ISSN No:-2456-2165

A Systematic Review of Differential Item Functioning of Educational Assessments in Africa

Corresponding Author; Tapela Bulala[1]
Botswana University of Agriculture and Natural Resources),
Botswana Tapela Bulala is a Senior Lecturer at Department of Agricultural Education and Extension.
He holds MEd in Educational Research and Evaluation. He is currently doing Ph.D.
(Measurement and Evaluation) at Cape Coast University, Ghana.

Eric Anane[2]
University of Cape Coast, Ghana
Eric Anane is an Associate Professor of Educational Assessment and Research Methodology in the College of Education
University of Cape Coast. His areas of expertise are psychometrics, cognition,
educational assessment, and curriculum & policy development.

Andrews Cobbinah[3]
University of Cape Coast, Ghana
Andrews Cobbinah is a Senior Lecturer in Educational Foundations Department. He holds PhD in Educational Research,
Measurement and Evaluation. His expertise incudes assessment in schools, critical thinking, and peer assessment skills training.
https://orcid.org/0000-0002-5315-659

Abstract:- DIF detection and correction are essential for fair and equitable assessment practices, especially when results are used for high-stakes decisions like college admissions. By identifying and addressing DIF, assessment developers can reduce the risk of discrimination or bias against certain groups. Using papers from 2013 to 2023, a systematic review of DIF in African educational assessments was conducted. To find and screen eligible papers, PRISMA 2021 guided the search. Qualitative synthesis was used to create a narrative summary of findings after data extraction. 15 papers met the criteria. Between 2013 and 2023, these English-language papers addressed educational assessment in Africa. WASSCE, BECE, SSCE, and BJCE exams were the focus of most studies. The most common DIF method was IRT. The groups favored in African educational assessments using Differential Item Functioning (DIF) varied by context, assessment, and subject. DIF has been found in African studies. Educational experts must critically assess African assessments and take steps to address these issues.

Keywords:- *Differential item functioning, Item response theory, educational assessment.*

I. INTRODUCTION

When constructing a measure, the constructor needs to assure that it measures the same way for different persons being measured, this is called measurement invariance (Bundsgaard, 2019). It means that the result of a test should not depend on anything else but the students' proficiency in the area the test is intended to measure. It should not matter what background the student comes from, or on the specific items used to test this specific student. An overarching objective in research comparing different sample groups is to ensure that the reported differences in outcomes are not affected by differences between groups in the functioning of the measurement instruments, i.e. the items have to work in the same way for the different sample groups to be compared. Lack of invariance across sample groups are commonly called Differential Item Functioning (DIF) (Hagquist, 2019). In educational and psychological testing, the term DIF 'means that the probability of a correct response among equally able test takers is different for various racial, ethnic, gender [or other] subgroups (Strobl, Kopf, & Zeileis, 2015). DIF can occur when an item is biased in favour of or against certain groups of test-takers based on their background or characteristics, such as gender, ethnicity, or language. Differential item functioning is a statistical technique in educational testing that examines whether an item in a test is functioning differently for different groups of test-takers. DIF is generally seen as a problematic phenomenon, i.e. as an indicator of item bias, and the solution is therefore often to remove items that show DIF, or to treat the items as not-administered for the groups where they showed DIF (Bundsgaard, 2019).

DIF can occur for various reasons, including differences in cultural background, language proficiency, and socioeconomic status. For instance, a test question that relies heavily on cultural references or language nuances that are unfamiliar to certain groups may be more challenging for them to answer correctly, even if they have the same level of knowledge or skills as other groups. Similarly, items that are related to certain life experiences or contexts that are more common among one group than another may also lead to DIF.

To detect DIF, researchers use various statistical methods that compare the performance of different groups on each item, while controlling for their overall ability or knowledge level. Most of these methods are based on the comparison of the item parameter estimates between two or more pre-specified groups of subjects, such as males and females, as focal and reference groups (Strobl, et al., 2015). These methods can be based on item response theory (IRT), logistic regression, or other approaches that estimate the probability of correct response for each group.

ISSN No:-2456-2165

McNamara and Roever (2006) classified the methods used to detect Differential Item Functioning (DIF) into four broad categories. The first category involves analysing item difficulty estimates to compare differences in item difficulty. The second category uses nonparametric methods, such as contingency tables, chi-square, and odd ratios. The third category is item-response-theory-based approaches, which use 1, 2, or 3-parameter analyses to compare the fit of statistical models. Finally, the fourth category includes other approaches, such as logistic regression, generalizability theory, and multifaceted measurement.

If a significant difference is found between the groups, further investigation is needed to determine the nature and extent of the DIF and its potential impact on the assessment results. Different methods to detect DIF seem to generate similar results. In addition to identifying DIF, researchers have suggested ways to minimize or eliminate its impact on assessment results. For instance, Mokobi and Adedoyin, (2014) suggest that test developers in Africa should always endeavour to create bias free items for testing and examination purposes and the connotations reflected in test or examination items should be relevant to the life experiences of examinees responding to the items whiles Annan-Brew, (2020) suggests that DIF studies should be conducted by test developers on their test so that the items exhibiting Differential Item Functioning (DIF) could be revised or eliminated to enhance fairness.

Despite the importance of DIF analysis for ensuring fair and valid assessment practices, there is limited knowledge about the extent and implications of DIF in African contexts, where cultural, linguistic, and socioeconomic diversity can create unique challenges and opportunities for assessment development and evaluation. Whiles there exist individual studies on DIF in Africa, researchers have not paid specific attention to the synthesis of such studies. Therefore, this systematic review aims to synthesize and critically evaluate the existing literature on DIF studies in Africa, with the goal of identifying types of assessments studied, methods used to detect and analyse DIF in Africa, findings of DIF for demographic groups and the recommended solutions to minimizing diff in Africa. This information can help assessment developers and practitioners in Africa to design and use assessments that are valid, reliable, and fair for all individuals, regardless of their cultural, linguistic, or socioeconomic background.

II. METHOD

A. Study Selection (Inclusion & Exclusion Criteria)

The methods for this review were based on guidelines developed by the preferred reporting items for systematic reviews and meta-analyses (the PRISMA statement) (Page, McKenzie, Bossuyt, Boutron, Hoffmann, Mulrow, & Moher, 2021). A systematic review of literature was conducted for studies that explored differential item functioning of educational measurement in Africa for the past 10 years (2013 to 2023) consistent with Palmatier, Houston, and Hulland's, (2018) assertion that the purpose of a review of literature is to provide an integrated, synthesized overview of the current state of knowledge.

The search for articles occurred between June 1st 2023 and July 5th 2023. To identify the studies included in this systematic literature review, a comprehensive search was conducted in several databases including using AJOL, Research Gate, JSTOR, Science direct databases and Google Scholar. The search was conducted using the following search terms: "differential item functioning in Africa", "differential item functioning of assessment in Africa". These wide search words gave more results which were narrowed by relevance and usefulness.

For instance, a search on Science Direct using "differential item functioning", yielded 172,024 results, however, upon adding the term in Africa, the result was reduced to 22,980. Furthermore, when the year range was restricted to 2013 to 2023, there were a total of 12,392 articles. Moreover when the articles were restricted to open access articles, there were a total of 1,894 articles. Furthermore, the search was restricted to research articles only and the results further reduced to 1,424 results. The titles of these articles were screened to identify relevant studies. Other search engines were also searched, titles and abstracts of articles were screened to identify relevant studies. Considering the databases searched, a total of 350 studies were identified and thoroughly screened to identify those that met the inclusion criteria. Also, reference lists of studies that were relevant were thoroughly examined to identify additional studies. This was done to ensure that the review is thorough as much as possible. Authors of articles were checked to ensure factually sound research and to prevent 'bad media' viewpoints (Gusenbauer & Haddaway, 2019).

Firstly, the researcher included studies with a title an English language-based abstracts. Reviewed articles were those that reported original research and that had unambiguous designs and methodologies, included study settings, specific assessment utilised, clear definitions of outcomes and study samples or populations. Studies were excluded if they were published before 2013, were not reported in English language, did not focus on students in academic contexts, or were related to health assessments. Also, studies that did not provide a clear methodology such as type of demographic studied, sample used, setting in which the study was conducted or inconsistent and ambiguous methodology and findings were excluded. Editorials, commentaries, or letters to the editor were excluded because they present opinions and judgments of the writers without presenting findings in relation to study populations. A total of 15 articles were included in the study.

B. Data Extraction, Quality Assessment & Data Analysis Procedure

Once eligible studies were identified, the researcher extracted data from them using a standardized data extraction form. The data extracted from the studies included the author and year of each study, study setting, the sample, method of DIF analysis, results of the analysis and recommendations for reducing DIF. After data extraction, the researcher synthesized the results using qualitative synthesis techniques.

ISSN No:-2456-2165

The analysis was only limited to systematic review. Thus, only a synthesis of the existing literature was done. A meta-analysis was not undertaken due to the diverse samples, measured outcomes, study designs, and instruments of data collection. The Cochrane Handbook warns against combining studies that do not use similar designs as this will cause real differences to be obscured (Higgins, & Green, 2008). Thus, meta-analyses has the potential to mislead, particularly if specific study designs, within-study biases, variation across studies, and reporting biases are not carefully considered (Deeks, Higgins, & Altman, 2019). Instead, a narrative summary of findings was adopted. Unlike clinical trials or cohort studies, cross-sectional surveys of attitudes and practices do not involve the study of an intervention, as such, there is limited awareness of any existing instrument that specifically addresses risk of bias in surveys of attitudes and practices (Agarwald, Guyatt, & Busse, 2019). Studies were therefore screened to identify its suitability to be included in the review. The researcher also assessed the quality of the studies by searching for these studies on google scholar and journals in which these studies were reported.

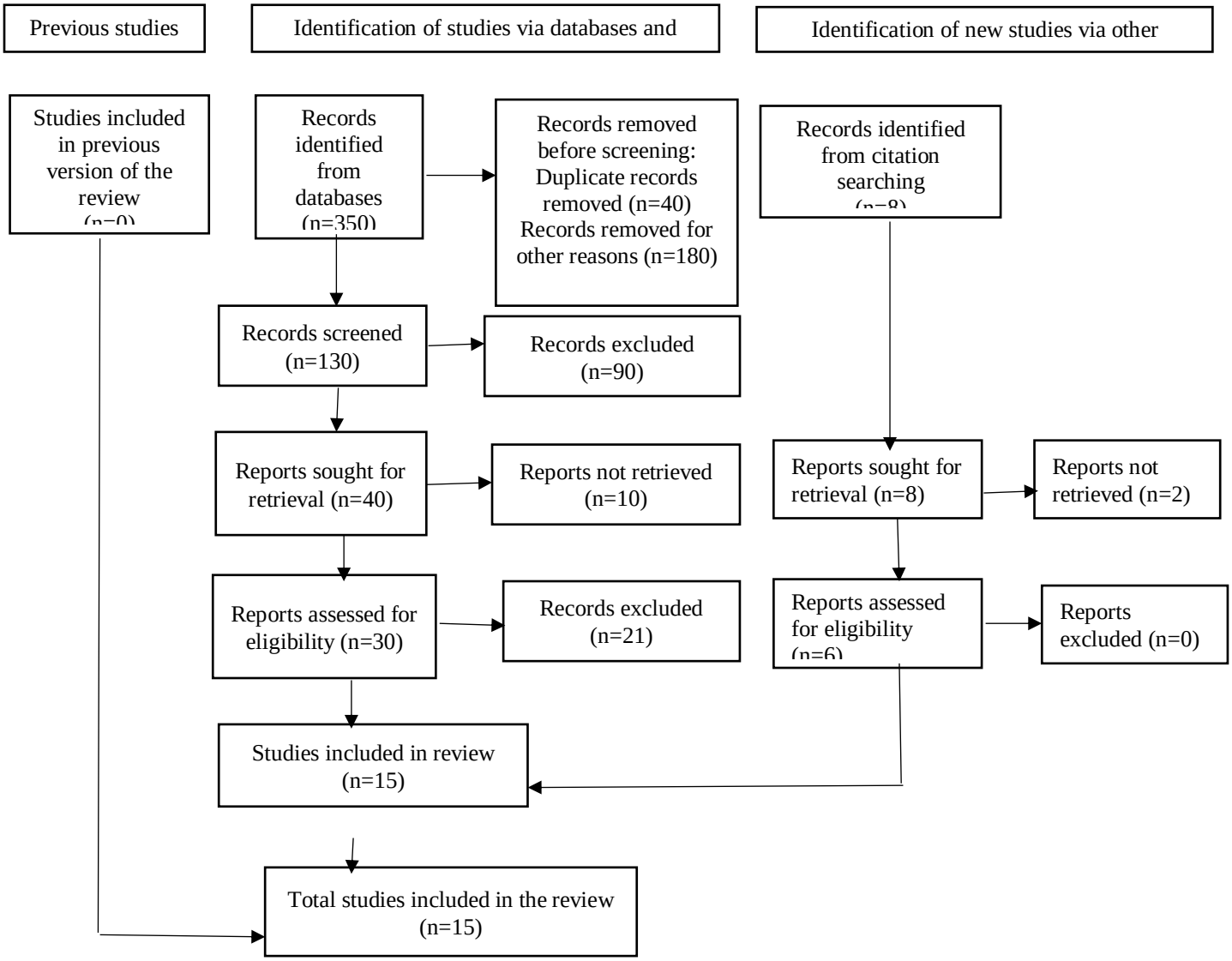

Fig. 1: Search and selection flow diagram of studies for the systematic review

III. FINDINGS

A total of 15 articles were reviewed. These studies were conducted in Africa from 2013 to 2023, focused on educational assessments and met the inclusion criteria outlined. The findings presented are guided by three research questions.

A. Research question one: What types of assessments have been studied for DIF in Africa?

Several types of assessments have been studied in the context of Differential Item Functioning (DIF) in Africa. These assessments encompass national level examinations such as the West African Examination Council (WAEC) Economics Multiple Choice Items (Ikeh, Ugwu, Mfon, Omosowon, Iketaku, Opa, Eze, Kalu, Ikwueze, & Ani, 2020), Mathematics multiple-choice items of the Basic Education Certificate Examination (BECE) (Evans, Mary, and Ekim, 2022), Mathematics for the Senior Secondary School Certificate Examination (SSCE) (Obiebi-Uyoyou, 2023), and Physics multiple-choice questions used by WAEC (Okagbare, Ossai, and Osadebe, 2023). Additionally, the Botswana junior certificate examination (BJCE) Mathematics paper one was investigated by Mokobi and Adedoyin (2014), while Annan-Brew (2020) focused on subjects including English Language, Mathematics, Integrated Science, and Social Studies in the West African Senior School Certificate Examination (WASSCE). State or regional-specific assessments were conducted, such as the analysis of Mathematics Joint Mock Multiple-Choice Test Items in Kwara State by Jimoh et al. (2020). Subject-specific assessments included performance-based assessments of mathematics among Senior High School students conducted by Gyamfi (2023) and a Mathematics Achievement Test examined by Effiom (2021). Language examinations explored reading achievement between English and isiXhosa languages in the prePIRLS 2011 assessment, as studied by Mtsatse and Van Staden (2021), and the Flowers on the Roof text of the PIRLS Literacy 2016 assessment across three languages, analyzed by Roux, van Staden, and Pretorius (2022). These studies, conducted by various researchers, have contributed to a better understanding of assessment practices in Africa, however, most of these studies have focused on national level assessments such as WASSCE, BECE, SSCE, and BJCE examinations.

Table 1: Types of assessments have been studied for DIF in Africa

Country	Authors	Assessment focus
Botswana	Mokobi and Adedoyin, (2014)	2010 Botswana junior certificate examination Mathematics paper one
Ghana	Annan-Brew (2020)	West African Senior School Certificate Examination (WASSCE) English Language, Mathematics, Integrated Science, and Social Studies from 2012 to 2016 Southern Ghana
	Gyamfi, (2023)	Performance-based assessment of mathematics among Senior High School (SHS) students in the Western Region of Ghana
Nigeria	Igomu (2014)	National Examinations Council (NECO) Biology questions for the year 2012
	Omoro, and Iro-Aghhedo, (2016).	National Business and Technical Examinations Board (NABTEB) 2015 Mathematics Multiple Choice Examination
	Sa'ad, et al., (2020)	2014 NECO English Language Examination in the North Senatorial District of Kano State, Nigeria
	Effiom (2021)	Mathematics Achievement Test
	Jimoh, et al., (2020)	Mathematics Joint Mock Multiple-Choice Test Items in Kwara State, Nigeria
	Ikeh, et al. (2020)	West African Examination Council Economics Multiple Choice Items in Nigeria
	Nathan, and Umoinyang, (2022)	Mathematics for the Basic Education Certificate Examination in Akwa Ibom State, Nigeria
	Evans, et al., (2022)	Mathematics multiple-choice items of the 2020 Basic Education Certificate Examination (BECE)
	Obiebi-Uyoyou (2023)	Mathematics for the Senior Secondary School Certificate Examination" in Nigeria
	Okagbare, et al., (2023)	Physics multiple choice used by WAEC in 2020
South Africa	Mtsatse and Van Staden (2021)	Reading achievement of prePIRLS 2011 between English and isiXhosa languages in South Africa
	Roux, et al., (2022)	Flowers on the Roof text of the PIRLS Literacy 2016 assessment across three languages in South Africa

B. Research question Two: What methods have been used to detect and analyze DIF in Africa?

Several studies have employed different approaches to analyze Differential Item Functioning (DIF) in their assessments. For instance, studies conducted by Roux et al. (2022), Mtsatse and Van Staden (2021), Evans et al. (2022), Nathan and Umoinyang (2022), Effiom (2021), Omoro and Iro-Aghhedo (2016), Gyamfi (2023), Annan-Brew (2020), and Mokobi and Adedoyin (2014) utilized item-response-theory-based approaches. These studies applied methods such as Rasch analysis, and Graded Response Model (GRM). Likelihood Ratio (LR) procedures to examine DIF was also used by Annan-Brew (2020). Mantel-Haenszel (MH) was used by Nathan and Umoinyang (2022), Jimoh et al. (2020), and Annan-Brew (2020) for their DIF analyses. Chi-square tests were employed by Okagbare et al., (2023) and Obiebi-Uyoyou (2023). Furthermore, logistic regression was utilized by Okagbare et al., (2023), Obiebi-Uyoyou (2023), Nathan and Umoinyang (2022), Ikeh et al., (2020), Sa'ad et al., (2020), and Igomu (2014) to investigate DIF in their respective assessments. Results indicated that themost commonly used DIF method was the IRT method.

Table 2: Methods have been used to detect and analyze DIF in Africa

Authors	Diff Detection Method
Mokobi and Adedoyin, (2014)	3PL (Multilog software) Item Response Theory (IRT)
Annan-Brew (2020)	Item Response Theory (IRT), Mantel-Haenszel (MH), and Likelihood Ratio (LR) procedures.
Gyamfi, (2023)	Item response theory (IRT, Graded Response Model, GRM)
Igomu (2014)	Logistic regression
Omoro, and Iro-Aghhedo, (2016).	Item response theory (IRT, Area Index (Raju) method)
Sa'ad, et al., (2020)	Logistic Regression (LR)
Effiom (2021)	Three-Parameter Logistic Model of Item Response Theory
Jimoh, et al., (2020)	Mantel-Haenszel (MH D-DIF).
Ikeh, et al. (2020)	Logistic Regression
Nathan, and Umoinyang, (2022)	One-parameter item characteristics model, logistic regression, and the modified Mantel-Haenszel Delta statistics
Evans, et al., (2022)	Item response theory (IRT)
Obiebi-Uyoyou (2023)	Logistic regression and chi-square test
Okagbare, et al., (2023)	Binary logistic regression and independent chi-square tests
Mtsatse and Van Staden (2021)	Rasch's item response theory (IRT)
Roux, et al., (2022)	Rasch's item response theory (IRT)

ISSN No:-2456-2165

C. Research question three: Which demographic groups have been studied and what are the findings of DIF for these groups?

Mokobi and Adedoyin, (2014) found that DIF in the 2010 Botswana junior certificate examination Mathematics paper one were found to be particularly challenging for both male and female students in rural schools. Annan-Brew (2020) investigated the WASSCE core subjects in Ghana and found gender DIF favoring females in English Language, while Mathematics and Integrated Science favored males. Gyamfi (2023) conducted a performance-based assessment in mathematics among SHS students in Ghana and found no presence of DIF. Igomu (2014) analyzed NECO Biology questions in Nigeria and discovered DIF favoring private schools and variations between urban and rural schools.

Other studies explored various subjects and assessments. Omoro and Iro-Aghhedo (2016) examined the NABTEB Mathematics Multiple Choice Examination in Nigeria and found DIF favoring females. Sa'ad, Ali, and Abdullah (2020) investigated the NECO English Language Examination in Nigeria, revealing DIF favoring males, boarding students, and urban school students. Effiom (2021) focused on a Mathematics Achievement Test, which showed bias for both male and female students. Jimoh et al. (2020) analyzed Mathematics Joint Mock Multiple-Choice Test Items in Nigeria and found DIF favoring males and rural schools. Ikeh et al. (2020) explored West African Examination Council Economics Multiple Choice Items in Nigeria, indicating DIF favoring urban school students. Nathan and Umoinyang (2022) studied Mathematics for the Basic Education Certificate Examination in Nigeria, finding DIF favoring females (based on one parameter) and males (based on Mantel-Haenszel method and logistic regression).

Additional, Evans, Mary, and Ekim (2022) investigated Mathematics multiple-choice items in the Basic Education Certificate Examination in Nigeria and found DIF favoring males, urban schools, and public schools. Obiebi-Uyoyou (2023) explored Mathematics for the Senior Secondary School Certificate Examination in Nigeria, revealing DIF favoring both males and females, as well as low socio-economic status test takers, rural test takers, and mixed schools. Okagbare et al. (2023) analyzed Physics multiple-choice questions used by WAEC in Nigeria, indicating DIF favoring males, urban schools, and private schools. Mtsatse and Van Staden (2021) focused on the reading achievement of prePIRLS 2011 in South Africa and found DIF favouring English test takers. Roux et al. (2022) investigated the PIRLS Literacy 2016 assessment across languages in South Africa, revealing no clear universal pattern of discrimination based on DIF. These studies contribute valuable insights into DIF and highlight the need for further analysis to ensure fair and equitable educational assessments in Africa.

Table 3: Findings of DIF for these groups

Authors	**Interest**	**Sample (N)**	**Results**
Mokobi and Adedoyin, (2014)	School location (rural and urban) with respect to sex (male and female)	4000	Favoured urban test takers
Annan-Brew (2020)	Sex (male and female) and Location (region).	36,035	Gender DIF favouring females in English Language, while Mathematics and Integrated Science, favoured males. No statistically significant gender DIF was found in Social Studies. Location DIF was found in favour of candidates from the Central and Greater Accra regions in all subjects.
Gyamfi, (2023)	Sex (male and female) and category of school (A, B and C)	750	No presence of DIF in the items for both gender and the category of school
Igomu (2014)	School location (urban and rural), and school ownership (public or private).	432	Diff favoured private schools with some items favouring urban schools and others favouring rural schools
Omoro, and Iro-Aghhedo, (2016).	Sex (male and female)	17,815	Diff favoured females
Sa'ad, et al., (2020)	Sex (male and female), accommodation status (day or boarding) and school location (urban and rural)	370	Diff favoured males, favoured boarding students, and favoured urban school students.
Effiom (2021)	Sex (male and female)	1,751	Items showed bias for both male and female students
Jimoh, et al., (2020)	Sex (male and female), and school location (urban rural)	1,062	Diff favoured males and rural schools
Ikeh, et al. (2020)	School location (urban and rural)	444	Diff favoured urban school students
Nathan, and Umoinyang, (2022)	Sex (male and female)	870	Diff favoured females (1 parameter), favoured males (Mantel-Haenszel method and logistic regression)
Evans, et al., (2022)	Sex (male and female), school location (urban and rural), and school ownership (public or private).	1,956	Diff favoured males, urban schools and public schools
Obiebi-Uyoyou (2023)	Sex (male and female), socio-economic status, location, school type (mixed vs single) and school ownership (public or private).	375	Diff favoured both males and females, 12 items each, low SES test takers, rural test takers and mixed schools.

ISSN No:-2456-2165

Okagbare, et al., (2023)	Sex (male and female), school location (urban and rural), and school ownership (public or private).	1,080	Diff favoured males, urban schools and private schools.
Mtsatse and Van Staden (2021)	Language use (English and isiXhosa)	819	Diff favoured English test takers.
Roux, et al., (2022)	Language use (isiZulu, Afrikaans and English students)	12,810	There was no clear universal pattern of discrimination against any one language based on the DIF

D. Research question four: What are the recommended solutions to minimizing diff in Africa?

Based on the findings of the studies, several recommendations were made:

In multilingual settings, it is crucial to ensure rigorous translation and quality assurance procedures when developing assessments to assess a diverse population accurately. Mtsatse and Van Staden (2021) emphasized the importance of considering differential item functioning (DIF) analysis to detect potential disparities between language groups and ensure fair treatment of all learners.

Test developers and measurement practitioners should conduct DIF analysis for all items after administering and scoring tests to validate the inferences made from the test results. Annan-Brew (2020) highlighted the importance of employing Item Response Theory (IRT), Mantel-Haenszel (MH), and Likelihood Ratio (LR) procedures to detect gender and location DIF in examinations, such as the West African Senior School Certificate Examination (WASSCE). Test developers should avoid including items with irrelevant clues or biases in assessments. Ikeh et al. (2020) recommended employing logistic regression analysis and DIF analysis to identify biased items, especially in high-stakes examinations like the West African Examination Council (WAEC) Economics Multiple Choice Items.

It is crucial to consider the presence of DIF and implement rigorous translation and quality assurance procedures when assessing a multilingual population. Roux et al. (2022) emphasized the importance of analyzing item-level performance among sub-groups to ensure fairness and validity in assessments. Test experts and developers should utilize DIF analysis, such as logistic regression, to detect biased items and ensure the development of fair and valid assessments. Effiom (2021) recommended conducting DIF analysis to identify gender-biased items in mathematics. Examination bodies, including the West African Examination Council (WAEC) and the National Examination Council (NECO), should incorporate DIF analysis in their pilot studies and quality assurance processes. Igomu (2014) and Sa'ad, Ali, and Abdullah (2020) highlighted the importance of identifying and addressing DIF to ensure fair treatment of different subgroups of examinees.

Test developers should conduct thorough field testing of all items before selecting them for examinations or any other educational decision-making instruments. Nathan and Umoinyang (2022) recommended considering the use of multiple methods, such as the one-parameter item characteristics model, logistic regression, and the modified Mantel-Haenszel Delta statistics, to identify biased items and ensure the validity of assessments. In high-stakes examinations, such as the Basic Education Certificate Examination (BECE), it is crucial to incorporate DIF analysis to identify and address potential biases related to gender, school location, and school ownership. Evans, Mary, and Ekim (2022) and Obiebi-Uyoyou (2023) stressed the importance of considering DIF analysis to ensure fairness and validity in assessments.

Test developers and measurement practitioners should explore the use of the differential item functioning (DIF) approach to detect bias in mathematics assessments, particularly in multilingual and diverse educational settings. Obiebi-Uyoyou (2023) highlighted the importance of considering DIF analysis to ensure fair assessment practices for different subgroups of test takers. The connotations reflected in test or examination items should be relevant to the life experiences of examinees responding to the items (Mokobi, & Adedoyin, 2014). By implementing these recommendations, educational assessments can be improved to ensure fairness, validity, and accurate measurement of student performance across different subgroups.

IV. DISCUSSION

Studies conducted by various researchers, have contributed to a better understanding of assessment practices in Africa; however, most of these studies have focused on national-level assessments (Mokobi and Adedoyin, 2014; Igomu, 2014; Omoro and Iro-Aghhedo, 2016; Sa'ad, Ali, and Abdullah, 2020; Annan-Brew, 2020; Ikeh et al., 2020; Evans, Mary, and Ekim, 2022; Obiebi-Uyoyou, 2023; Okagbare, Ossai, and Osadebe, 2023). These assessments include WASSCE, BECE, SSCE, and BJCE examinations. Expanding the scope of research to include other levels of assessment, such as classroom-level and regional-level assessments, is crucial for developing a comprehensive understanding of assessment practices in Africa. While national-level assessments provide valuable insights into the overall examination systems, they may not capture the intricacies and nuances present at lower levels.

Classroom-level assessments play a vital role in evaluating students' learning progress on a day-to-day basis. These assessments can provide valuable information about instructional practices, student engagement, and the effectiveness of teaching strategies. Exploring classroom-level assessments can help identify any biases or inequalities that may arise within specific classrooms or among teachers. Factors such as teaching styles, instructional materials, and classroom environments can impact students' performance and perceptions, which may not be fully captured by national-level assessments alone. By expanding the research scope to include classroom-level and regional-level assessments, a more holistic understanding of assessment practices in Africa

can be achieved. This broader perspective will enable researchers and policymakers to identify and address issues of bias, inequity, and disparities that may exist at various levels. It will also facilitate the development of assessment systems that are fair, inclusive, and responsive to the diverse needs of students across different contexts.

The item-response-theory-based approach stands out as the most commonly used method for analysing Differential Item Functioning (DIF) in the studies mentioned. This approach is highly regarded for its capacity to model the characteristics of individual test items and assess the differential performance of these items among various groups. By employing sophisticated statistical models, such as Rasch analysis and the Graded Response Model, researchers were able to effectively identify items that exhibited DIF and delve deeper into the underlying factors contributing to this differential functioning.

The item-response-theory-based approach offers several advantages in analysing DIF. It allows for a comprehensive understanding of the relationship between item characteristics, such as item difficulty and discrimination, and the performance of different groups of test takers. This enables researchers to pinpoint specific items that may disproportionately favor or hinder certain groups, providing valuable insights into potential sources of bias in assessments. Moreover, this approach takes into account the entire item response pattern, considering both correct and incorrect responses, which enhances the accuracy and reliability of DIF detection.

However, it is worth noting that the studies also employed other methods to complement the item-response-theory-based approach. The Likelihood Ratio procedure, for instance, was utilized by Annan-Brew (2020) to examine DIF. This statistical technique compares the likelihood of the data under different models, allowing researchers to assess the significance of DIF and identify items that contribute significantly to the observed differential performance.

The Mantel-Haenszel method was employed by Nathan and Umoinyang (2022), Jimoh et al. (2020), and Annan-Brew (2020) in their DIF analyses. This method involves stratifying the data by a relevant background variable and evaluating the conditional association between item responses and group membership. By controlling for potential confounding factors, researchers can ascertain whether the observed DIF is robust across different strata.

Chi-square tests were utilized by Okagbare et al. (2023) and Obiebi-Uyoyou (2023) as an alternative method for examining DIF. This statistical test assesses the association between item responses and group membership by comparing the observed frequencies to the expected frequencies under the assumption of no DIF. It offers a straightforward and easily interpretable approach to detecting DIF, particularly when the focus is on categorical item responses.

Logistic regression, a widely used statistical technique, was employed by multiple studies including Okagbare et al. (2023), Obiebi-Uyoyou (2023), Nathan and Umoinyang (2022), Ikeh et al. (2020), Sa'ad et al. (2020), and Igomu (2014) to investigate DIF. By modelling the probability of a correct response as a function of group membership and other relevant covariates, logistic regression allows for a more nuanced exploration of the factors influencing item performance across different groups.

The utilization of these diverse methods highlights the flexibility and adaptability in analysing DIF based on the specific characteristics of the data and the research objectives. While the item-response-theory-based approach offers a comprehensive framework, the inclusion of alternative methods provides a robust and comprehensive analysis of DIF, enhancing the reliability and validity of the findings.

The presence of Differential Item Functioning (DIF) in assessments in Africa has profound implications for educational practices and policies. It raises concerns about fairness and equity, as DIF can introduce biases favouring or disadvantaging certain student groups based on their characteristics. Addressing DIF is crucial to ensure equal opportunities for all students to demonstrate their knowledge and skills. Moreover, DIF undermines the validity of assessment results, affecting the accuracy of measurements and interpretations. By understanding and mitigating DIF, assessment developers can enhance the reliability of scores. DIF findings also shed light on curriculum and instructional practices, informing targeted interventions and reducing performance disparities. Policymakers can utilize DIF studies to design inclusive assessment frameworks and allocate resources equitably. Furthermore, addressing DIF contributes to quality assurance in assessment systems, fostering trust and confidence among stakeholders. Lastly, DIF research generates valuable knowledge, advancing educational research and guiding future efforts to improve assessment methodologies and practices. Overall, addressing DIF in assessments is essential for promoting fairness, equity, and quality in education in Africa.

V. CONCLUSIONS

In conclusion, the studies on Differential Item Functioning (DIF) in educational assessments across Africa shed light on the presence of bias towards certain groups. The findings highlight the complex nature of DIF, as different subjects, assessments, and contexts exhibit varying patterns of bias. While some assessments showed no clear bias, others revealed DIF favouring specific groups such as males, females, urban schools, rural schools, private schools, and students of different socio-economic backgrounds. These studies underscore the importance of ensuring fair and equitable educational assessments, as DIF can have significant implications for students' opportunities and outcomes. Further research and analysis are needed to better understand the factors contributing to DIF and to develop appropriate strategies for mitigating its effects. By addressing DIF, education systems in Africa can strive towards providing inclusive and unbiased assessments that promote equal educational opportunities for all students.

ISSN No:-2456-2165

VI. LIMITATIONS

While this study employed rigorous and comprehensive methods for the review, there are several limitations that should be acknowledged. Firstly, the exclusion of grey literature sources may have introduced bias and potentially overlooked valuable studies or unpublished works. Additionally, the decision to focus only on articles related to education may have limited the number of studies addressing validity and reliability, thereby impacting the comprehensiveness of the review. The exclusion of non-English articles also restricts the generalizability of the findings, as relevant information reported in other languages might have been missed. Lastly, the inability to contact authors of included studies might have resulted in the omission of important details or additional insights regarding the psychometric properties of their research. These limitations highlight the need for future research to address these gaps and provide a more comprehensive understanding of differential item functioning in educational assessments in Africa.

- **Data Availability:** The data (primary) used to support the findings of this study are available from the corresponding authors upon reasonable request.

DECLARATION

- **Conflicts of Interest:** No conflict of interest exists in the study. We wish to state categorically that there are no known conflicts of interest associated with this publication, and there has been no any financial support for this work that could have influenced the results.
- **Ethics Approval and Consent to Participate:** Not applicable.
- **Funding:** No funding was received for this study.

REFERENCES

[1.] Agarwal, A., Guyatt, G., & Busse, J. (2019). Methods commentary: Risk of bias in cross-sectional surveys of attitudes and practices. McMaster University

[2.] Aituariagbon, K. E., & Osarumwense, H. J. (2022). Non-parametric method of detecting differential item functioning in Senior School Certificate Examination (SSCE) 2019 economics multiple choice items. Kashere Journal of Education, 3(1), 146-158.

[3.] Ajeigbe, T. O., & Afolabi, E. R. I. (2017). Assessing unidimensionality and differential item functioning in qualifying examination for Senior Secondary School Students, Osun State, Nigeria. World Journal of Education, 4(4), 30-37.

[4.] Akanwa, U. N., Ihechu, K. J. P., & Nkwocha, P. C. (2020). Effect of language manipulation on differential item functioning of agricultural science multiple choice test items. Journal of the Nigerian Academy Of Education, 14(1).

[5.] Annan-Brew, R. (2020). Differential item functioning of West African Senior School Certificate Examination in core subjects in Southern Ghana. University of Cape Coast

[6.] Bundsgaard, J. (2019). DIF as a pedagogical tool: analysis of item characteristics in ICILS to understand what students are struggling with. Large-scale Assessments in Education, 7(1), 9.

[7.] Cameron, I. M., Scott, N. W., Adler, M., & Reid, I. C. (2014). A comparison of three methods of assessing differential item functioning (DIF) in the Hospital Anxiety Depression Scale: ordinal logistic regression, Rasch analysis and the Mantel chi-square procedure. Quality of life research, 23, 2883-2888.

[8.] Deeks, J. J., Higgins, J. P., Altman, D. G., & Cochrane Statistical Methods Group. (2019). Analysing data and undertaking meta-analyses. Cochrane handbook for systematic reviews of interventions, 241-284.

[9.] Effiom, A. P. (2021). Test fairness and assessment of differential item functioning of mathematics achievement test for senior secondary students in Cross River state, Nigeria using item response theory. Global Journal of Educational Research, 20(1), 55-62.

[10.] Evans, G., Mary, U. K. O., & Ekim, R. (2022). Evaluation of differential item functioning of basic education certificate examination (BECE) mathematics items in Akwa Ibom State, Nigeria. The African Journal of Behavioural and Scale Development Research, 4(1).

[11.] Goliath-Yarde, L., & Roodt, G. (2011). Differential item functioning of the UWES-17 in South Africa. SA Journal of Industrial Psychology, 37(1), 01-11.

[12.] Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Research synthesis methods, 11(2), 181-217.

[13.] Gyamfi, A. (2023). Differential Item Functioning of Performance-Based Assessment in Mathematics for Senior High Schools. Jurnal Evaluasi dan Pembelajaran, 5(1), 20-34.

[14.] Hagquist, C. (2019). Explaining differential item functioning focusing on the crucial role of external information–an example from the measurement of adolescent mental health. BMC Medical Research Methodology, 19(1), 1-9.

[15.] Halcomb, E. J., & Andrew, S. (2005). Triangulation as a method for contemporary nursing research. Nurse researcher, 13(2).

[16.] Higgins, J. P., & Green, S. (Eds.). (2008). Cochrane handbook for systematic reviews of interventions.

[17.] Igomu, A. C. (2014). An assessment of item bias using differential item functioning technique in NECO biology conducted examinations in Taraba state Nigeria. American International Journal of Research in Humanities, Arts and Social Sciences, 95.

[18.] Ikeh, F. E., Ugwu, F.C, Mfon, T. E., Omosowon, V. O., Iketaku, R. I., Opa, F. A., Eze, B. A., Kalu, I. A., Ikwueze, C. C., and Ani, M. I (2020). Analysis of differential item functioning in economics multiple choice items administered by West African examination council using logistic regression procedure. Journal of Critical Reviews 8(1), 980-986

[19.] Jimoh, M. I., Daramola, D. S., Oladele, J. I., Adaramaja, L. S., & Ogunjimi, M. O. (2020). Differential Item Functioning of Mathematics Joint Mock Multiple-Choice Test Items in Kwara State, Nigeria.
[20.] McNamara, T., & Roever, C. (2006). Psychometric Approaches to Fairness: Bias and DIF. Language learning.
[21.] Mokobi, T., & Adedoyin, O. O. (2014). Identifying location biased items in the 2010 Botswana junior certificate examination Mathematics paper one using the item response characteristics curves. International Review of Social Sciences and Humanities, 7(2), 63-82.
[22.] Mtsatse, N., & Van Staden, S. (2021). Exploring differential item functioning on reading achievement: A case for isiXhosa. South African Journal of African Languages, 41(1), 1-11.
[23.] Mustapha, M. L. A., & Olaoye, J. (2021). Determining Career Development Needs of College of Education Students in Kwara State, Nigeria.
[24.] Nathan, N. A., & Umoinyang, E. I. (2020). Differential item functioning in Basic Education Certificate Examination in Mathematics in Akwa Ibom State, Nigeria. Journal of Research in Education and Society, 13(3)
[25.] Nathan, N. A., & Umoinyang, E. I. Differential Item Functioning In Basic Education Certificate Examination in Mathematics in Akwa Ibom State, Nigeria.
[26.] Obiebi-uyoyou, o. (2023). Assessment of differential item functioning in mathematics multiple choice test questions in senior secondary school certificate examination in delta central senatorial district. Assessment, 3(2).
[27.] Okagbare, F., Ossai, P. A. U., & Osadebe, P. U. (2023). Assessment of differential item functioning of physics multiple choice used by WAEC in 2020 in relation to sex, location and school ownership in Delta State. IJARIIE 9(2)
[28.] Omoro, K. O., & Iro-Aghhedo, E. P. (2016). Determination of differential item functioning by gender in the National Business and Technical Examination Board (NADTEB) 2015 Mathematics Multiple choice Examination. International Journal of Education Learning and Development, 4(10), 25-35.
[29.] Osadebe, P. U., & Agbure, B. (2020). Assessment of differential item functioning in social studies multiple choice questions in basic education certificate examination. European Journal of Education Studies.
[30.] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. International journal of surgery, 88, 105906.
[31.] Palmatier, R. W., Houston, M. B., & Hulland, J. (2018). Review articles: purpose, process, and structure. Journal of the Academy of Marketing Science, 46, 1-5.
[32.] Roux, K., van Staden, S., & Pretorius, E. J. (2022). Investigating the differential item functioning of a PIRLS Literacy 2016 text across three languages. Journal of Education (University of KwaZulu-Natal), (87), 135-155.
[33.] Sa'ad N., Ali A. A. & Abdullahi, A. (2020). Differential item functioning of 2014 NECO English Language Examination in North Senatorial District of Kano State, Nigeria. Kano Journal of Educational Psychology 2(2)
[34.] Strobl, C., Kopf, J., & Zeileis, A. (2015). Rasch trees: A new method for detecting differential item functioning in the Rasch model. Psychometrika, 80, 289-316.
[35.] Ugwuozor, F. O. (2020). Constructivism as pedagogical framework and poetry learning outcomes among Nigerian students: An experimental study. Cogent Education, 7(1), 1818410.