



## A Comparative Analysis of Students' Academic Performance in WAEC and NECO in Mathematics and English Language in Jos South, Nigeria

Ojodale Oswald Ameh

Department of Educational Foundations, Federal University Lokoja, Nigeria

\*Corresponding Author E-mail; amehoswald@gmail.com, ORCI ID: <https://orcid.org/0009-0009-5228-7123>

James Tokgak Wutenwa

Department of Educational Foundations, University of Jos, Nigeria

E-mail; jamestokgak1@gmail.com, ORCID ID: <https://orcid.org/0009-0007-2087-170X>

Received 30/03/2026

Revised 29/04/2026

Accepted 07/05/2026

### Abstract

**Background and Aims:** Comparability in large-scale examinations is essential for fairness in educational assessment, certification, and access to higher education. In Nigeria, public examination results play a major role in determining students' eligibility for graduation and admission into tertiary institutions. The West African Examinations Council (WAEC) and the National Examinations Council (NECO) conduct parallel senior secondary school examinations that are often treated as equivalent in content coverage, difficulty level, and grading standards. However, concerns remain about whether performance outcomes across the two examination bodies are consistently comparable. This study examined differences in students' performance in WAEC and NECO Mathematics and English Language examinations in Jos South Local Government Area.

**Methodology:** The study adopted an ex post facto research design because it relied on existing examination records without manipulating any variables. Secondary data were obtained from selected secondary schools in Jos South Local Government Area for the years 2019 to 2021. The sample comprised 1,800 students, with 600 candidates drawn from each year. The data covered students' performance in Mathematics and English Language in both WAEC and NECO examinations. Descriptive statistics, particularly percentages, were used to summarize performance trends across subjects and years. Chi-square tests of independence were applied to determine whether significant differences existed between WAEC and NECO performance outcomes at the 0.05 level of significance. Cramér's V was also used to assess the strength of association. A major limitation of the study was its ex post facto design, which did not allow control over factors such as students' socio-economic background, instructional quality, school environment, or examination conditions. Therefore, the findings indicate associations rather than causal relationships.

**Results:** The findings showed variations in students' performance across the two examination bodies during the three years. In 2019, a statistically significant difference was found in English Language performance between WAEC and NECO candidates. In 2020, significant differences were observed in both Mathematics and English Language, indicating variations in performance patterns between the two examinations. However, the effect sizes across these significant results were consistently small, suggesting that the observed differences had limited practical significance. In 2021, no statistically significant differences were found in either Mathematics or English Language, indicating closer alignment in student performance outcomes for that year. Descriptively, NECO recorded a relatively higher proportion of passes than WAEC across some subjects and years.

**Conclusion:** The study concludes that although statistically significant differences existed between WAEC and NECO performance in some years and subjects, the differences were generally small in practical terms. This suggests a substantial level of functional comparability between the two examination bodies. The findings support the continued use of WAEC and NECO results as broadly equivalent credentials for certification, admission, and employment purposes. However, they also highlight the need for sustained standardization, quality assurance, and closer coordination between both examination bodies to strengthen the validity, reliability, and fairness of senior secondary certification in Nigeria.

**Keywords:** WAEC, NECO, Academic Performance, Mathematics, English Language, Assessment Comparability

### Introduction

Assessment is a component of the education systems in various countries worldwide, as a structured manner of establishing what learners know, understand, and can do. Besides its direct classroom functions, assessment has accountability, certification, selection, and placement functions that affect the future of individuals both in academia and profession





(Elmore, 2019). In modern-day schooling, most particularly within a system that has been adversely affected by standardised testing, the final results of the assessment are usually absorbed as a gate-keeping tool to propel oneself further to the subsequent level of education and employment opportunities (Black and Wiliam, 1998). Among the matters that are important to the discussion on educational measurement are the sincerity, consistency, and fairness of the assessment systems.

On the international scale, mass-scale standardised tests are meant to provide consistency in assessment, enabling similar assessments to be made in different populations (Sireci, 2020). But the attainment of comparative effectiveness in assessment has become a long-standing problem, and more so when there are multiple examining bodies within the same educational arena. Newton (2007) points out that comparability is how much the scores collected in one test or assessment system could be interpreted in the same context when compared to other tests or assessment systems. Without strong comparability evidence, the inferences made on the basis of such scores are questionable (Kane, 2013). This is one area of concern, particularly where there are different examination bodies that may claim to be assessing similar content areas but may differ in presentation of the tests, grading, or moderation.

Nigeria has two large exam boards: the West African Examinations Council (WAEC) and the National Examinations Council (NECO), which carry out the senior secondary school certification system. The two bodies have a part to play in administering standardised exams that make secondary education worthwhile and eligibility to join tertiary institutions (Emmanuel et al., 2020). WAEC was established in 1952, and since then, it has been functioning in the member countries of West Africa and is traditionally believed to be the more established examination board (Oguguo et al., 2020). NECO was created in 1999 due to the perceived deficiencies in access opportunities to entry and exams in the Nigerian context (Agwu et al., 2020). Despite the different histories of institutions at both organizations, both assess the same curricula and deliver outcomes that are usually regarded as being equal regarding academic progression.

The assumption of the interchangeability of NECO and the WAEC results has taken very deep roots in the educational practice and policy in Nigeria. The two certificates can be equally accepted as an admission and selection tool by universities, employers, and other stakeholders. However, this alleged equivalence is of major concern within educational measurement. According to Wiliam (2010), two supposedly interchangeable systems of assessment that produce systematically different scores in similar groups of students are dubious on fairness, validity, and equity. Koretz (2008) contends that fairness in evaluation goes beyond the design of a single test to encompass consistency of measurement between parallel systems that are utilised to make similar decisions.

Empirical studies in assessment research have repeatedly shown that the differences in test construction, testing conditions, examiner expectations, and grading policies have the capability of influencing the student performance outcomes within examination institutions. Comparability in situations with such differences cannot be assumed and should be proven with the help of stringent statistical and psychometric analysis. The evidence on the differences in WAEC and NECO performance in the Nigerian context, however, is mostly anecdotal or of a generalist nature as opposed to being a systematic empirical study (Abdullahi and Oladosu, 2021). This limits the level at which the stakeholders can make informed decisions on the reliability of either of the examination systems.

This issue is more severe in high-stakes courses such as Mathematics and English Language, the compulsory requirements of admission to tertiary studies. These are not only





that are the basic measures of academic proficiency but also come to be the norms of cognitive and communicative abilities. These discrepancies in the results of performance of WAEC and NECO in these subjects may thus be largely dependent on the lack of students to receive the education. Nevertheless, educators, students, and the general population still have a perception that performance may be different in the two examination bodies, and some feel that one of them may be tougher or easier than the other. These beliefs, either true or false, can undermine the trust in the examination system, and the problem of standardization can be brought up.

Nevertheless, even though this issue is significant, there is a significant lacuna in localised empirical research, which directly compares the results of performance in WAEC and NECO in particular geographical settings in Nigeria. A lot of the literature available summarises the data at a national level, therefore covering up the differences that might occur at the contextual level across regions, school types, or groups of students. It is a focus of educational measurement literature that validity evidence must be context-sensitive and based on more than one level of analysis to draw strong conclusions (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014). In the absence of such localised evidence, the conclusions in regard to comparability are not complete and can be misleading.

Moreover, current studies tend to consider pass rates or overall performance patterns in general without disaggregating the data by subject area or analysing systematic variation in performance patterns. This gives a conceptual disjuncture to the interpretation of whether the differences seen are caused by the examination bodies per se or other intervening factors like the quality of instruction, preparedness of the students, or examination conditions. According to Braun (2005), the differences in performance in standardised testing can be a complicated combination of both measurement design and contextual factors, as opposed to the difference in ability of the students.

It is on this basis that the need to consider the comparability of the results of the WAEC and NECO results in a structured and context-specific way emerges. This research is thus aimed at the performance of students in WAEC and NECO exams in Mathematics and English Language in Jos South Local Government Area. Through the focus on these main areas, the research will aim to offer a better insight into the question of whether there is a significant difference in performance results between two examination authorities in a specific educational situation.

In an ideal scenario of a standardised assessment system, the various examination bodies undertaking the same curriculum should come up with the same and interchangeable outcomes for similar candidates. Within this kind of system, the stakeholders are assured that they can interpret scores without worrying about bias or variability among testing agencies. But when empirical results show that performance results are substantially different between WAEC and NECO, then the supposition of equivalence will be challenged, and policy interventions might be necessary to correct underlying factors. This is a case of what can be termed as an equivalence paradox, whereby the results of WAEC and NECO are treated as interchangeable due to the policy, but are often perceived as varying in terms of difficulty and credibility among students, educators, and the general populace (Olutola et al., 2025).

Although past research has made efforts to compare the performance of the WAEC with that of NECO, most have been narrow in their research and usually aimed at comparing the whole nation or one year of data. Others have not determined the statistical consistency of differences observed among subjects or contexts. This creates an important gap in the literature





on local, subject-specific comparative analysis that can better inform educational policy and practice.

Thus, the current research fills this gap by giving an empirical comparison between the performance of the students on WAEC and NECO Mathematics and English Language examinations in Jos South from 2019 to 2021. Jos South Local Government Area is an appropriate case study because it has both a mixture of public secondary schools and a regularity in taking both the WAEC and NECO exams, making it convenient for the study of localised patterns of assessment comparability in the Nigerian educational system. Of particular importance is the choice of the 2019-2021 period, which reflects the performance trends just before and during the COVID-19 pandemic, a time when teaching, administration of assessments, and standardisation of examinations were disrupted in the whole world (Korshunova et al., 2024). Therefore, the outcome of the study adds to the discussion on the comparability of assessment by providing evidence in a particular educational setting, which will lead to a better understanding of whether the two examination systems can be reasonably deemed as similar in practice. It is hoped that the results will enlighten education stakeholders such as policymakers, school administrators, and examination bodies about the challenges of standardization and fairness, as well as the validity of high-stakes examination results in Nigeria.

## Objectives

This study aims to determine whether significant differences exist in students' academic performance in WAEC and NECO in Mathematics and English Language.

## Literature Review

The issue of comparability of large-scale examination systems has continued to be a focal point in educational assessment, especially where there are several examining agencies that determine similar amounts of academic accomplishment. In Nigeria, parallel Senior School Certificate Examinations (SSCE) are administered by the West African Examinations Council (WAEC) and the National Examinations Council (NECO), and are both essential qualifications to higher education and to securing a job. Although these examinations are formally equivalent, there has been concern that the academic achievements of students in such examinations can be used interchangeably without jeopardising fairness and consistency in high-stakes decision-making.

In educational measurement, comparability is how scores on one assessment can be interpreted in a similar manner when measured in different assessments, when they are intended to measure the same construct, and when the measurement is conducted under similar conditions (Kolen and Brennan, 2014). Practically, this assumption is usually contested by discrepancies in the ways of administration, scoring, and contextual aspects, which can be linked to the differences in the performance results of students.

Empirical research of WAEC and NECO offers valuable information on how much these expectations are fulfilled. Structurally, there is a study indicating that there is a reasonable level of congruence between the two examination bodies. According to Bandele and Adewale (2013b), the reliability and validity of WAEC and NECO exams are acceptable, meaning that they are comparable in terms of psychometric quality. Their comparison of the item difficulty also showed that no significant differences between examination bodies were present, hence neither examination is more challenging than the other. Likewise, Udofia and Udoh (2017) observed that WAEC and NECO Mathematics examinations have similar distributions of items





in cognitive domains and syllabus topics and are indicative of the following related curricular frames.

Nonetheless, the picture is more complex and inconsistent based on the evidence provided by the academic performance of students. Abdullahi and Oladosu (2021) reported that there was a positive correlation between the performance of students in WAEC and NECO, but there were consistent worries among various stakeholders about whether they were equivalent. The research also showed that the preference of students to take one test rather than the other could be linked to their belief in the difficulty and scoring criteria. On the same note, Egbunu and Penda (2022) have found differences in the academic performance of students in both WAEC and NECO Biology examinations, implying that there is a possibility of the differences being attributed to contextual factors like the quality of teaching, school facilities, and examination conditions.

On the topic of English Language, Akanya and Omachonu (2019) note that despite the fact that the syllabi of the WAEC and NECO are largely similar, some aspects of assessment might not be able to adequately test the communicative competence of students. This is indicative of the fact that variations in both implementation and assessment practices could be linked to differences in learning outcomes.

This difference between structural consistency and variability in the performance of students points to the intricacy of examination comparability. The similarities in the test design do not always imply equal results, especially in areas where there are varying educational circumstances (Creswell, 2014).

Critical analysis of literature indicates that there are some limitations. Most researchers use ex-post facto designs that have a relatively small or regionally localised sample, thus restricting the generalisability of their results. Some of them concentrate on individual subjects without considering various variables, leading to biased evaluations of comparability. Moreover, there has been very little focus on the difference in performance with time by different examination bodies, and also context-specific analysis of various localities in Nigeria. Moreover, there is a lack of research that combines both structural and performance-based approaches to give a holistic view of comparability. This leaves a loophole in the explanation of how similarities in examination design can be associated with differences in the observed outcomes.

Hence, context-sensitive research studies are required that look at the academic performance of students in a variety of core subjects and through a variety of examination cycles. This research fills this gap by offering a comparison of the performance of WAEC and NECO in Mathematics and English Language in Jos South, Nigeria. This disparity highlights the importance of context-dependent and multi-subject analyses to gain a more accurate idea of the character of examination comparability in various educational contexts.

### **Psychometric vs Empirical Comparability**

An important difference between psychometric comparability and empirical comparability is a key difference in assessment research. Psychometric comparability is the similarity in test design, reliability, validity, and content coverage, which are constructed to measure the same constructs under similar technical conditions (Bandeale & Adewale, 2013b; Kolen & Brennan, 2014). In contrast, empirical comparability is concerned with whether these similarities have the same outcomes in performance when the tests are given to actual groups of students. This difference is significant in terms of validity, with similar test designs not necessarily implying similar score interpretations (Kane, 2013).





The Nigerian contextual evidence indicates that, despite the significant overlap in the coverage of the syllabus and psychometric qualities of WAEC and NECO, there remains a difference in the performance outcomes of students as observed in different examination bodies. This implies that psychometric equivalence does not necessarily lead to empirical equivalence, especially when there are differences in conditions of administration, grading procedures, and preparation of candidates that can be used to affect results (Linn, 1993).

Moreover, most of the current research is based on national aggregates or yearly datasets, which restricts them in the ability to measure both temporal changes and localised differences in performance outcomes. These methods blur the dynamic quality of comparability and decrease analysis sensitivity to local variations. This shortcoming highlights the importance of multi-year, context-based research involving a combination of psychometric and empirical attitudes, as is being done in the current research.

The debate surrounding comparability in educational assessment has been further complicated by recent developments in educational assessment, especially in regard to the struggle between standardisation and contextualisation in the digital age. Standardisation focuses on consistency in test construction, administration, and scoring to promote fairness and comparability among the candidates, and contextualisation acknowledges the effect of local learning conditions, resources, and learner attributes on the performance in the test (Canché et al., 2025). Practically, it is difficult to strike a balance between the two principles, particularly in the developing setting where the inequality in the quality and teaching materials used in schools is acute.

The context of the Nigerian setting further complicates the issue of validity and comparability of examination results with the aspect of examination malpractices. Agwu et al. (2020) emphasized that the number of so-called miracle centres and other malpractices can skew the performance results, thus devaluing such large-scale examinations as WAEC and NECO. In some situations, this artificially raises pass rates, creating apparent performance differences that are not always indicative of true academic ability. On a validity side, it brings about issues of whether a difference in examination bodies can be ascribed to actual variation in the standards of assessment, or to contextual anomalies of examination administration.

These new concerns imply that comparability cannot be considered as a technical characteristic of a test design, but rather as a role of a more systemic and contextual nature. Therefore, it follows that any empirical analysis of the performance of WAEC and NECO has to be viewed in the broader context of education within which these examinations are conducted.

### **Theoretical Foundations of Assessment Comparability**

The concept of assessment comparability is based on the theory of validity, whereby test scores are expected to be reflective of the constructs they are supposed to measure and make reasonable interpretations and applications (Messick, 1989; Kane, 2013). Comparability, in this case, is the issue of whether the scores using different examination systems can be construed to reflect the same levels of student achievement.

Next to the notion of validity, there is the notion of comparability, which is the extent to which scores of various assessments are interpretable in a similar manner (Newton, 2007). Comparability is critical in systems where there are more than two examination bodies whose purpose is similar to certify education, in order to facilitate fairness in the education decision-making process. When the results of examinations cannot be compared, students can be either advantaged or disadvantaged depending on the body of examination that they are taking.





Classical test theory can give further insight into this problem since it conceptualises observed scores as a combination of true scores and measurement error. Any disparities in performance outcomes can thus be attributed to differences in the test difficulty or reliability or in scoring procedures (Linn, 1993). Building on this line of thinking, the generalisability theory emphasises that various sources of measurement error, such as items, raters, and testing conditions, may affect observed scores. During massive tests like the WAEC and NECO, irregularities in all these aspects can compromise the consistency of the outcomes.

The theory of fairness also offers a significant perspective on comprehending comparability. In a fair assessment, all candidates should be assessed in an environment that does not favour or discriminate against a group of people systematically (Koretz, 2008). Variations in the performance trends among the examination bodies could therefore be a matter of concern regarding fairness and uniformity in the evaluation processes.

Moreover, the construct equivalence concept stresses that various examinations ought to assess the same underlying competencies to an equal degree of cognitive difficulty. In other subjects like Mathematics and English Language, this means that not only will you have some content that is similar in the curriculum, but it is also the structure and the complexity of the tasks set in the assessment. Differences in the manner in which constructs are operationalised can be linked to differences in the performance outcomes of the students.

Collectively, the above theoretical views indicate that assessment comparability is a multidimensional concept comprising validity, reliability, fairness, and construct alignment. They also highlight the difficulties in attaining parallelism in examination systems. These theories form a framework for examining whether observed differences in WAEC and NECO performance are associated with variations in assessment practices rather than inherent differences in students' abilities.

## Conceptual Framework

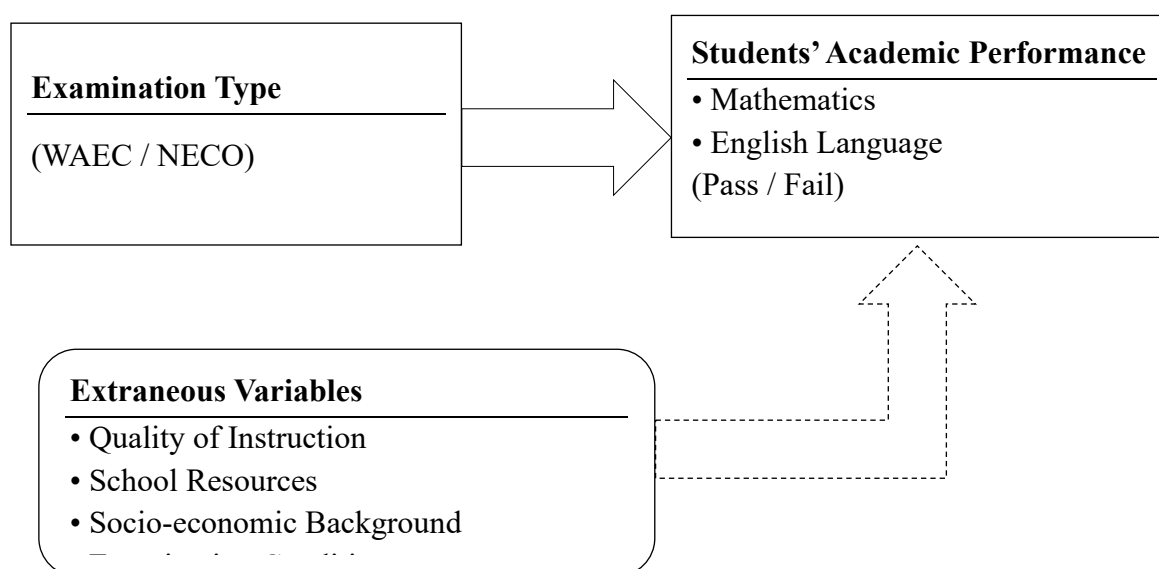


**Figure 1:** Conceptual model showing the relationship between examination type (WAEC and NECO) and students' academic performance in Mathematics and English Language, as well as some extraneous variables which may be confounders.





This conceptual framework shows the correlation between the type of examination and the academic performance of students in chosen subjects. The model recognises examination type, in this case, the West African Examinations Council (WAEC) and the National Examination Council (NECO) as the grouping variable and the academic performance of the students in Mathematics and the English language as the dependent variables, each operationalised as a dichotomous variable with two categories: Pass and Fail.

The model is based on the assumption that despite the similarity between the content of the curriculum covered by both the WAEC and the NECO tests, differences in their assessment processes can be linked with the observed performance by students. These could be due to variations in test construction, difficulty, marking scheme, and standardisation procedures. Such dissimilarities can cause differences in the level of pass and pass rates among the students in the two systems of examination.

The outcome variables are Mathematics and English Language because they are core subjects in senior secondary and because they are used in determining eligibility for admission into tertiary education. The model presumes that the performance of students in these subjects is related to the examination system in which the students are evaluated.

The conceptual relationship thus indicates that, on the assumption that WAEC and NECO are completely comparable, there should not be a significant difference in the performance of students in the two examination systems. Nevertheless, the evident variations in the performance results suggest the possible problem of the equivalence of the examination bodies. This framework offers the foundation of empirically testing whether academic performance is significantly associated with examination type.

Though not directly quantified in this study, the framework recognizes the existence of extraneous variables like instructional quality, school resources, socioeconomic background, and examination conditions that could influence the variations in the performance outcomes among students between the types of examination. The fact that they are included enhances the explanatory capabilities of the model and demonstrates the complexity of the system of educational assessment.

## Methodology

### Data Source

The data used in this research were obtained through secondary sources, which were examination records of students of some chosen public secondary schools in Jos South Local Government Area of Plateau State, Nigeria. In deciding Jos South, the availability of consistent school records and the availability of students who had taken both WAEC and NECO exams during the same period informed the decision. The research targeted the performance of two main subjects (Mathematics and English Language), which are mandatory subjects for a senior secondary school certification and university entry in Nigeria.

The database was from 2019 to 2021. This timeframe was chosen to give an adequate scope of observations to compare and analyse them in various examination cycles, but retain the relevance and completeness of data. The target population comprised students in senior secondary schools who took both the WAEC and NECO examinations within the given timeframe. The data involved the grades of students as reported by the schools, and performance was dichotomised into the two categories of performance based on grading. Those students who scored A1 to C6 were coded Pass, and those who scored D7, E8, or F9 were coded Fail. This dichotomisation of grades into pass or fail categories was done in line with the Nigerian University admission requirement that a credit or above is taken as Pass in





Mathematics and English Language, and any grade below a credit is Fail (Agbaire, 2021). This approach does not reflect the variations in the quality of grades, such as A1 vs C6, which is recognised as a limitation, but it is perfect for Chi-square analysis. The sample consisted of 600 students per year and was deemed sufficient, as it meets the conditions of minimum expected cell frequencies and improves the statistical power of the Chi-square test (Pallant, 2020). In the sampling of the study period, which was the secondary schools within Jos-South LGA of Plateau State, a purposive sampling method was used in picking up 10 secondary schools that had full and available WAEC and NECO records. In the chosen schools, a census of available cases was achieved by enrolling all the able students who had taken both examinations.

### **Instrument for Collecting Data**

A data extraction template was created, and it was utilised as the main tool for gathering the corresponding data in the school records. The researcher created the template to provide uniformity, accuracy, and consistency in the extraction of data in various schools. It included the fields of recording the demographic identifiers of students (coded to provide anonymity), the type of examination (WAEC or NECO), subject area (Mathematics and English Language), year of examination (2019, 2020, and 2021), and the performance results.

The results of performance were classified into pass and fail according to the grading criteria used by the WAEC and NECO in the years considered. The instrument also incorporated coding instructions to standardise responses and minimise ambiguity when entering the data. To strengthen the validity, the template was checked by educational measurement and evaluation experts to make sure that it was able to measure all the variables that were pertinent to the study objectives. Their response was used to optimise the instrument, especially with regard to the readability of classification and correspondence to examination grading frameworks.

### **Data Collection Process**

The collection of data was done in a well-organised and ethical fashion. The first step that the researcher undertook was to seek official permission from the school authorities of the sampled secondary schools in Jos South Local Government Area. This measure was made to adhere to institutional policies on access to the academic records of students and also uphold ethical standards in the conduct of research.

After agreement, school examination officers and record personnel were involved to determine and collect relevant WAEC and NECO result records of the given years. The researcher collaborated with these staff to ensure proper identification of the students who had both taken the examinations. The records of all qualified students were closely compared between the WAEC and NECO databases, with the unique identifiers given by the schools, and a high level of confidentiality was maintained by anonymisation of the personal data.

A structured template was used to extract the data manually to maintain consistency. All records were checked during extraction to prevent duplicate and lost records. Where incomplete records were observed, they were not included in the dataset to maintain the integrity of the analysis. The data collected were anonymised to maintain the confidentiality of students. Ethical permission was provided by the relevant school authorities. No identifying information was collected or reported. The whole process was implemented within a specified duration so as to have a systematic coverage of all the chosen schools, as well as reducing errors caused by haste in data collection.





## Data Analysis

The data collected were analysed using both descriptive and inferential statistical techniques. The performance of students in WAEC and NECO examinations in Maths and English Language was summarised using descriptive statistics; that is, frequencies and percentages, per year under review.

The Chi-square ( $\chi^2$ ) test of independence was used for inferential analysis to show whether there were significant differences based on examination type (WAEC and NECO) and performance outcome (pass and fail) of students. The assumptions of the chi-square test were satisfied, including adequate expected cell frequencies. Chi-square test was found to be the right test because the variables were categorical in nature.

Along with the statistical significance testing, the effect sizes were also examined with reference to calculating Cramer's V in order to identify the strength of the relationship between the type of examination and performance outcomes. This was needed since statistical significance does not reveal the size of observed differences. The following range of Cramer's V values is used for the interpretation of effect sizes. 0.00 to 0.10 (very small), 0.10 to 0.29 (small), 0.30 to 0.49 (moderate), and 0.50 and above (large) (Cohen, 1988).

The testing of all hypotheses was done at the level of 0.05. The p-value of less than or equal to 0.05 was used to reject the null hypothesis, whereas a p-value of more than 0.05 was used to retain the null hypothesis. Both the significance testing and the estimation of the effect size were included to make the interpretation of the results more complete.

Both descriptive and inferential approaches made it possible to clearly present the performance patterns as well as thoroughly analyse the differences between the results of WAEC and NECO. This method of analysis gave a strong foundation for tackling the research objective and drawing significant conclusions regarding the comparability of assessment.

## Results

### Result 1: Performance Differences in 2019

Table 1: WAEC and NECO Performance in Mathematics and English Language (2019)

| Subject          | Exam | Pass n (%) | Fail n (%) | $\chi^2$ | df | p-value | Cramer's V |
|------------------|------|------------|------------|----------|----|---------|------------|
| Mathematics      | WAEC | 162 (54.0) | 138 (46.0) | 1.738    | 1  | 0.187   | 0.054      |
| Mathematics      | NECO | 178 (59.3) | 122 (40.7) |          |    |         |            |
| English Language | WAEC | 150 (50.0) | 150 (50.0) | 4.196    | 1  | 0.041   | 0.084      |
| English Language | NECO | 175 (58.3) | 125 (41.7) |          |    |         |            |

In Table 1, the results show that students performed better in NECO compared to WAEC in both Mathematics and English Language in 2019. NECO in both subjects performed better in passing as compared to WAEC, with a similar trend across exam boards.

In the case of Mathematics, the test provided a  $\chi^2 (1) = 1.738$  with  $p = 0.187$ , thus showing that there is no significant difference between the results of the WAEC and the NECO, as well as a very negligible effect size (0.054). However, in English Language,  $\chi^2 (1) = 4.196$ ,  $p = 0.041$  indicates that there is a significant difference between the two examination bodies. The magnitude of the associated effect, in terms of Cramer V values for the English language, is 0.084 (very small). This suggests that although the difference is statistically significant, the magnitude of this difference is weak. In a realistic sense, the examination type accounts for less than 1% of the difference in the performance of students.





On the whole, NECO seems to provide somewhat higher performance results in comparison to WAEC in 2019, especially in English Language, but the effect size is small, which means that the practical impact of such differences is minimal.

## Result 2: Performance Differences in 2020

Table 2: WAEC and NECO Performance in Mathematics and English Language (2020)

| Subject          | Exam | Pass n (%) | Fail n (%) | $\chi^2$ | df | p-value | Cramer's V |
|------------------|------|------------|------------|----------|----|---------|------------|
| Mathematics      | WAEC | 140 (46.7) | 160 (53.3) | 13.594   | 1  | 0.000   | 0.151      |
|                  | NECO | 185 (61.7) | 115 (38.3) |          |    |         |            |
| English Language | WAEC | 145 (48.3) | 155 (51.7) | 8.224    | 1  | 0.004   | 0.117      |
|                  | NECO | 180 (60.0) | 120 (40.0) |          |    |         |            |

Table 2 shows that there is a steady trend of improved performance by the student in NECO than in WAEC in both Math and English Language in 2020. In both subjects, NECO has reported superior performance results, which implies that there is a systematic difference between the two examination bodies.

Chi-square analysis proves the fact that these differences are significant. In the case of Mathematics,  $\chi^2(1) = 13.594$ ,  $p = 0.000$  implies that there is a significant difference in performance between NECO and WAEC. Likewise, in the case of English Language,  $\chi^2(1) = 8.224$ ,  $p = 0.004$  indicates that there is a statistically significant difference in the two exams. Nevertheless, these differences are not huge, which is indicated by the Cramer V values (0.151 in Mathematics and 0.117 in English Language). This indicates that despite the fact that examination type is related to variations in student performance, its practical effects are minimal. This means that in 2020, examination type only accounts for about 1% and 2% of the variance in performance in Mathematics and English Language, respectively, due to the low V values.

Generally, the 2020 results support the trend of the last year, according to which NECO candidates seem to be more successful than their WAEC counterparts. However, the small effect sizes imply that the differences observed are statistically significant but might not represent substantial practical differences in performance.

## Result 3: Performance Differences in 2021

Table 3: WAEC and NECO Performance in Mathematics and English Language (2021)

| Subject          | Exam | Pass n (%) | Fail n (%) | $\chi^2$ | df | p-value | Cramer's V |
|------------------|------|------------|------------|----------|----|---------|------------|
| Mathematics      | WAEC | 170 (56.7) | 130 (43.3) | 0.027    | 1  | 0.869   | 0.007      |
|                  | NECO | 168 (56.0) | 132 (44.0) |          |    |         |            |
| English Language | WAEC | 165 (55.0) | 135 (45.0) | 0.169    | 1  | 0.681   | 0.017      |
|                  | NECO | 170 (56.7) | 130 (43.3) |          |    |         |            |

Based on the statistics reported in Table 3, the performance of students in WAEC and NECO in both Mathematics and English Language in 2021 was mainly similar. The performance trends of the two examination bodies are observed to be in a near-synchronous manner, indicating little difference between the two.

The Chi-square test shows that these differences are not statistically significant. In the case of Mathematics, the  $\chi^2(1) = 0.027$ ,  $p = 0.869$ , which means that there is no significant difference between the results of WAEC and NECO. Likewise, in English Language,  $\chi^2(1) = 0.169$ ,  $p = 0.681$  indicates that the difference is not statistically significant.

Moreover, the effect sizes are insignificant, with the Cramer V values of 0.007 and 0.017 for Mathematics and English Language, respectively. This suggests that the type of examination body has virtually no relationship with students' performance in 2021.

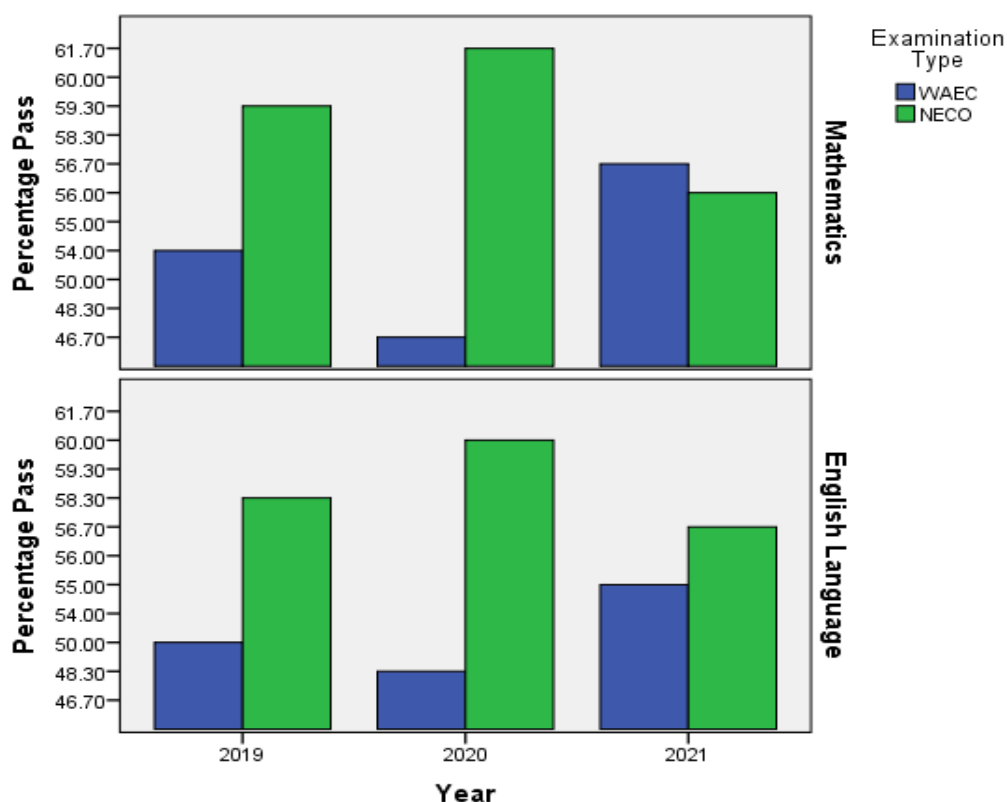


**Figure 2:** Three-Year Trend of WAEC and NECO Percentage Pass Rates in Mathematics and English Language

In general, a year-on-year analysis for three years presented in Figure 2 reveals that although NECO's pass rate was consistently a bit higher than WAEC, the difference between the two contracted over time, and in 2021, they performed almost equally, which suggests a convergence in performance outcomes across the two examination bodies.

## Discussion

The results of this research showed that comparability of the outcomes of the WAEC and NECO examinations in Mathematics and English language changes over the three years under study. In particular, significantly different scores in 2019 (English Language) and 2020 (Mathematics and English Language) were found between the two examination bodies, but not in 2021. But the low effect sizes exhibited over the years indicate that, despite the statistically significant differences, the practical role of examination type on the performance of students is



low. The significance of this discovery cannot be underrated, as it suggests that the perceived difference between WAEC and NECO could be exaggerated in real-life situations because statistically, there are differences. Comprehensively, this trend suggests that the equivalence between the performance outcomes of the WAEC and the NECO is not a timely phenomenon but has an up-and-down fluctuating nature that depends on the circumstances of context and system. This dynamic inconsistency is consistent with educational measurement issues that comparability is not a permanent property of a system of assessment, but a state that needs to be continually achieved and sustained by regular processes of standardisation (Kolen & Brennan, 2014).

The high disparities witnessed in 2019 mean that the performance of students varies between the examination bodies in which they were assessed, with NECO always registering higher pass rates in Mathematics and English Language than in WAEC. However, it can be argued that due to the small magnitude of difference, it is crucial to interpret them with some caution, as they might not reflect differences in the actual academic ability of students. The variation observed could be due to the differences in the level of difficulty, standards, or administration of the exam by the two bodies in that year. On the measurement front, this stands to support the view of Newton (2007) that comparability is only met when the various tests give the same interpretation of performance. In the cases where such equivalence does not exist, the meaning of scores is not interchangeable but can be determined within the context. It also compares to the structural analyses described by Bandele and Adewale (2013a), which, however, conclude the lack of significant differences between item difficulty and stipulate that WAEC and NECO exams are similar regarding the technical quality. This deviation is indicative of a crucial difference between comparability at the level of test construction and comparability at the level of performance outcomes.

The same trend was recorded in 2020, as there was once again a wide gap in the two subjects, with NECO recording higher pass rates compared to WAEC. However, in 2019, the differences in question were rather small, which implies that there were no practical implications in spite of statistical significance. This supports the evidence that statistical findings are to be interpreted with effect sizes to prevent overgeneralising the true implications of such results. The fact that these differences in two successive years remain to this point indicates that the variation seems to exist in a consistent manner between years, although perhaps small. This observation is not surprising given the fact that, according to Linn (1993), test-design variations, administrative processes, and scoring criteria can result in differences in scores when tests are made using the same curriculum. It is also consistent with the results of Abdullahi and Oladosu (2021), who found that despite positive correlations between the results of both tests, the comparability of the two tests is still doubted by stakeholders because of the perceived differences in the difficulty of the tests and grading systems. On the same note, Egbunu and Penda (2022) noticed differences in the academic performance of students in WAEC and NECO Biology exams but credited their findings to the contextual and teaching priorities. Combined, these studies support the perspective that the issue of comparability is not isolated to the design of the test itself but is much more holistic in terms of the educational and administrative settings.

Moreover, the differences in 2019 and 2020 could be related to candidate perceptions and strategic behaviour. Abdullahi and Oladosu (2021) report that there are situations where students select NECO because of the perception that it is not as demanding or is more practical in marking. Although such perceptions could have an impact on preparation strategies and motivation, small effect sizes observed in this study provide indications that these behavioural





variables may not have an enormous difference in actual performance outcomes. This presents a significant nuance to the comparability debate that evaluating perceived differences among examining organizations can be greater than empirically measured impacts.

The 2021 performance is, however, a different tale where there are no considerable differences in Mathematics or English Language between the WAEC and NECO. The performance data of that year was very much the same, and this implies that there was convergence of the two examination systems. This result proves the hypothesis that parallel examination bodies can be compared using certain conditions. It can be an indication of better standardisation procedures, more appropriate matching of the test specifications, or better moderation and marking methods. This overlap is consistent with structural comparability, which was found by Udofia and Udoh (2017) to have parallel distributions of items in cognitive domains and syllabus content in physics and biology tests based on NECO and WAEC. It also gives empirical support to the argument that Kolen and Brennan (2014) put forward, where they state that when the assessment system is consistent in its design and implementation, then equivalence is possible.

This convergence in 2021 can be partially attributed to the influence of the COVID-19 pandemic that shocked the educational systems of all countries and caused changes in the way examinations were conducted, the system of their grading, and additional integration of the activities of different examination authorities. The pandemic in Nigeria required changes in examination time, marking, and quality assurance activities, which might have facilitated increased alignment of the WAEC and NECO results in this time frame.

The aspect of comparability that is dynamic also comes out in the various outcomes of the three years. Validity, as argued by Messick (1989), is not just a quality of the test instrument but also inferences drawn from the test scores under specific circumstances. It is worth noting that the small effect sizes found over the course of the study suggest that even in instances where there are differences, they would most likely not have any substantial impact on what student achievement means. The variability observed in this research does not only signify the mere statistical differences but also the situational influences such as administering tests, marking consistency, school preparedness, and even the socio-economic status of a particular situation. Creswell (2014) also had a similar point that educational outcomes are influenced by various variables interacting in multiple ways, and the inability to consider them can result in making incomplete conclusions regarding the equivalence of assessment.

The mixed results of this study, when put in the context of the overall literature, indicate the current controversy about the similarity of parallel examination systems. Despite some researchers, such as Bandele and Adewale (2013a) and Udofia and Udoh (2017), indicating that the level of the test construction can be considered comparable, other researchers, such as Abdullahi and Oladosu (2021) and Egbunu and Penda (2022), refer to the discrepancies in the outcome of student performance. The current research paper expands this debate by showing that even though statistical differences can be present, they have little practical implications because they have small effect sizes. Hence, the two perspectives of the literature taken separately are not complete.

Such results have significant policy and practice implications for the assessment practice and policy in education in Nigeria. Because the outcomes achieved by the WAEC and NECO are used in certification and admission to tertiary institutions, there are discrepancies in the performance outcomes, which raise concerns of fairness and equity in the selection. However, the fact that the differences are shown to have small effect sizes is, to some degree, encouraging that such differences may not be harmful to the candidates in a real-world scenario. This





subtlety is important because it indicates that no matter how much is done to enhance comparability, the issue of huge differences might be exaggerated. This is close to the same issue that Messick (1989) raised, that the validity of the assessment systems is based on the relevance and reasonableness of using and interpreting the scores.

In addition to the statistical aspects, the results also have implications for the social validity of the examination systems. The continued belief among students, teachers, and the general population that WAEC is a more challenging examination than NECO could affect examination preferences and practices. Nonetheless, the effect sizes in this study have been small all the time, which points to the fact that such perceptions might not be empirically backed strongly. The perceptions need to be addressed to enhance the public trust in the equality and fairness of both examination systems. Moreover, the convergence that was observed in 2021 indicates that it can be improved by a conscious effort towards standardisation. This implies that policy interventions to increase the congruence of coordination between WAEC and NECO, such as the harmonisation of test specifications, alignment of marking schemes, and common moderation processes, can lead to greater comparability and reduce disparities in outcomes. The findings hence support literature demands that there is a need to constantly check and review large-scale examination systems so as to attain consistency and equity.

Furthermore, the time gap that was apparent in this experiment gives a good reason to think that a longitudinal approach should be considered when assessing the comparability of exams. Conclusions drawn based on a single year of data may be misleading, as shown in the figure above, since comparability may be inconsistent over time. In the meantime, the general minimal large effect sizes, which are observed steadily over the years, justify the argument in favour of the necessity to complement statistical significance testing with tests of practical significance in educational studies.

By and large, the paper demonstrates that, despite the fact that the WAEC and NECO exams are supposed to be used to measure the same academic constructs, the results of both tests are not necessarily equal over time. However, the magnitude of these differences is not usually significant, which means that there is not much practical impact on the performance of students. The results have empirical evidence that comparability is a dynamic and context-dependent phenomenon that is affected by structural and contextual factors. The research assists in the creation of a more complex vision of what examination comparability is and why all the differences between the two systems cannot be eliminated, and why a lot needs to be done to make the two systems closer to each other to ensure that the evaluation of student achievement might be credible and fair.

### Knowledge Contribution

The study is relevant to the increasing literature on the comparability of educational assessment since it offers localised empirical data on the outcomes of the WAEC and NECO examinations in Mathematics and English Language. In contrast to most of the past research, which may use national aggregates or generalised assumptions, the study is based upon a particular Nigerian situation, providing a more contextualised understanding of how parallel examination systems actually work. These results add to the current body of literature by demonstrating that, despite the observed variance in performance of students between WAEC and NECO, over the years and subjects, the levels of these variances tended to be small, indicating that there was little practical difference in the performance of students between the two examination systems.





Theoretically, the research enhances the current frameworks of assessment validity and comparability by showing the relationship between the variance in measurement systems and the interpretation of student achievement scores. It supports the opinion that comparability is not a natural characteristic of examinations but an empirical assertion that must be continually tested by a variety of types of evidence, including both measures of statistical significance and measures of effect size. This is in line with the modern theory of validity that focuses on the combination of various sources of evidence to aid in the interpretation of scores and their application.

In practice, the research has significant implications for educational stakeholders, such as policymakers, examiners, school administrators, and higher education institutions. The results indicate that there are some slight differences in performance, but they might not be significant enough to justify high levels of preferential treatment to either of the examination bodies. This fact can be used to streamline the process of standardisation, better moderate practices, and to improve the alignment of WAEC and NECO assessment systems. The study helps to foster justice, credibility, and transparency in high-stakes tests and encourages more equitable certification and admission judgments by putting the statistical results in the context of practice. The findings, therefore, empirically support the use of WAEC and NECO results as synonymous by tertiary institutions' admissions boards due to the small effect sizes found over time.

### Recommendations

In line with the findings of this study, to enhance comparability and fairness in the results of the WAEC and NECO examinations, the following recommendations are offered.

Examination bodies (WAEC and NECO) ought to further intensify the cooperation in developing the tests, standardisation, and marking schemes to ensure the comparability of the results in the two examinations by forming a joint WAEC- NECO marking and moderation committee, which convenes twice a year to harmonise marking requirements and scoring rubrics.

The policymakers are encouraged to facilitate the creation of joint moderation and quality assurance systems to enhance uniformity in the level of difficulty, the content to be covered, and the grading criteria among the examination bodies.

The secondary schools are to implement uniform instructional delivery and examination preparation methods in order to reduce the small differences in performance of students through the WAEC and NECO.

Students should be sensitised through guidance and counselling units about the best preparation strategies that can be used in both the WAEC and NECO examinations, and these should be geared towards ensuring that the performance outcomes are stable. This can be achieved by creating a collective WAEC-NECO student preparation guide to harmonise examination preparation strategies.

Future studies must broaden the study to other fundamental topics and across wider geographic regions, and use longitudinal designs or more sophisticated methods of analysis to further explore comparability trends over time.





## References

- Abdullahi, M. S., & Oladosu, A. G. A. S. (2021). Relationship between students' learning outcomes in WAEC and NECO Arabic language examinations in North-Central Nigeria. *International Journal of Islamic Educational Psychology*, 2(2), 127–141. <https://doi.org/10.18196/ijiep.v2i2.12324>
- Agbaire, J. (2021). *Accessing higher education in Nigeria: Everyday experiences of "fair" admissions policy as an arena of struggle* (Doctoral dissertation, University of Sussex). University of Sussex Research Repository. [https://sussex.figshare.com/articles/thesis/Accessing\\_higher\\_education\\_in\\_Nigeria\\_everyday\\_experiences\\_of\\_fair\\_admissions\\_policy\\_as\\_an\\_arena\\_of\\_struggle/23479505?file=41188580](https://sussex.figshare.com/articles/thesis/Accessing_higher_education_in_Nigeria_everyday_experiences_of_fair_admissions_policy_as_an_arena_of_struggle/23479505?file=41188580)
- Agwu, P., Orjiakor, T., Odii, A., Onalu, C., Nzeadibe, C., & Okoye, U. (2020). *Nature and drivers of "miracle examination centres" in private schools in Nigeria: A systematic review of literature on examination malpractice* (ACE Research Consortium Working Paper No. 26). <https://ace.soas.ac.uk/wp-content/uploads/2022/05/WP26-Miracle-Examination-Centres-in-Nigeria.pdf>
- Akanya, J., & Omachonu, C. G. (2019). Effects of the West African Examinations Council (WAEC) and the National Examinations Council (NECO) oral English syllabi on secondary school leavers' spoken English in Kogi State, Nigeria. *British Journal of Education*, 7(2), 105–120. <https://ejournals.org/bje/vol-7-issue-2-february-2019/effects-of-the-west-africa-examination-council-waec-and-the-national-examination-council-neco-oral-english-syllabi-on-secondary-school-leavers-spoken-english-in-kogi-state-nigeria/>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association. <https://www.testingstandards.net/open-access-files.html>
- Bandele, S. O., & Adewale, A. E. (2013a). Comparative analysis of the item difficulty levels of WAEC, NECO, and NABTEB mathematics achievement examinations. *Mediterranean Journal of Social Sciences*, 4(2), 761–765. <https://doi.org/10.5901/mjss.2013.v4n2p761>
- Bandele, S. O., & Adewale, A. E. (2013b). Comparative analysis of the reliability and validity coefficients of WAEC, NECO, and NABTEB examinations. *Journal of Educational and Social Research*, 3(2), 397–402. <https://doi.org/10.5901/jesr.2013.v3n2p397>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Braun, H. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Educational Testing Service. <https://www.ets.org/Media/Research/pdf/PICVAM.pdf>
- Canché, M. S. G., Zheng, K., Song, Y., & Liang, Y. (2025). Standardized testing for diverse talent identification: A framework to address geographical bias in standardized testing and increase diversity in college admissions in the post-affirmative action/race-neutral admissions era. *Research in Higher Education*, 66(2), 1–35. <https://doi.org/10.1007/s11162-024-09824-4>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203771587>





- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage Publications.
- Egbunu, R., & Penda, B. M. (2022). *A comparative study of students' academic performance and achievement in Biology WAEC and NECO in the Federal Capital Territory, Abuja, Nigeria. FUOYE Journal of Education*, 5(2).  
[https://fjed.fuoye.edu.ng/index.php/public\\_html/article/view/55](https://fjed.fuoye.edu.ng/index.php/public_html/article/view/55)
- Elmore, R. F. (2019). The future of learning and the future of assessment. *ECNU Review of Education*, 2(3), 328–341. <https://doi.org/10.1177/209653111987896>
- Emmanuel, O. O., John, A. E., & John, O. (2020). Analysis of WAEC and NECO performance of senior secondary school students in North Central Nigeria. *International Journal of Humanities and Social Science Invention*, 9(2), 49–55.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking* (3rd ed.). Springer.  
<https://doi.org/10.1007/978-1-4939-0317-7>
- Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Harvard University Press.
- Korshunova, O. V., Telegina, N. V., Petrova, L. A., Pashanova, O. V., Latysheva, V. V., & Prokopyev, A. I. (2024). The unified state examination in mathematics in urban and rural schools: The impact of the 2020–2021 pandemic. *Eurasia Journal of Mathematics, Science and Technology Education*, 20(2), Article em2403.  
<https://doi.org/10.29333/ejmste/14240>
- Linn, R. L. (1993). Linking results of distinct assessments. *Applied Measurement in Education*, 6(1), 83–102. [https://doi.org/10.1207/s15324818ame0601\\_5](https://doi.org/10.1207/s15324818ame0601_5)
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). American Council on Education/Macmillan.  
[https://www.ets.org/research/policy\\_research\\_reports/publications/chapter/1989/hwrq.html](https://www.ets.org/research/policy_research_reports/publications/chapter/1989/hwrq.html)
- Newton, P. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149–170.  
<https://doi.org/10.1080/09695940701478321>
- Oguguo, B. C., Agah, J. J., Ene, C. U., Acholonu, V. N., Azubuike, R. N., Okeke, M. A., & Agbo, L. P. (2020). Content validity of West African Examinations Council financial accounting questions. *International Journal of Learning, Teaching and Educational Research*, 19(6), 161–178. <https://doi.org/10.26803/ijlter.19.6.10>
- Olutola, A. T., Bagudu, A., & Enetairo, R. (2025). Analysis of linear equating in WAEC, NECO, and NABTEB senior school certificate multiple-choice economics examinations in North West, Nigeria. *Federal University Gusau Journal of the Faculty of Education*, 5(5), 177–187. <https://doi.org/10.64348/zije.2025176>
- Oyetunde, T. O. (2010). Educational standards in Nigeria: The state of the art. In T.O Oyetunde, J.S. Aliyu, M.P Haggai, J.M. Musa (Eds). *Improving educational standards in Nigeria. Perspectives, challenges & strategies*. Zaria. Institute of Education, Ahmadu Bello University.
- Pallant, J. (2020). *SPSS survival manual: A step-by-step guide to data analysis using IBM SPSS* (7th ed.). McGraw-Hill Education.





- Sireci, S. G. (2020). Standardization and UNDERSTANDardization in educational assessment. *Educational Measurement: Issues and Practice*, 39(3), 100–105. <https://doi.org/10.1111/emip.12377>
- Udofia, N. A., & Udoh, K. I. (2017). Comparative analysis of WAEC and NECO mathematics examination. *Mathematical Theory and Modelling*, 7(6), 41–52. <https://iiste.org/Journals/index.php/MTM/article/view/37688>
- Wiliam, D. (2010). What counts as evidence of educational achievement? The role of constructs in the pursuit of equity in assessment. *Review of Research in Education*, 34(1), 254–284. <https://doi.org/10.3102/0091732X09351544>

