

Assessment of Differential Item Functioning by School Ownership in 2020 WAEC SSCE Economics Multiple Choice Items

Jackson Joseph NKEKI

Department of Educational Evaluation and Counselling Psychology

Faculty of Education, University of Benin, Benin City

jacksonnkeki2@gmail.com

Theresa Ikhiaonotse ASEKOMHE

Department of Guidance and Counselling

Ambrose Alli University, Ekpoma

Tessyonotse2007@yahoo.com

Abstract

The study aimed at assessing differential item functioning by school ownership in 2020 WAEC SSCE Economics multiple choice items. To achieve the objectives of this study, four research questions were raised and two hypotheses were tested. The Survey Research Design was adopted. Multi-stage Sampling Procedure was used to select the sample for the study. The population of the study comprised 41,830 students in SS3 in Benin metropolis who wrote the 2020 WAEC SSCE. The sample size for the study was four thousand one hundred and eighty-three (4,183) SS3 students. The research instruments used for the study was 2020 WAEC SSCE Economics multiple choice questions. Logistic Regression Analysis was used to detect DIF while the hypotheses were tested using Chi-Square Statistical Analysis. All hypotheses were tested at 0.05 level of significance. The findings showed that out of the fifty (50) items in WAEC 2020 Economics examination question paper, eight items function differentially by school ownership. From the results of the hypotheses of the 2020 WAEC SSCE Economics, the null hypotheses of no significant difference in the number of items functioning differentially by ownership was retained. Based on the findings, it was recommended that test experts and developers should explore the use of Logistic Regression Analysis to detect items with DIF.

Key Words: Differential Item Functioning, Multiple Choice Item, Logistic Regression Analysis, West African Examinations Council, Economics

Introduction

Assessment is the process of gathering and interpreting evidence to make judgments about student learning. In order for assessment to facilitate learning, students need to receive information about their performance and the existing discrepancy between the actual and the desired state, and effectively process that information. Assessment is the only method through which classroom teachers can measure the knowledge of their students after they have been exposed to course of instruction (Nkeki & Osarumwense, 2021). It serves as means of controlling the quality of education, its product and progress. The term assessment, in its broadest meaning, connotes a

method of collecting and interpreting data about learning and students' performance that are used to provide information to students and their parents about the progress in acquiring knowledge and skills.

Assessment can be both formative and summative process. Formative assessment is an assessment for learning. It is used at the beginning of an instructional period and during the process of instruction as teachers check for student understanding. Dylan (2011) stated that formative assessment enables students to learn through the process of feedback and opportunities to practice and improve. Summative assessment on the other hand, is assessment of learning. It is used towards and at the end of the instruction period. Therefore, towards the end of secondary school education, students register for external examinations, which are conducted by different examination bodies of which West African Examinations Council (WAEC) is one. The West African Examinations Council (WAEC) is an examination body in West Africa countries of which Nigeria is one that conducts the Senior Secondary Certificate Examination and certifies the testees based on their abilities. One of the functions of WAEC is to certify the testees. This certification can be given after conducting examination for the testees to test their ability level.

Testing in the society today has become one of the most important parameters in detecting the standard of the educational system. Testing has consistently been an essential part of the educational system that all those who are not always present in school normally turn up to school during an examination period. The reason for testing is to show the latent ability of testee which is based on the testee's likelihood of success on a latent trait. This means that testing helps the tester to know whether testees have learned what they were expected to learn or the degree to which testees have learned the materials and to measure learning progress and achievement and to assess the effectiveness of educational programs. At the school level, educators create tests to measure their students' understanding of specific content or the effective application of critical thinking skills. Such tests are used to evaluate student learning and academic achievements at the end of an instructional period, such as the end of a project, unit, course, semester, program or school year. According to the Glossary for Educational Reform (2015) tests are used to determine whether students have learned what they were expected to learn or the degrees to which students have learned the material.

Testing as a process is a programme of exercise that has been predetermined after a careful taught and planning. Testing as a programme of exercise consists of many stages which include: Pre-determination of test purposes such as selecting a suitable test; administration of the tests to the testees; assigning scores to the testees' response; analysing, interpretation of the scores and application of the result. Most modern societies have seen testing as the most acceptable method in decision making in schools, government organisations and even industries. Testing is used for teaching, placement, admission, promotion, recruitment, guidance, evaluation and research purpose among others (Emaikwu, 2011). Testing according to Odili (2010) is the act of using a test to obtain data about a given attribute.

Test therefore, refers to a set of questions or statements which have been structured to elicit and measure responses about an attribute. Malcolm (2003) views test as an examination structured to evaluate how much knowledge an individual has acquired in a particular field. Kanjee, (2007) states that an important step in the construction of assessment tests is to make sure that no individual or group answering the question or instrument is disadvantaged in anyway. Test may be used to measure learning progress and achievement and to evaluate the effectiveness of educational programs and also measure student progress toward stated goals or to determine student placement in programs. They are recorded as scores or grades in a student's academic record as report card or for admission to higher education.

Another importance of tests at the school level is to determine student strengths and weaknesses. One effective example of this is when teachers' uses pre-tests at the beginning of units to find out what students already know and figure out where to focus the lesson. There is an assortment of literacy tests that can help target a weakness in accuracy as well as learning style and multiple intelligences tests to help teachers learn how to meet the needs of their students through instructional techniques. Tests have various advantages such as proper assessment that is tests provide a basis for finding out the suitability of candidates for various jobs, to evaluate the extent to which the objectives of education are being achieved, help to classify school objectives, decide for proper classification of students and selection of better candidate and reduction of labour turnover.

Examination bodies such as West African Examinations Council (WAEC) assess the testees with the use of Essay and Objective test questions. An Essay Test is an assessment technique that

requires students to respond to a question by developing, organizing, and writing an original composition. The purpose of an Essay Test is to assess testees' abilities to construct a logical, cohesive and persuasive writing piece. Essay Tests provide a better indication of examinee's achievement in learning. The response to the test items provides a clue to the nature and quality of the examinee's thought process. That is, assessing how the examinees present their ideas (whether their manner of presentation is coherent, logical and systematic) and how they conclude. Essay Tests have various advantages such as test takers can elaborate and provide detailed answers to a question, test takers are not able to guess and select an answer, test takers can review individualized responses from each examinee, and it can be used for all types of subjects and take less time to create questions. The problem associated with essay type of test is that it is difficult to score objectively and accurately and also, it is difficult to assess objectively and partially because the answers are not fixed like the answers of objective items because of the variability in the tester judgment regarding the contents of the answers. These notwithstanding, Essay test have proved to be incredible in testing. However, the Objective test items prove to be more objective in scoring.

Therefore, the focus of this study is on objective tests. Objective test is one of the assessment instruments used in assessing students' academic achievement in a given instruction. In objective tests, testees are required to choose the best possible answer (s) out of the options from a list (Okoro, 2006). Objective test items are those that require a specific answer. Objective tests can be easily scored objectively and accurately. The scoring will also be objective because when the answers are fixed there will obviously be complete interpersonal agreement among the testees. Test scores obtained from the objective tests are used to evaluate the competence of the examinees. Bush (2001) states that objective tests can boost the test takers chances of guessing the right answer to a question by removing unlikely choices. The objective tests generally are much more objective, because they are uniformly scored irrespective of the number of scorers. Objective questions usually have only one potential correct answer. Objective test includes matching, true/false, and multiple choices among others. In objective test, the answers are either right or wrong and require no interpretation or judgment on the part of the scorer as it is the case with subjective tests like essays.

Objective test items no matter how well they are constructed permit and encourage guessing by the examinee and the probability of guessing cannot be fully eliminated. The effect of the

guessing is the increase of the actual score obtained on the testee. Guessing is most obvious when the length of the test is short and the two alternative objectives type is used or when difficult alternative responses are included in multiple choice items or matching items and the length of the test is short. External examination bodies such as the WAEC assess the ability of the testees using essay and objective tests in their subjects such as Mathematics, English Language, Chemistry, Biology, Economics, Government, Commerce, Agricultural Science, and Literature and among others.

Economics is one the subjects conducted WAEC. It is one of the senior secondary school subjects that require assessment to ascertain students' basic knowledge, skills and understanding of the concepts and the objective of economic problems in a given society. The objectives of studying Economics according to Asadu (2001) are: to help examinees to gain knowledge for the practical solution to economic problem of the society, developing the nation and also the world at large, to educate students with the fundamental principle of economics essential that is required for useful living and to increase students respect for the dignity of labour.

To ensure that economics achievement test is fair for all examinees, the instrument should be fair. Fairness of a test is one which ensures that all test takers with equal ability level should have the same opportunity to display the knowledge and skills that they have obtained and which are important to the test purpose. Roever (2005) opines that a fair test is one that allows every testee the same chance to exhibit the knowledge and skills which they have attained or gained and which are useful to the purpose of the test. A test is said to be fair or unbiased when two groups of equal ability level testees with respect to the construct measured by the test earn the same score on each item of the test. If there is difference between scores of examinees with equal ability level, it gives indication of items that are functioning differentially for distinct populations of test takers. If the test is not fair or yield different scores from subgroups, it is said to flag Differential Item Functioning (DIF).

Differential Item Functioning refers to the situation where testees from different groups on the same level of the latent trait have a different probability of giving a certain response to a particular item. Differential Item Functioning (DIF) occur whenever a group of testees with identical level of abilities, taught and measured on equal construct of interest display different chances of answering an item correctly in the test (Camilli, 2006; Osterlind & Everson, 2009).

Differential Item Functioning (DIF) would occur in Economics achievement test if the item response function (IRF) for an item is different for two groups of equal ability levels. A test item is considered to flag DIF when a dimension on the examination is deemed to be irrelevant to the construct that is being measured, placing one group of the examinees at a disadvantage in taking the test (Hambleton & Roger, 2005). Consequently, if there is a distinction in the performance of groups of testees with distinct ability level on a particular item then, there is no evidence that the item has DIF; rather it is considered item impact (Schumacher, 2005). For instance, if SS2 test takers put in for the same examinations like WAEC with that of SS3 test takers, the test takers in SS3 would perform significantly better than the test takers in SS2. This is because the test takers in SS3 have had the opportunity to learn the materials and when this happened, the item is not considering to flag DIF, it is considering to be item impact (Nkeki & Orheruata 2023). But, if a test item differs in parameters among distinct subgroups of equal ability level, this might be considered to have DIF.

Several demographic variables or attributes were determined for potential differential item functioning. These variables are school ownership, ethnicity, socio-economic background, age, school location, race, sex and religion. The researcher's focus variables for the study is school ownership. School ownership factor is the reason for the study because of the widespread presence of privately owned schools in Nigeria in recent time. Alutu and Eraikhuemen (1999) reported that there was appreciable difference in academic performance in favour of private schools in 1996 and 1998 for JSS3 test takers. At every level of educational system, privately owned schools are nearly surpassing public owned schools. In Nigeria educational system today, there are different levels of education such as nursery, primary, secondary and tertiary. Privately owned schools are in functioning in all these levels of education. According to the study carried out by Anigbo (2006) on effect of school ownership on pupils' educational performance in the primary school level mathematics in Enugu State, pupils from privately owned primary schools was said to be at an upper hand than those pupils in the public owned schools. The public owned schools are found to have highly qualified teachers than the privately owned schools but examinees of the privately owned schools are found to have a higher grade in external examinations (Nkeki & Orheruata, 2023). This is because the teachers in privately owned schools are strictly supervised by the school owners to make a good name which might not be found in the public owned schools.

There are several procedures of detecting Differential Item Functioning (DIF). Some of these procedures are Logistic regression procedure, Mantel Haenszel, Item characteristic curve, Lord's chi-square method, Transformed item difficulty method, Likelihood ratio test, Purification method, Simultaneous Item Bias (SIBTEST) and Raju's area measures. Most of these procedures provide similar but not identical fact about DIF. This study focuses on Logistic regression procedure. Logistic regression procedure used in the detection of DIF was first proposed by Swaminathan and Roger in 1990 and it is a well-known statistical procedure. Logistic Regression Procedure is one of the forms of regression analysis which is needed to predict the result of a categorical dependent variable, that is, a variable that takes on a restricted number of cases or categories centered on one or more predictor variables. It is used to obtain odds ratio in the presence of more than one explanatory variable. The result is the impact of each variable on the odds ratio of the observed event of interest. The chances explaining the likely outcome of a trial are modelled as a function of explanatory variables. Logistic regression measures the relationship that exists among categorical dependent variables and continuous independent variables by changing the dependent variable to likelihood scores.

This study is hinged on Classical test theory (CTT). Classical test theory (CTT) is a body of related psychometric theory that predicts outcomes of psychological testing such as the difficulty of items or the ability of test-takers. It is a theory of testing based on the idea that a person's observed or obtained score on a test is the sum of a true score (error-free score) and an error score. The aim of classical test theory is to understand and improve the reliability of psychological tests. Classical test theory may be regarded as roughly synonymous with true score theory. The term "classical" refers not only to the chronology of these models but also contrasts with the more recent psychometric theories, generally referred to collectively as item response theory, which sometimes bear the appellation "modern" as in "modern latent trait theory".

Classical test theory was codified by Novick (1966) and described in classic texts such as Lord and Novick (1968) and Allen and Yen (1979). Classical test theory (CTT) has been the foundation for measurement theory for over 80 years. The conceptual foundations, assumptions, and extensions of the basic premises of CTT have allowed for the development of some excellent psychometrically sound scales. Classical Test Theory (CTT) is the underlying theoretical framework that underpins conventional psychometric testing. The broad objective of CTT is to

ensure reliability, precision, and accuracy of psychometric test scores by minimizing error. CTT is best exemplified by the following formula:

$$\text{Observed score } (X) = \text{True Score } (T) + \text{Error } (E)$$

For example, if a test taker completes an Economics test, and scores 16 / 20, his “Observed score” is 16. However, no psychometric assessment is 100% reliable, as error always influences the result, meaning this candidate’s observed score will differ from their “True score”. This true score is the test taker’s true level of answering the test questions, which is unknowable from a CTT perspective. The magnitude of difference between the observed score and the true score is determined by the level of error associated with that assessment, with unreliable assessments showing greater levels of error. The goal of CTT based assessments therefore, is to minimize the error component, ensuring maximum congruence between the observed score and the true score.

Under CTT, error is estimated using reliability coefficients, particularly test-retest reliability and internal consistency. The most commonly used estimate of internal consistency is the famous “Cronbach’s Alpha” statistic, which ranges from 0-1, with scores of .7 or above generally indicating a sufficient level of reliability. Higher levels of reliability generally indicate lower levels of error, and thus greater congruence between the true score and the observed score. Low levels of reliability however, show greater levels of error, meaning the observed score is likely to differ significantly from the true score, making the results invalid. Increasingly, CTT is being replaced by the more complex Item Response Theory (IRT), or modern psychometric test theory. Although CTT works well when assessments utilize a uniform set of questions, CTT is very limited when creating item-banked assessments. Because CTT posits that only two factors influence a person’s observed score i.e. their true score and error, CTT cannot account for differences in question difficulty, item discrimination, and guessing, all of which require parameterization in item banked assessments. To account for this additional complexity, IRT factors in these parameters into the observed score, freeing assessments from requiring fixed-forms. Classical test theory was laid down only after the following three ideas were conceptualized: recognition of the presence of errors in measurements; conception of that error as a random variable; and conception of correlation and how to index it.

Nigeria is a heterogeneous society made up of many sub groups (different ethnic groups, different religious groups, even different sexes, different location, different school owners and so on) where the various test takers who sit for standard examinations conducted by examination bodies like WAEC, NECO and NABTEB emerge. As such, the problem of differential item functioning (DIF) might likely arise in these examinations. The differences in performance could

be as a result of the nature of the items used in the test which could make members of one group perform better than members of the other group. The issue of test fairness and differential item functioning (DIF) have become increasingly vital in current researches and standard examinations.

One of the important considerations in the selection and use of any test is that test must not be biased, that is test must be fair to all testees. If the results of tests are to be used for decisions, then, the quality of test items has to be investigated so as to ensure that the test do not possess differential item functioning (DIF). That is, the items should be fair to all sub groups in terms of sex, race and many others. Roever (2005; as cited in Perrone 2006) points out that a fair test is one being valid for all groups, individuals and society providing each test taker with an equal opportunity of demonstrating his or her skills and knowledge relevant to the purpose of the test. In other words, test takers with similar knowledge of material on a test (based on their total scores) must perform similarly on individual examination items irrespective of their sex or race, otherwise it is said to have differential item functioning (DIF). In other words, test items are considered to be having differential item functioning (DIF) if they contain sources of difficulty which is irrelevant to the construct measured. Thus, items containing sources of difficulty beyond those of interest which results in a discrimination against particular groups are regarded as bias.

In Nigeria, the external examinations are the possible examinations that may have Differential Item Functioning because it is conducted across different subgroups. Researchers such as Amuche & Akpan, 2011; Omorogiuwa & Iro-Aghedo, 2016; Emaikwu, 2012; Enunwah, & Akwa, 2014; Ogbemor, 2012; and many others tend to investigate differential item functioning (DIF) in science-based subjects like physics, chemistry, mathematics and biology. Recommendations are that differential item functioning (DIF) been carried out in other subjects such as Economics being one of the central social science subjects. Hence the need to determine whether or not differential item functioning (DIF) exists in WAEC Economics. Therefore, this study seeks to empirically determine Differential Item Functioning (DIF) 2020 WAEC SSCE by school ownership.

Purpose of the Study

The purpose of this study was to determine the assessment of differential item functioning (DIF) in 2020 WAEC SSCE Economics Multiple Choice Items. Specifically, the study aimed to:

- i. determined items of 2020 WAEC SSCE Economics multiple choice items functioning differentially by School Ownership.

Research Questions

The following Research Questions were raised to guide the study:

- i. How many items of the 2020 WAEC SSCE Economics function differentially by school ownership?
- ii. Will there be a difference in the number of items functioning differentially in 2020 WAEC SSCE Economics test items by school ownership?

Hypothesis

Research Question 1 was hypothesized as follows:

- i. There is no significant difference in the number of items functioning differentially in 2020 WAEC SSCE Economics test items by school ownership

Methodology

The Survey Research Design was adopted in the study. The population comprised 41,830 candidates that enrolled and sat for the WAEC Senior Secondary School Certificate Examination in 2022 from 525 schools in Benin Metropolis. This number comprised 3,510 SS 3 students in all the 39 public owned secondary schools and 38,320 SS 3 students in all the 486 registered privately owned secondary schools in Benin metropolis made up of Egor, Ikpoba Okha and Oredo Local Government Area of Edo State. This population is the entire students who offered Economics in Benin Metropolis. The statistical population of the study is 50 Economics objective test items of WAEC SSCE. The sample size for this study comprised 4,183 SS3 students. Simple random sampling technique was used to select 2,093 male examinees from both public and privately owned secondary schools and 2,090 female examinees from public and privately owned schools representing 10% of the total population.

The instruments for this study are 2020 WAEC SSCE Economics objective questions. The Economics objective questions are made up of 50 items. These instruments are made up of four options A-D. The instrument used for this study was standardized tests conducted by West African Examinations Council (WAEC). So, they are presumed to have undergone the process of

validation. The reliability of the instruments was determined using test-retest method and the reliability co-efficient of 0.79 was obtained from 2020 WAEC Economics multiple choice test items. The instrument was administered to the students by the senior Economics teachers of the selected schools visited under the supervision of the researcher. The instrument was administered under similar conditions as given by West African Examinations Council (WAEC).

Logistic Regression Analysis Technique was used to analyze the data collected. An item is considered to flag Differential Item Functioning when the alpha value of 0.05 is greater than the significant value (Sig) on the other hand; an item does not flag Differential Item Functioning when the significant value is greater than the alpha value. According to Ling and Lau (2004) when the parameter *b* (item difficulty) for a particular group (for instance, Public School Examinees) is greater than the other group (for instance, Private School Examinees), it shows that such item is more difficult for the public school examinees group and then the item is considered to favour the private school examinees group vice versa, while the hypothesis was tested using the chi-square statistical analysis. The hypothesis was tested at 0.05 level of significance.

Results

Research Question One: How many items of the 2020 WAEC SSCE Economics function differentially by school ownership?

Table 1: Summary of Logistic Regression in detecting DIF by School Ownership of the 2020 WAEC SSCE Economics

| Items | B | S.E | Wald | Df | Sig | Exp(B) | Decision |
|-------|-------|------|-------|----|------|--------|----------|
| Q1 | .050 | .113 | 0.167 | 1 | .762 | 0.783 | NO DIF |
| Q2 | -.048 | .101 | 0.544 | 1 | .771 | 0.633 | NO DIF |
| Q3 | -.061 | .107 | 0.017 | 1 | .814 | 0.586 | NO DIF |
| Q4 | .108 | .105 | 0.009 | 1 | .967 | 0.872 | NO DIF |
| Q5 | -.324 | .102 | 1.337 | 1 | .468 | 0.780 | NO DIF |
| Q6 | .032 | .109 | 5.270 | 1 | .900 | 1.356 | NO DIF |
| Q7 | .146 | .114 | 1.652 | 1 | .199 | 1.158 | NO DIF |
| Q8 | .004 | .113 | 0.001 | 1 | .973 | 1.004 | NO DIF |
| Q9 | -.326 | .109 | 1.998 | 1 | .766 | 0.782 | NO DIF |
| Q10 | .140 | .111 | 1.579 | 1 | .209 | 1.150 | NO DIF |
| Q11 | -.037 | .100 | 0.129 | 1 | .794 | 0.764 | NO DIF |
| Q12 | -.026 | .105 | 0.164 | 1 | .717 | 1.023 | NO DIF |
| Q13 | -.057 | .103 | 9.517 | 1 | .010 | 0.962 | DIF |
| Q14 | .021 | .101 | 0.017 | 1 | .712 | 0.989 | NO DIF |

| Items | B | S.E | Wald | Df | Sig | Exp(B) | Decision |
|-------|-------|------|-------|----|------|--------|----------|
| Q15 | .001 | .101 | 0.000 | 1 | .988 | 1.001 | NO DIF |
| Q16 | -.171 | .110 | 4.593 | 1 | .039 | 1.632 | DIF |
| Q17 | .150 | .112 | 1.811 | 1 | .178 | 1.162 | NO DIF |
| Q18 | -.096 | .103 | 0.855 | 1 | .355 | 0.909 | NO DIF |
| Q19 | -.114 | .104 | 1.207 | 1 | .272 | 0.892 | NO DIF |
| Q20 | -.138 | .107 | 1.656 | 1 | .198 | 0.871 | NO DIF |
| Q21 | .328 | .109 | 0.582 | 1 | .727 | 1.796 | NO DIF |
| Q22 | .000 | .104 | 0.000 | 1 | .992 | 0.999 | NO DIF |
| Q23 | .373 | .108 | 9.905 | 1 | .009 | 0.878 | DIF |
| Q24 | .154 | .112 | 1.886 | 1 | .170 | 1.167 | NO DIF |
| Q25 | -.214 | .103 | 0.672 | 1 | .620 | 0.207 | NO DIF |
| Q26 | .053 | .105 | 0.254 | 1 | .614 | 1.054 | NO DIF |
| Q27 | -.343 | .109 | 1.936 | 1 | .102 | 0.710 | NO DIF |
| Q28 | .145 | .106 | 9.874 | 1 | .032 | 1.550 | DIF |
| Q29 | -.180 | .112 | 2.566 | 1 | .109 | 0.836 | NO DIF |
| Q30 | .031 | .099 | 0.101 | 1 | .751 | 1.032 | NO DIF |
| Q31 | .063 | .105 | 9.990 | 1 | .045 | 1.192 | DIF |
| Q32 | .087 | .106 | 0.667 | 1 | .414 | 1.091 | NO DIF |
| Q33 | -.050 | .108 | 0.211 | 1 | .646 | 0.952 | NO DIF |
| Q34 | .008 | .105 | 0.005 | 1 | .941 | 1.008 | NO DIF |
| Q35 | -.177 | .109 | 2.648 | 1 | .104 | 0.838 | NO DIF |
| Q36 | -.173 | .115 | 2.265 | 1 | .068 | 0.841 | NO DIF |
| Q37 | .174 | .111 | 2.438 | 1 | .118 | 1.190 | NO DIF |
| Q38 | .261 | .116 | 5.054 | 1 | .025 | 1.299 | DIF |
| Q39 | -.081 | .120 | 0.459 | 1 | .498 | 0.922 | NO DIF |
| Q40 | -.157 | .104 | 2.302 | 1 | .129 | 0.855 | NO DIF |
| Q41 | .020 | .106 | 0.037 | 1 | .848 | 1.021 | NO DIF |
| Q42 | .041 | .107 | 0.145 | 1 | .703 | 1.042 | NO DIF |
| Q43 | -.142 | .106 | 9.880 | 1 | .011 | 0.963 | DIF |
| Q44 | .070 | .111 | 0.400 | 1 | .527 | 1.073 | NO DIF |
| Q45 | .011 | .115 | 5.619 | 1 | .048 | 1.911 | DIF |
| Q46 | -.007 | .125 | 0.003 | 1 | .953 | 0.993 | NO DIF |
| Q47 | .094 | .121 | 0.603 | 1 | .437 | 1.098 | NO DIF |
| Q48 | -.295 | .115 | 1.741 | 1 | .371 | 0.765 | NO DIF |
| Q49 | .341 | .146 | 0.459 | 1 | .120 | 1.476 | NO DIF |
| Q50 | .124 | .135 | 1.729 | 1 | .708 | 0.378 | NO DIF |

Table 1 shows the summary of logistic regression in detecting the number of items that function differentially by school ownership in 2020 WEAC Economics multiple choice items. The finding

shows that eight items that is items 13, 16, 23, 28, 31, 38, 43 & 45 out of fifty WAEC Economics multiple choice items functioned differentially by school ownership.

Table 2: Chi-square (χ^2) Summary of Differential Item Functioning in favour of Public and Private school examinees

| School Ownership | N | df | Chi-square | P-value |
|------------------|---|----|------------|---------|
| Public | 6 | | | |
| Private | 2 | 1 | 14.69 | .079 |
| Total | 8 | | | |

$\alpha = 0.05$

Table 2 shows a chi-square value of 14.69 and a p-value of .079 and an alpha value of 0.05. The p-value of .079 is greater than the alpha value of 0.05; the null hypothesis of no significant difference in the number of items functioning differentially by school ownership in the 2020 WAEC Economics multiple choice items is retained. The finding shows that eight items that is items 13, 16, 23, 28, 31, 38, 43 & 45 out of fifty WAEC Economics multiple choice items functioned differentially by school ownership. It was revealed that eight items representing 13.3% functioned differentially by school ownership with the alpha value of 0.05 greater than the P-value and fifty-two (42) items representing 86.7% do not function differentially with P-value greater than 0.05. The findings showed that out of the eight (8) items that functioned differentially by school ownership, six items that is items 13, 16, 28, 31, 38 & 45 representing 10% were in favour of public school examinees while two items that is items 23 & 43 representing 3.33% functioned against the private school examinees.

Discussion of Findings

The findings of this study revealed that out of 50 multiple choice questions by school ownership in 2020 Economics, eight items (13%) functioned differentially by school ownership with alpha value of 0.05 greater than the P-value and forty-two (87%) do not function differentially with alpha value less than 0.05. Out of the eight items that functioned differentially in Economics 2020 multiple choice items by school ownership, six items (75%) were in favour of the public school examinees while two items (15%) were in favour of the private school examinees. The finding agree with the findings of Pedrajita and Talisayan (2009) who found out that 22 items in Chemistry Achievement test in Public and Private schools, 11 items in favour of public school examinees with 18.3% while 11 items in favour of private school examinees with 18.3%. This finding disagrees with the finding of Alade, Aletan and Sokenu (2020) which stated that out of 50

items, 44 items in favour of private school examinees with 88% while 6 items in favour of public school examinees with 12%.

The finding from the hypothesis showed that there is no significant difference in the number of items functioning differentially by school ownership in favour of public school examinees and those in favour of private school examinees in the 2020 WAEC multiple choice Economics examinations. The result of the analysis also showed that the school ownership DIF items are as a result of the presence of certain sources of difficulty that are irrelevant to the construct being measured and placing one group of examinees at disadvantaged and the potential sources of DIF include expert reviews of items for content, cognitive complexity, cultural load or linguistic differences in multiple language versions of items as well as cognitive interviews. These efforts are intended to identify the presence of DIF, that is, whether testees' responses not only reflect ability on the construct of interest, but also signal underlying variance outside the measured construct (i.e., construct-irrelevant variance). These irrelevant factors affect testees' performance.

Conclusion

It was concluded based on the findings of this study that there is an existence of DIF in WAEC 2020 Economics multiple test items with respect to school ownership. However, there were no significant difference in the number of items that functioned differentially by school ownership.

Recommendations

Based on the findings of the study, the researcher made the following recommendations:

- i. Examination bodies should ensure that examinations administered to students are bias free by ensuring that there is no evidence of DIF.
- ii. Test developers should write Economics test items that would not favour one group against the other. They should be put into consideration the heterogeneous nature of Nigeria.
- iii. Test developers should subject items constructed to verification for the presence of DIF and ensure that items that flagged DIF are corrected.
- iv. NECO should analyze students' responses to test items for differential functioning before administering to the examinees.

References

- Alade, O.M, et al (2020). Assessing the Differential Item Functioning of WASSCE Mathematics Achievement tests, *AJB-SDR*, 2 (3).18-24
- Alutu, A. N. G. & Eraikhuemen, L. (1999). A Comparative Study of the Academic Performance of Some Selected Private and Public Junior Secondary School Students in Egor Local Government Area of Edo State, *African Journal of Educational Research*. 5 (2). 121 – 130.
- Amuche, C. & Akpan, F. (2011). An Assessment of Item bias using Differential Item Functioning Technique in NECO Biology. *American International Journal of Research in Humanities, Arts and Social Sciences*
- Anigbo, L. C. (2006). Development and Standardization of Mathematics Achievement Test Batteries for primary four pupils in Nigeria. Unpublished doctoral dissertation, University of Nigeria Nsukka
- Asadu, I. N. (2001). Trend in student’s enrolment and performance in senior secondary certificate examination in Economics. Unpublished doctoral dissertation, University of Nigeria, Nsukka.
- Bush, L.D. (2001). Culturally Competent Assessment: More than Non-biased Tests. *Journal of Child and Family*, 11, 61-78.
- Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.) (221–256). American Council on Education/Praeger Publishers.
- Dylan, W. (2011). *Embedded Formative Assessment*. Hawker Brownlow Education.
- Emaikwu, S. O. (2012). Evaluation of student’s ability in schools. Being a paper presented at a workshop on teaching practice on Friday, 29th July in the College of Agricultural & Science Education, Federal University of Agriculture Makurdi, Benue State.
- Enunwah, C.I & Akwa, A.M. (2014). Differential Items and Group Functions of secondary students Achievements in Mathematics in Cross Rivers State. *African Education Indices*, (7) 2276-982
- Glossary of Educational Reform for Journalists, Parents and Community Members (2015). *Passion in Education at the Heart of Teaching*
- Hambleton, R., & Rodgers, J. (2005). Item bias review. *Practical Assessment, Research, and Evaluation*, 4(6).
- Kanjee, A. (2007). Using logistic regression to detect bias when multiple groups are tested. *South African Journal of Psychology*, (37) 47-61.
- Ling, S.E & Lau, S.H. (2004). Detecting differential item functioning (DIF) in standardized multiple-choice test: An application of item response theory (IRT) using three parameter logistic model. *Journal of Applied Psychology*, 94 (7), 452-459.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Malcolm, T. (2003). An achievement test. Retrieved November, 20, 2013, from <http://www.wisegeek.com/what-is-an-achievement-test.htm>
- Nkeki, J.J. & Osarumwense, H.J. (2021): Influence of Feedback on Ikpoba Okha Senior Secondary School Students Academic Achievement in Economics: *Journal of Educational Evaluation and Counselling Psychology* 6 (41).

- Nkeki, J.J & Orheruata, M.U (2023): Determination of Differential Item Functioning by sex in the 2017 and 2018 National Examination Council SSCE Economics Multiple Choice Items using Logistic Regression Analysis: *UniAfrica Journal of Education*, 2 (1).
- Odili, J.N (2010) Reducing Differential Item Functioning through Language of Test Items: An Avenue for Promoting Peace Through Psychological Testing, Readings in Conflict Management and Peace Building in Africa (1) 341-357
- Ogbebor U.C (2012). Differential Item Functioning Biology Question Paper of National Examinations Council in Taraba State Nigeria. Unpublished M.ed thesis. University of Ibadan.
- Okoro, O.M (2006). Measurement and evaluation in education. Uruowulu-Obosi: Pacific Publishers Ltd.
- Omoroguiwa, K.O. & Iro-Aghedo, E.P. (2016). Determination of Differential Item Functioning by gender in the National Business and Technical Examination Board 2015 Mathematics multiple choice Examinations; *International journal of Education, Learning & Development*, 25-35.
- Osterlind, S.J., & Everson, H.T. (2009) *Differential item functioning*. Thousand Oaks, CA: Sage Publishing.
- Pedrajita, J.Q., & Talisayon, J.L (2009). Using logistic regression to detect biased test items. *The international of educational and psychological assessment journal*
- Perrone, M. (2006). Differential item functioning and item bias: Critical considerations in test fairness. *Columbia University Working Papers in TESOL & Applied Linguistics*. 6 (2), 1-3.
- Roever C (2005). "That's not fair!" Fairness, bias, and differential item functioning in language testing. Retrieved from <http://www2.hawaii.edu/~roever/brownbag.pdf>
- Schumacker, R (2005). Test bias and differential item functioning *Journal of Educational Measurement*, (4) 97-118