their work 'Formative vs. summative assessment: impacts on academic motivation, attitude toward learning, test anxiety, and self-regulation skill' . retrieved from: https://doi.org/10.1186/s40468-022-00191-4

Iwuji, V.B.C. (2007), M*easurement and evaluation for effective teaching and learning*, Onitsha: Summer Educational Publishers (Nig.) Limited

Krejcia, R.V., & Morgan, D.W, (1970). Determining sample size for research activities. *Educational Psychological Measurement,* 30(3), 607-610

Kyung, T. H. (2013). *Windows software that generates IRT parameters and item responses: Research and Evaluation Program Methods (REMP).* University of Massachusetts Amherst.Doi:http://www.umass.edu/remp/software/simcata/weingen/homeF.html.(20 21)

Lane, R. (2018). Explainer: what's the difference between formative and summative assessment in schools?.

Linn, R. L., (2008). *Measurement and assessment in teaching*. Pearson Education India.

Mboto, F. A. & Bassey, S. W. (2024). Attitude and gender in science, technology and mathematics (STM) student's performance. *International Journal of Research in Education 1*(1&2), 34-37.

Medina-Díaz, M. del R., & Verdejo-Carrión, A. (2020). Validity and reliability in student learning evaluation throughout active methodologies. *Alteridad, 15*(2), 263-275. https://doi.org/10.17163/alt.v15n2.2020.10

National Examination Council (2024). *2024 SSCE External Results.* NECO website

National Examinations Council. (2024). 2024 SSCE External Results.

National Policy on Education (2016). *Federal Republic of Nigeria*

Nenty, H. J. (2020). Classical test theory's and item response theory's perceptions of the validity of test scores. *BOLESWA Journal of Educational Research*. *2*(7), 1-23.

Nering, M. L., & Ostini, R. (2020). *Handbook of polytomous item response theory models*. New York: Routledge.

Obarasua, L. O. & Amadioha, A. (2022). Analyzing the reliability and content validity of basic education certificate examination mathematics objective test items from 2018-2020. *Nigeria Journal of Empirical Studies in Psychology and Education (NJESPE) 23(2), 142-150.*

Obilor, E. I., & Obubere, M. (2020). Development and Validation of Social Studies Cognitive Achievement Test for Junior Secondary Schools in Rivers State. *International Academic Journal of Educational Research*. 5(3): 1-20,

Ohiozua & Omorogiuwa (2025). DETERMINATION OF ABERRANT Examinees' responses Of 2018 And 2019 May/June Biology Objective Questions Of National Business And Technical Certificate Examination. *Journal of Evaluation*, *10*(1), 185-199.

Okoro, I. F. (2021). Curriculum Implementation: A driving force for achieving gender equity in education. *Journal of International Gender studies (JIGS), 6*, 55-64.

Olela et al, (2021) titled 'The Influence of Summative Evaluation on Classroom Practices in Rarieda Sub-County, Kenya'. https://doi.org/10.46606/eajess2021v02i03.0105

Omole, O. A., & Onuka, A. O. (2021). Bias in standardized tests: The influence of socio-cultural context on item performance. *African Journal of Educational Measurement, 13*(1), 21–34.

Omoruan, B. E. (2017). Construction, Validation and Application of Mathematics Language Mastery Achievement test for Junior Secondary Schools Students in Delta and Edo States, Nigeria*.Unpublished Doctoral Thesis*, University of Port, Port Harcourt.

Opong, K. N. (2006). Construction and Validation of an Economics Achievement Test for Senior Secondary School Students in Ethiope East and Ukwuani Local Government Areas of Delta State. *An Unpublished M.Ed. Dissertation.* Delta State University, Abraka.

Osadebe, P. U. (2016). Practical Guide to Item Generation*. A seminar presented at Delta State University, Abraka*.

Osterlind, S. J. (2012). Item response theory. Journals of home school and academic learning. Retrieved November, 30, 2022, from http://www.education.com>Home>School and Academics>classroom learning

Oyetunde, K. C. (2022). *Application of generalizability theory in estimating measurement error in NECO Mathematics examination in Osun State.* Obafemi Awolowo University Repository. https://ir.oauife.edu.ng

Pande, S. S., Pande, S. R., Parate, V. R., Nikam, A. P., and Agrekar, S. H. (2013). Correlation between difficulty & discrimination indices of MCQs in formative exam in Physiology. *South-East Asian Journal of Medical Education* 7, 45–45. doi: 10.4038/seajme.v7i1.149.

Paniagua, M. A., & Swygert, K. A. (2016). *Constructing written test questions for the basic and clinical sciences*. Philadelphia, PA: National Board of Medical Examiners.

Reeve, B. B. (2020). Item and scale-level analysis of clinical and non-clinical sample responses to the MMPI-2 depression scales employing item response theory. *Unpublished Doctoral Dissertation*, University of North Carolina at Chapel Hill.

Robert-Okah, I. (2021). Gender Equity and women empowerment in Nigeria: A paradigm for natural development. *Journal of International Gender Studies (JIGS). 6*, 167-179.

Samaila, N., & Alhassan, N. (2022). School location and gender as determinants of senior secondary students' performance in biology examination in Kaduna State. *ATBU Journal of Science, Technology and Education*, *10*(1), 300-310.

Santrock, J. W. (2016). Life span development. New York: McGraw-Hill.

Shavelson, R. J., Baxter, G. P., & Pine, J. (2019). Performance assessments: Reform, pitfalls,

Shirani, R. J. (1997). Cousine for estimating the standard error and confidence intervals. Bootstrap Methods and Their Application. Cambridge University Press.

Stcharlesedu. (n.d.). NECO Mathematics Past Questions PDF Download, Objective, Theory.

Study & Scholarships. (n.d.). National Examination Council (NECO) Exams.

Udo, E., Uyoata, U. E., Inyon, A. U., & Ekanem, I. E. (2021). Instructional strategies for enhancing gender equity in learning primary science: Cooperative small group instructional mode. Journal of International Gender Studies (JIGS), 6, 1-10.

Ukwuije, R. P. I. & Opara, I. (2012). *Test and Measurement for Teachers,* Alakahia-Akpor: Chadik Printing Press.

Uwadiae, I. (2018). WAEC released result. Saturday Punch September, 27:10.

Vincent, W., & Shanmugam, S. K. S. (2020). The role of classical test theory to determine the quality of classroom teaching test items. *Pedagogia: Jurnal Pendidikan*, *9*(1), 5-34.

Wozencraft (2023). Sex comparison of certain abilities. *Journal of Educational Research,* 11,21-27.

Wushishi, O. (2025). NECO *Releases 2024 SSCE results written from Wednesday June 19 to Thursday July 24, 2024*. National Television Authority (NTA)

Yashim, A. U., Mhab, L. C., & Waziri, J. A. (2021). Measurement errors in educational assessment. *Journal of Educational Theory and Practice (JETP) Online Journal*, *1*(1).

Yashim, A.U., Mhab, L.C. & Waziri, J.A. (2021). Measurement errors in educational assessment. *JETP Online Journal.1*(1). 1, 030. http://independent.academia.edu/leagueofresearchers

Yilwa, A. V., & Olarinoye, R. D. (2014). The influence of location, proprietorship, sex and grade level on Junior secondary schools students' performance in the skill of observing. *Nigerian Journal of Curriculum Studies, 11*(1), 54-63.

Young, D. J. (2018). Rural and urban differences in students achievement in science and mathematics: A multilevel Analysis. School Effectiveness and School Improvement, *Review Journal9*(4). 386-412

APPENDIX A

Table 1: Sample from Population Distribution of SSS III Students in Ethiope East LGA

SN	NAME OF SCHOOL	Sample		Total
		Male	Female	
1.	Igun Secondary School, Igun	18	10	28
2.	Oviore Sec. Sch. Oviore	30	21	51
3.	Okurekpo Secondary School, Okurekpo	15	12	27
4.	Orhoakpo Secondary School, Orhoakpo	23	40	63
5.	Okpara Boys Sec. School, Okpara Inland	56		56
6.	Abraka Grammar School Abraka	28	30	58
7.	Eku Girls Sec. Sch., Eku	0	64	64
8.	Umiaghwa Sec. Sch. Abraka	27	26	53
9.	Kokori Mixed Secondary School, Abraka	24	29	53
10.	Ojeta Secondary School, Abraka	24	20	44
11.	Okpara mixed secondary school, Okpara waterside.	32	33	65
12.	Otorho secondary school, Otorho Ab raka	27	31	58
13.	Ovu college, Urhodo Ovu	19	23	42
14.	Ovu Grammar School, Ovu Inland	27	31	58
15.	Agbon College, Okpara Inland		49	49
16.	Agbon Secondary School, Isiokolo	21	20	41
17.	Baptist Medical Center, Eku	31	33	64
18.	Egbo Commercial Grammar School, Kokori	21	34	55
19.	Ekpan Ovu Secondary School, Ekpan Ovu	19	32	51
20.	Eku Girls Secondary School, Abraka	35	22	57
21.	Erho Secondary School, Abraka	35	22	57
22.	Ibruvwe Secondary School, Samagidi, Kokori	32	31	63
23.	Isiokolo Girls Secondary School, Isiokolo		34	34
24.	Kokori Girls Secondary School, Kokori		79	79
	Total	**544**	**726**	**1270**

Source: Ministry of Basic and Secondary Education, Asaba, Delta State, 2025

APPENDIX B

2024 NECO MATHEMATICS MULTIPLE CHOICE TEST QUESTIONS

Answer **all** questions.

Each question is followed by **five** options lettered **A** to **E**. Choose the correct option for each question and shade in **pencil** on your answer sheet the answer space that bears the same letter as the option you have chosen. Give only **one** answer to each question and erase completely any answer you wish to change. Do **all** rough work on this question paper.

An example is given below:

The product of three numbers is 3876. Two of the numbers are 17 and 19. **What** is the third number?

A. 57
B. 12
C. 6
D. 3
E. 2

The correct option is '12' which is lettered **B**. Therefore, answer space **B would** be shaded as shown below:

[A] [B] [C] [D] [E]

1. Express 384.126 to the nearest hundred.

   A. 480
   B. 400
   C. 384
   D. 380
   E. 300

2. If A = {natural numbers between 3 and 15}, find n(A).

   A. 15
   B. 14
   C. 13
   D. 11
   E. 10

3. Evaluate $3241_{five} - 1342_{five}$.

   A. 2341_{five}
   B. 1344_{five}
   C. 1342_{five}
   D. 1324_{five}
   E. 1234_{five}

4. Simplify $(4\sqrt{3}-6)(4\sqrt{3}+6)$.

   A. 12
   B. 4
   C. $2\sqrt{3}$
   D. 2
   E. $\sqrt{3}$

5. The third term of a Geometric Progression is – 5 and the seventh term is – 80. Find the common ratio.

   A. 16
   B. 8
   C. 5
   D. 3
   E. 2

6. A man with an annual salary of ₦1,300,000.00 is to pay an income tax of 22%. Calculate his tax, if his allowances amount to ₦137,000.00:

A. ₦254,760.00
B. ₦255,860.00
C. ₦258,860.00
D. ₦350,140.00
E. ₦907,140.00

7. A wire of length 25 cm was measured by a student to be 24.4 cm. Find the percentage error.

A. 4.4
B. 3.4
C. 2.4
D. 1.4
E. 0.4

8. How many years will ₦12,000.00 saved in a bank amount to ₦12,960.00 at 2% per annum simple interest?

A. 6
B. 5
C. 4
D. 3
E. 2

9. Rationalize $\frac{5}{\sqrt{2}+\sqrt{3}}$.

A. $5(\sqrt{3}-\sqrt{2})$
B. $5(\sqrt{2}+\sqrt{3})$
C. $5(\sqrt{2}-\sqrt{3})$
D. $\sqrt{3}+5\sqrt{2}$
E. $\sqrt{3}+2\sqrt{5}$

10. In a class of 80 students, every student studies Mathematics or Geography or both. If 65 students study Mathematics and 50 study Geography, how many study both subjects?

A. 45
B. 35
C. 30
D. 20
E. 15

11. The 11th term of an Arithmetic Progression is 63. Find the first term, if its common difference is 3.

A. 65
B. 63
C. 35
D. 33
E. 30

12. Evaluate 15⊗26 in modulo 5.

A. 0 (mod 5)
B. 1 (mod 5)
C. 2 (mod 5)
D. 4 (mod 5)
E. 5 (mod 5)

13. Solve $2^{3x} = 16^{\frac{3}{4}}$.

A. 8
B. 4
C. 3
D. 2
E. 1

14. If Olu, Tony and Tunde share ₦240,000.00 in the ratio 2:3:5 respectively, what is two-thirds of Tunde's share?

A. ₦120,000.00
B. ₦80,000.00
C. ₦72,000.00
D. ₦48,000.00
E. ₦40,000.00

15. Arrange the following fractions in ascending order of magnitude: $\frac{2}{3}, \frac{3}{5}, \frac{5}{12}, \frac{4}{15}, \frac{3}{10}$.

A. $\frac{4}{15}, \frac{3}{10}, \frac{5}{12}, \frac{3}{5}, \frac{2}{3}$
B. $\frac{3}{10}, \frac{4}{15}, \frac{5}{12}, \frac{3}{5}, \frac{2}{3}$
C. $\frac{4}{15}, \frac{5}{12}, \frac{3}{10}, \frac{3}{5}, \frac{2}{3}$
D. $\frac{4}{15}, \frac{5}{12}, \frac{3}{5}, \frac{3}{10}, \frac{2}{3}$
E. $\frac{2}{3}, \frac{3}{5}, \frac{5}{12}, \frac{4}{15}, \frac{3}{10}$

16. Find the values of x, y and z respectively for which $\begin{bmatrix} x & 2y \\ z & 9 \end{bmatrix} = \begin{bmatrix} 4 & 12 \\ 3 & 9 \end{bmatrix}$.

A. (6, 4, 3)
B. (4, 6, 3)
C. (6, 3, 4)
D. (3, 4, 6)
E. (4, 4, 6)

17. If $P = \begin{bmatrix} -1 & 2 \\ 3 & 1 \end{bmatrix}$ and $Q = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}$, find PQ.

A. $\begin{bmatrix} 2 & 1 \\ -8 & -10 \end{bmatrix}$
B. $\begin{bmatrix} 2 & 1 \\ 8 & 10 \end{bmatrix}$
C. $\begin{bmatrix} -2 & 1 \\ 8 & 10 \end{bmatrix}$
D. $\begin{bmatrix} -2 & -1 \\ 8 & 10 \end{bmatrix}$
E. $\begin{bmatrix} 2 & -1 \\ 8 & 10 \end{bmatrix}$

18. Evaluate $\log_4 16 + \log_3 27 - \log_8 4096$.

A. $\frac{1}{9}$
B. $\frac{1}{3}$
C. 1
D. 2
E. 3

19. Solve the equation $\frac{4x}{5} - \frac{7}{3} = \frac{5x}{12}$.

A. $-6\frac{2}{23}$
B. $-3\frac{1}{2}$
C. $5\frac{2}{23}$
D. $6\frac{2}{23}$
E. $6\frac{3}{23}$

20. If $f(x) = 3x^2 - 9x - 5$, find $f(-3)$.

A. -5
B. 5
C. 27
D. 49
E. 54

21. Find u in terms of f and v in the relation $\frac{1}{v} = \frac{1}{f} - \frac{1}{u}$.

A. $u = \frac{-f}{(v-f)}$

B. $u = \frac{fv}{(f-v)}$

C. $u = \frac{fv}{(v+f)}$

D. $u = \frac{fv}{(v-f)}$

E. $u = \frac{fv}{2(f+v)}$

22. Find the quadratic equation whose roots are -1 and 5.

A. $x^2 - 4x - 5 = 0$
B. $x^2 - 4x + 1 = 0$
C. $x^2 - 4x + 5 = 0$
D. $x^2 + 4x - 5 = 0$
E. $x^2 + 4x + 5 = 0$

23. The product of two numbers is 40 and their sum is 13, find the numbers.

A. 2 and 20
B. 2 and 8
C. 4 and 10
D. 4 and 8
E. 5 and 8

24. If $(x+6)$ is a factor of $x^2 + 4x - 12$, find the other factor.

A. $(x - 2)$
B. $(x - 6)$
C. $(x + 2)$
D. $(x + 4)$
E. $(x + 6)$

25. Calculate the mid-point of the line joining $(8, -3)$ and $(-2, 3)$.

A. $(-3, 0)$
B. $(0, 5)$
C. $(3, 0)$
D. $(0,3)$
E. $(5, 0)$

26. Calculate the gradient of a line joining the points $(-2, -5)$ and $(4, 8)$, correct to 1 decimal place.

A. 4.2
B. 3.4
C. 3.2
D. 2.2
E. 2.1

27. Given the statements:
p: All terrorists are guilty
q: All terrorists are criminals
Write the following statement in symbolic form; "All terrorists are **not guilty but criminals**".

A. $p \vee q$

B. $p \wedge q$

C. $p \wedge \sim q$

D. $\sim p \vee q$

E. $\sim p \wedge q$

28. Simplify $\frac{x+\frac{1}{3}}{x+\frac{1}{2}}$.

A. $\frac{2(3x+1)}{3(2x+1)}$

B. $\frac{2+x}{3+x}$

C. $\frac{x+1}{x-1}$

D. $\frac{3x+1}{2x-1}$

E. $\frac{x^2}{x-1}$

29. Find the equation of a line whose gradient is 6 and y-intercept is -7.

A. $y = 8 + 7x$
B. $y = -7 - 6x$
C. $y = 7 - 6x$
D. $y = -7 + 6x$
E. $y = 7 + 6x$

30. Find the sum of the roots of the quadratic equation $x^2 - 5x + 6 = 0$.

A. 10
B. 5
C. -2
D. -3
E. -5

31. Find the roots of the equation $8x^2 - 6x - 9 = 0$.

A. $x = \frac{3}{2}$ or $\frac{-4}{3}$

B. $x = \frac{3}{2}$ or $\frac{-3}{4}$

C. $x = \frac{3}{2}$ or $\frac{3}{4}$

D. $x = \frac{-3}{2}$ or $\frac{3}{4}$

E. $x = \frac{-3}{2}$ or $\frac{-3}{4}$

32. Expand $(2x - 3)(3x + 4)$.

A. $6x^2 - 17x - 12$
B. $6x^2 - x - 12$
C. $6x^2 - x + 12$
D. $6x^2 + x - 12$
E. $6x^2 + 17x - 12$

33. The difference between the present ages of two brothers is 6 and their product is 135. What is the sum of their ages?

A. 27
B. 24
C. 22
D. 21
E. 15

34. In **Fig. 3.1**, PQS is a circle with centre O. RST is a tangent at S and ∠ SOP = 120°. Find ∠ PST.

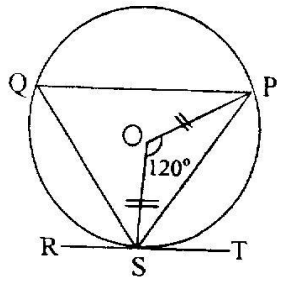

Fig. 3.1

A. 64°
B. 60°
C. 35°
D. 31°
E. 29°

35. In **Fig. 3.2**, A, B, C and D are points on a circle with centre O. $\overline{BA}$ is produced to M.
If ∠ MAD = **82°** and ∠ ADO = 74°, find ∠ ABO.

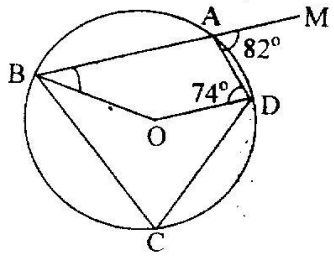

Fig. 3.2

A. 24°
B. 74°
C. 82°
D. 98°
E. 164°

36. Calculate the area of trapezium ABCD in **Fig. 3.3**.

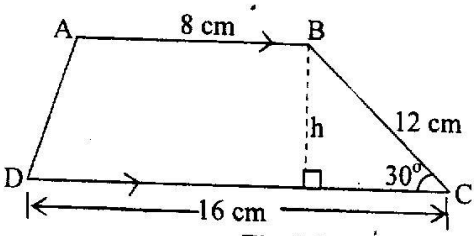

Fig. 3.3

A. 27 cm^2
B. 36 cm^2
C. 72 cm^2
D. 92 cm^2
E. 144 cm^2

37. A boy walks 4 km due West. He then changes direction and walks on a bearing of 214° until he is South-West of his starting point. How far is he from his starting point? Correct your answer to one decimal place.

A. 5.1 km
B. 8.3 km
C. 11.1 km
D. 15.7 km
E. 17.4 km

38. Two points P and Q lie on the same great circle. P is on latitude 70°N and Q is on latitude 55°N. Calculate their difference in latitudes.

A. 15°
B. 25°
C. 45°
D. 65°
E. 125°

39. A bird which is on top of a building 35 m high observes a prey 25 m away from the foot of the building. Calculate the angle of depression of the prey from the bird.

A. 75.00°
B. 60.00°
C. 54.46°
D. 35.25°
E. 10.00°

40. Calculate the length of an arc which subtends an angle of 66° at the centre of a circle of radius 8 cm, correct to one decimal place.

A. 3.4 cm
B. 5.2 cm
C. 7.4 cm
D. 9.2 cm
E. 12.3 cm

41. Find the value of y in **Fig. 3.4**.

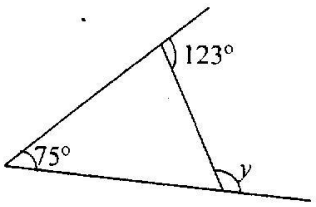

Fig. 3.4

A. 48°
B. 57°
C. 105°
D. 132°
E. 198°

42. The base radius and height of a cone are 4 cm and 6 cm respectively. Calculate its volume, correct to the nearest whole number.

A. 75 cm^3
B. 86 cm^3
C. 98 cm^3
D. 101 cm^3
E. 110 cm^3

43. Find the distance between the points $(3, -4)$ and $(-5, 2)$.

A. 18
B. 16
C. 14
D. 12
E. 10

44. Calculate the surface area of a sphere with diameter 10.4 cm, correct to 3 significant figures.
(Take $\pi = 3.142$)

A. 439 cm^2
B. 400 cm^2
C. 340 cm^2
D. 339 cm^2
E. 338 cm^2

45. Obi walks 400 m to the top of a hill which slopes at angle 30° to the horizontal. Determine the height of the hill.

A. 430 m
B. $400\sqrt{3}$m
C. $200\sqrt{3}$ m
D. 200 m
E. 100 m

46. The volume of a spherical ball is 114 cm^3. Find its radius to the nearest whole number.

A. 8 cm
B. 6 cm
C. 5 cm
D. 4 cm
E. 3 cm

47. Find the equation of a line with gradient $\frac{-1}{4}$, passing through the point (3, 2).

A. $y + x = 3$
B. $y - x = 1$
C. $5y + 2x = 2$
D. $4y + x = 11$
E. $3y + 4x = 7$

48. Calculate the value of y in **Fig. 3.5**, leaving your answer in surd form.

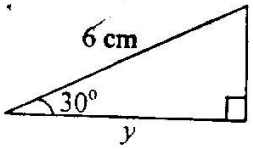

Fig. 3.5

A. $3\sqrt{3}$ cm
B. $2\sqrt{3}$ cm
C. $\sqrt{3}$ cm
D. $\frac{3\sqrt{3}}{2}$ cm
E. $\sqrt{2}$ cm

Use the information below to answer questions **49- 51**.

Given the following set of numbers; 11, 8, 9, 6, 4, 3, 10, 2, 6, 5.

49. Calculate the mean.

A. 8.0
B. 7.2
C. 6.4
D. 6.0
E. 5.8

50. Find the median.

A. 8
B. 6
C. 5
D. 4
E. 3

51. Find the mode.

A. 11
B. 8
C. 6
D. 4
E. 3

52. Find the range of the following set of numbers; 4, 9, 6, 3, 2, 8, 10, 7, 11.

A. 11
B. 9
C. 7
D. 4
E. 2

53. Calculate the variance of the following set of numbers;
30, 28, 35, 25, 37.

A. 31.0
B. 25.1
C. 19.6
D. 19.5
E. 15.3

54. If the tickets numbered 1 to 16 inclusive are mixed up and a ticket is drawn at random, what is the probability that the ticket drawn is a multiple of 2 or 3?

A. $\frac{5}{16}$
B. $\frac{1}{2}$
C. $\frac{5}{8}$
D. $\frac{11}{16}$
E. $\frac{13}{16}$

55. The following are the lucky numbers in a raffle draw; 4, 5, 12, 20, 2, 8, 3, 6, 10, 9, 7, 25, 12, 10, 14, 27.
If a number is picked at random, what is the probability that it is a perfect cube?

A. $\frac{11}{16}$
B. $\frac{5}{8}$
C. $\frac{7}{16}$
D. $\frac{3}{8}$
E. $\frac{1}{8}$

56. The probability that it will rain in Lagos and Oyo on the same day are $\frac{3}{4}$ and $\frac{1}{2}$ respectively. Find the probability that it will not rain in both towns on the same day.

A. $\frac{1}{12}$
B. $\frac{1}{8}$
C. $\frac{1}{6}$
D. $\frac{1}{4}$
E. $\frac{1}{2}$

57. **Table 3.1:** Ages of students

Age (years)	16	17	18
No of students	4	8	6

Table 3.1 shows the ages of students in a particular class. What is the probability that a student chosen at random is less than 18 years?

A. $\frac{9}{10}$
B. $\frac{4}{5}$
C. $\frac{2}{3}$
D. $\frac{1}{2}$
E. $\frac{1}{3}$

Use the information below to answer questions **58 - 60**.

A particle moves a distance of S metres, where $S = 5t^3 - 12t^2 + 7$.

58. At what time is its acceleration equals zero?

A. 0.8 sec
B. 1.2 sec
C. 1.6 sec
D. 2.0 sec
E. 2.5 sec

59. Find the velocity **after 3** seconds.

A. 70 m/s
B. 63 m/s
C. 50 m/s
D. 40 m/s
E. 30 m/s

60. Find the acceleration after 12 seconds.

A. 400 m/s^2
B. 360 m/s^2
C. 336 m/s^2
D. 300 m/s^2
E. 180 m/s^2

APPENDIX C

Formula for Computation

Note :Random error – 1 – r

= 1 - .844

=0.156

Systematic error = $\frac{mean\ differnece}{mean\ score} \times 100$

Mean of score = 19.91 +21.19 ÷ 2

=41.1÷ 2 = 21

= $\frac{-1.28}{21}$ = 0.0609×× 100

= 6.09%

Note : Total error of variance = 15.60 + 1.64 =17.24

% of total error = $\frac{random\ error\ of\ variance}{total\ variance}$

= $\frac{systematic\ error}{total\ variance}$

APPENDIX D

ANALYSIS OUTPUT

Correlations

Notes

Output Created		25-OCT-2025 14:19:26
Comments		
Input	Active Dataset	DataSet0
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	320
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics for each pair of variables are based on all the cases with valid data for that pair.
Syntax		CORRELATIONS /VARIABLES=TEST1 TEST2 /PRINT=TWOTAIL NOSIG /MISSING=PAIRWISE.
Resources	Processor Time	00:00:00.02
	Elapsed Time	00:00:00.19

[DataSet0]

Correlations

		TEST1	TEST2
TEST1	Pearson Correlation	1	.626**
	Sig. (2-tailed)		.000
	N	320	320
TEST2	Pearson Correlation	.626**	1
	Sig. (2-tailed)	.000	
	N	320	320

**. Correlation is significant at the 0.01 level (2-tailed).

T-Test

Notes

Output Created		25-OCT-2025 14:20:53
Comments		
Input	Active Dataset	DataSet0
	Filter	<none>
	Weight	<none>

	Split File	<none>
	N of Rows in Working Data File	320
Missing Value Handling	Definition of Missing	User defined missing values are treated as missing.
	Cases Used	Statistics for each analysis are based on the cases with no missing or out-of-range data for any variable in the analysis.
Syntax		T-TEST PAIRS=TEST1 WITH TEST2 (PAIRED) /CRITERIA=CI(.9500) /MISSING=ANALYSIS.
Resources	Processor Time	00:00:00.00
	Elapsed Time	00:00:00.03

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	TEST1	19.9188	320	8.49449	.47486
	TEST2	21.1906	320	7.27588	.40673

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	TEST1 & TEST2	320	.626	.000

Paired Samples Test

		Paired Differences			
					95% Confidence Interval of the Difference
		Mean	Std. Deviation	Std. Error Mean	Lower
Pair 1	TEST1 - TEST2	-1.27188	6.90358	.38592	-2.03115

Paired Samples Test

		Paired Differences				
		95% Confidence Interval of the Difference				
		Upper	t	df	Sig. (2-tailed)	
Pair 1	TEST1 - TEST2	-.51260	-3.296	319	.001	

Reliability

Notes

Output Created	25-OCT-2025 14:36:46

Comments		
Input	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	320
	Matrix Input	
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on all cases with valid data for all variables in the procedure.
Syntax		RELIABILITY /VARIABLES=VAR00001 VAR00002 VAR00003 VAR00004 VAR00005 VAR00006 VAR00007 VAR00008 VAR00009 VAR00010 VAR00011 VAR00012 VAR00013 VAR00014 VAR00015 VAR00016 VAR00017 VAR00018 VAR00019 VAR00020 VAR00021 VAR00022 VAR00023 VAR00024 VAR00025 VAR00026 VAR00027 VAR00028 VAR00029 VAR00030 VAR00031 VAR00032 VAR00033 VAR00034 VAR00035 VAR00036 VAR00037 VAR00038 VAR00039 VAR00040 VAR00041 VAR00042 VAR00043 VAR00044 VAR00045 VAR00046 VAR00047 VAR00048 VAR00049 VAR00050 VAR00051 VAR00052 VAR00053 VAR00054 VAR00055 VAR00056 VAR00057 VAR00058 VAR00059 VAR00060 /SCALE('Random Error') ALL /MODEL=ALPHA /STATISTICS=DESCRIPTIVE SCALE /SUMMARY=TOTAL.
Resources	Processor Time	00:00:00.02
	Elapsed Time	00:00:00.05

[DataSet1]

Scale: Random Error

Case Processing Summary

		N	%
Cases	Valid	319	99.7
	Excluded[a]	1	.3
	Total	320	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.844	60

Item Statistics

	Mean	Std. Deviation	N
VAR00001	.4263	.49532	319
VAR00002	.3135	.46464	319
VAR00003	.3135	.46464	319
VAR00004	.5831	.49383	319
VAR00005	.2288	.42075	319
VAR00006	.1912	.39388	319
VAR00007	.2633	.44113	319
VAR00008	.3605	.48090	319
VAR00009	.4263	.49532	319
VAR00010	.4608	.49925	319
VAR00011	.3041	.46074	319
VAR00012	.3448	.47606	319
VAR00013	.2069	.40572	319
VAR00014	.4232	.49484	319
VAR00015	.2382	.42668	319
VAR00016	.3197	.46711	319
VAR00017	.4608	.49925	319
VAR00018	.2821	.45074	319
VAR00019	.2194	.41451	319
VAR00020	.4765	.50023	319
VAR00021	.1850	.38887	319
VAR00022	.4295	.49578	319
VAR00023	.4483	.49810	319
VAR00024	.4734	.50007	319
VAR00025	.4514	.49842	319
VAR00026	.3323	.47177	319
VAR00027	.4357	.49663	319
VAR00028	.5423	.49899	319
VAR00029	.2069	.44278	319
VAR00030	.4389	.49703	319
VAR00031	.3009	.45939	319
VAR00032	.2633	.44113	319
VAR00033	.2476	.43232	319
VAR00034	.6364	.48180	319
VAR00035	.4890	.50066	319
VAR00036	.2194	.41451	319
VAR00037	.1787	.38369	319
VAR00038	.4639	.49948	319
VAR00039	.3417	.47502	319
VAR00040	.1630	.36995	319
VAR00041	.1254	.33168	319
VAR00042	.2539	.43594	319
VAR00043	.1097	.31303	319

VAR00044	.1881	.39140	319
VAR00045	.2884	.45373	319
VAR00046	.2100	.40797	319
VAR00047	.2727	.44606	319
VAR00048	.4075	.49215	319
VAR00049	.5549	.49776	319
VAR00050	.5235	.50023	319
VAR00051	.4671	.49970	319
VAR00052	.4451	.49776	319
VAR00053	.1755	.38103	319
VAR00054	.4608	.49925	319
VAR00055	.2790	.44921	319
VAR00056	.3950	.48962	319
VAR00057	.1975	.39873	319
VAR00058	.1755	.38103	319
VAR00059	.1755	.38103	319
VAR00060	.1818	.38630	319

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
VAR00001	19.5517	71.512	.158	.844
VAR00002	19.6646	70.117	.353	.840
VAR00003	19.6646	70.161	.347	.841
VAR00004	19.3950	71.001	.220	.843
VAR00005	19.7492	69.566	.475	.838
VAR00006	19.7868	73.219	-.044	.847
VAR00007	19.7147	72.437	.059	.846
VAR00008	19.6176	68.841	.502	.837
VAR00009	19.5517	71.732	.131	.845
VAR00010	19.5172	71.420	.167	.844
VAR00011	19.6740	71.535	.171	.844
VAR00012	19.6332	72.755	.012	.847
VAR00013	19.7712	71.014	.278	.842
VAR00014	19.5549	69.587	.393	.840
VAR00015	19.7398	71.564	.185	.844
VAR00016	19.6583	72.251	.077	.846
VAR00017	19.5172	68.993	.463	.838
VAR00018	19.6959	72.281	.077	.846
VAR00019	19.7586	72.309	.085	.845
VAR00020	19.5016	68.427	.532	.837
VAR00021	19.7931	71.643	.195	.843
VAR00022	19.5486	68.657	.508	.837
VAR00023	19.5298	68.225	.560	.836
VAR00024	19.5047	68.408	.534	.837
VAR00025	19.5266	70.080	.329	.841
VAR00026	19.6458	68.582	.547	.837
VAR00027	19.5423	69.161	.444	.838
VAR00028	19.4357	68.542	.519	.837
VAR00029	19.7712	72.240	.085	.845
VAR00030	19.5392	68.727	.498	.837
VAR00031	19.6771	72.678	.024	.847
VAR00032	19.7147	71.072	.243	.843
VAR00033	19.7304	72.688	.028	.846
VAR00034	19.3417	71.050	.221	.843

VAR00035	19.4890	70.785	.242	.843
VAR00036	19.7586	72.938	-.005	.847
VAR00037	19.7994	73.073	-.022	.847
VAR00038	19.5141	70.408	.289	.842
VAR00039	19.6364	70.767	.261	.842
VAR00040	19.8150	71.837	.176	.844
VAR00041	19.8527	72.780	.033	.845
VAR00042	19.7241	69.943	.404	.840
VAR00043	19.8683	72.027	.179	.843
VAR00044	19.7900	73.808	-.131	.848
VAR00045	19.6897	70.051	.371	.840
VAR00046	19.7680	70.097	.412	.840
VAR00047	19.7053	70.051	.379	.840
VAR00048	19.5705	68.919	.479	.838
VAR00049	19.4232	68.251	.557	.836
VAR00050	19.4545	70.123	.323	.841
VAR00051	19.5110	68.226	.558	.836
VAR00052	19.5329	69.948	.346	.840
VAR00053	19.8025	73.549	-.094	.848
VAR00054	19.5172	68.942	.469	.838
VAR00055	19.6991	68.444	.596	.836
VAR00056	19.5831	70.571	.276	.842
VAR00057	19.7806	71.933	.146	.844
VAR00058	19.8025	73.694	-.116	.848
VAR00059	19.8025	73.492	-.086	.848
VAR00060	19.7962	73.804	-.132	.848

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
19.9781	73.078	8.54857	60