International Journal of Advanced Educational Research
ISSN: 2455-6157; Impact Factor: RJIF 5.12
Received: 05-03-2019; Accepted: 07-04-2019
www.educationjournal.org
Volume 4; Issue 3; May 2019; Page No. 31-35

Identification of school type differential item functioning of 2010 senior school certificate (NECO) English language examination among students of Dawakin Kudu Education Zone-Kano state, Nigeria

Ibrahim Shu'Aibu[1], Binta Abba MRS[2]
[1] Department of Education, Sule Lamido University, Kafin-Hausa, Jigawa State, Nigeria
[2] Department of Education, Bayero University Kano, Nigeria

Abstract
The study investigated the incidence of Differential Item Functioning (DIF) on items in NECO/SSCE English language multiple-choice examination in 2010 due to school type differences among senior secondary school students in Dawakin Kudu Education Zone, Kano State, Nigeria. An expost-facto design was employed for the study. Population of the study consisted of 3,076 SSIII students in public secondary schools in Dawakin Kudu Education Zone in Kano state, Nigeria that sat for NECO/SSCE examination in 2010. Three hundred and forty-six (346) students were randomly selected as the sample. The analysis of DIF was done on person-by-item matrix on 100 English language multiple-choice items by NECO of 2010 with regards to school type. The DIF indices of each item was detected using Logistic Regression (LR) analysis and percentage of correct response to each item was obtained through descriptive statistics using cross tabulation. Items that favoured each group were detected and presented in accordance with the research questions. The results from the study shows that, 42% of the items were having negligible level of DIF; 20% were having moderate level of DIF while 38% of the items were having magnitude level of DIF. Out of the 38 items, 33 items were in favour of day school students while5itemwere in favour of boarding school students. From the findings, it was recommended that Examination bodies, experts in Education sector and those that are saddled with the responsibility of developing, validation and administration of test(s) need to carry out measurement bias on items at all times.

Keywords: item bias, differential item functioning (DIF)

Introduction

In Nigeria, achievement at any level of education is awarded with certification of those who successfully completed a course of study with good academic records. Educational test results is an important yardstick by which society pronounces the product of its educational system. As such, it is of paramount important that educational institutions and Examination bodies should conduct tests that will enable them establish the desired characteristics of their examinees. The National policy on Education, (FRN, 2014) [5] point out that a test should measure the knowledge and skills taught in schools and this should be to all students irrespective of gender, socio-economic background, location, school type and ethnic group etc. Similarly, for any examination to be judged good; the conduct of the examination and the published results should be deemed fair and achieve high level of acceptance by the public. More so, the examination should ensure that no particular candidate or candidates has/have an unfair advantage over others.

Thus, in order to draw valid conclusions on the results obtained from an achievement test, it is necessary that the test is a valid measurement of what it is intended to measure.

The national objective of creating an egalitarian society as stated in the National Policy on Education (FRN, 2014) [5] is one of the foundations upon which education should operate. And for these noble objectives to be achieved, test items at any level must discriminate only on the basis of the subject matter ability, not by any extraneous variables such as gender, educational programmers, socio-economic status, language, location etc.

In Nigeria, , at the end of secondary education, students are expected to sit for public examinations such as the West African Senior School Certificate Examinations (WASSCE), conducted by the West African Examinations Councils (WAEC), Senior School Certificate Examination (SSCE), conducted by the National Examination Councils (NECO), and the National Technical and Business Certificate Examination (NTCE/NBCE) also conducted by the National Business and Technical Examination Board (NABTEB) and Joint Admission Matriculation Board (JAMB).

English language examination is one of the tests constructed by NECO. There are three (3) kinds of items generated by NECO in English language for SSCE. These are: Paper one (1); Paper two (2) and Paper three (3). Paper two (2) of the examination is usually multiple-choice (MC) type with options the examinees are expected to choose from. Such items appear to be the most flexible and probably most effective of the objective item types. This is due to their versatility in measuring all levels of cognitive skills. They also permit sampling of wide contents and objectives. As a result of objectivity in scoring the MC items, scores tend to be reliable. In the same vein, the items can be scored quickly and accurately using machine. Despite these numerous advantages, MC items can provide unprepared students the opportunity to guess, and with guesses that are right, they get credit for things they do not know. This can

consequently expose students to misinformation that can influence subsequent thinking about the content and thus likely to make the test not to measure what it was designed to measure.

Thus, it is therefore necessary, after constructing and reviewing test items, based on some common item writing rules, to subject them to a series of judgmental and statistical evaluation techniques for validation purposes. This is done in order to identify poor quality items that might be in need of revision or outright rejection. The accepted or revised items are then trial tested. The responses from the trial test would be subjected to series of statistical analyses (Item analyses). The results of the analyses would be used to further identify and/or finalize the most appropriate test items to be incorporated in the final version of the test as suggested by Suen and McClellan (2003) [16].

The question one may ask is; are these test items in English language fair enough for all groups? This is because, Nigeria is composed of more than 250 ethnic groups situated at different locations (some are urban-based while some from rural areas). Thus, it is obvious; that most central examination being administered to Nigerian students may not be fair to both groups of examinees. Questions of test bias are related to questions of test validity. A test is valid if it measures what it purports to measure and invalid if it does not. Bias is a kind of invalidity that arises relative to groups. According to Gregory, (2006) [9], a test is biased if the characteristics of the measures, the measurement process or the interpretation of the results of measurement lead to inaccurate inferences about the knowledge, skills and other attributes of an individual or a group of testee.

One of the ways of investigating test bias at the item level is through differential item functioning (DIF) analysis. DIF is said to be present in a test item when, examinees from two sub- populations with same trait/ability level have different expected scores on the same item (Kamata & Vaughn, 2004) [11].

It has been claimed that some of the national examinations unfairly favor examinees of some particular groups e.g., cultural or linguistic groups to the extent that it is now believed that a particular section of the country perform most woefully in these national examinations as reported by numerous studies such as Garba (2017) [7]; Adedoyin (2014) [1], Amuche and Fan (2014); Nworgu and Odili (2005) [14]. A critical look at the perception of people on such national examination in Nigeria indicated that there is serious nature of item bias. A given educational or psychological test consisting of many items with significant DIF may be unfair to certain subgroups and it is important to identify such items for subsequent review as mentioned earlier.

It is against this background, that a study of this nature needs to be carried out, that is to analyze NECO 2010English language multiple choice test items with regards to school type.

Statement of the Problem

The multi-lingual and multi-cultural nature of Nigerian entity and the absence of a national unifying indigenous language led to the adoption of English language as a medium of intra-national and inter-national communication (Fakeye, 2006) [4]. More importantly English Language has become the spindle on which the educational wheel of Nigeria revolves. The language is the medium of instruction for almost all the school subjects from the primary school level to the University, in addition to being a compulsory school subject that must be passed at all levels of education in Nigeria, passing English Language at credit level upon completing secondary school education is a pre-requisite for getting admission into Nigerian Universities irrespective of whether the University is federal, state or privately owned.

The poor performance of students in English language at public examinations in recent times has been explained as a major cause of the decline in academic achievement and standard of education in Nigeria (Fakeye, 2009).

Underachievement in English Language has remained a recurrent feature of annual school certificate results released by National Examinations Council (NECO), resulting in high failure rate. The results of statistics carried out by the Nigerian Federal Ministry of Education from 2006 to 2010 public schools revealed the number of candidates and percentages of those with five credit passes and above including English language in NECO examinations. In Kano State, the percentages of those with five credit passes and above including English language in NECO examinations for public school results were obtained against the entry of five years (2006-2010) in which 1.155% in 2006; 2.378% in 2007; 4.492% in 2008; only 0.8204% in 2009 and 1.947% in 2010 respectively.

From the above result, it is clear that the percentage of student's performance of Kano State public schools in NECO 2010 is quite alarming in which only 1.947% of the candidates have five credit passes and above, these figures indicated the magnitude of the massive failure in NECO/SSCE examination including that of English language.

Given a regional spread of candidates covered by the National Examinations Council (NECO) Senior School Certificate Examination (SSCE), it is useful to analyze English multiple examination items for possible differentially item functioning. Thus, test with differential functioning items cannot be used to achieve the goals stipulated in the Nigerian National policy on Education concerning development of individuals and building a virile nation where every citizen has an equal opportunity. Therefore, for the results of examinations conducted by NECO to be used to achieve such objectives of educational system, there is a need to study the response patterns of examinees in English language examination by using method of DIF to establish whether the examinees functioned differently in English language multiple choice examination items set and administered by NECO, June/July, 2010, with respect to school type.

Objectives of the Study

1. To find out which items exhibits magnitude level of DIF with regards to school type in NECO/SSCE English language multiple-choice examinations of 2010
2. To find out which of the items in NECO/SSCE English language multiple-choice examinations of 2010 is/are in favour of each group.

Research Questions

1. Which items exhibits magnitude level of DIF with regards to school type in NECO/SSCE English language multiple-choice examinations of 2010?
2. Which of the items in NECO/SSCE English language multiple-choice examinations in 2010 is/are in favour of each group?

Literature Review

Concept of Differential Item Functioning (DIF)

DIF is a statistical method to explore whether test items function differentially across different groups of test takers who are matched on ability. DIF exists when different groups of learners have differing response probabilities of either (a) successfully answering an item (i.e., in multiple choice) or (b) receiving the same item score in an assessment (Ferne& Rupp, 2007; Zumbo, 2007) [6, 17].

DIF occurs when a test or a survey item functions differently for a reference group (e.g., males) of examinees or respondents compared to a focal group (e.g., females) of examinees or respondents, after controlling for the level of an attribute being measured (Dodeen & Johanson, 2003 [2]; Kamata & Vaughn, 2004) [11]. For example, an item exhibits DIF if the probability of males responding to a specific category differs from females when they both are operating at the same overall level on a given construct.

DIF is the displaying differences of the probability of answering item correctly according to subgroups, in every ability of psychological structure that is aimed to be measured with the item (Embretson & Reise, 2000) [3]. In DIF studies, the performances of different groups are compared according to test items related to demographical specifications such as men-women in the same ability level, Asian-European, and so on (Greer, 2004) [8].

Approaches for DIF Detection

There is no any single "best method" of DIF analysis which is effective and useful for all purposes as put forward by Lai, Teresi and Gershon, (2005) [12]. Various methods exist to examine DIF to estimate the level of disability, disease, capability, etc., among which some methods assume the existence of a latent variable which is estimated by using marginal maximum likelihood, some other methods assume a "valid" target dimension etc. However, according to McNamara and Roever (2006) [13] the following four broad categories of methods are used for DIF detection: (1) analysis based on item difficulty (comparing item difficulty estimates); (2) nonparametric approaches (procedures using contingency tables, chi-square, and odd ratios); (3) item-response-theory-based approaches (these approaches including 1, 2, and 3-parameter analyses which frequently compare the fit of statistical models); and (4) other approaches (including logistic regression, generalizability theory, and multifaceted measurement). In addition, Salehiand Tayebi, (2012) [15], classifies different DIF detection methods according to whether they (a) are parametric or non-parametric; (b) are based on latent or observed variables; (c) treat the disability dimension as continuous; (d) can model multiple traits; (e) can detect both uniform and nonuniform DIF; (f) can examine polytomous responses; (g) can include covariates in the model, and (h) must use a categorical studied (group variable).

DIF Detection with Logistic Regression (LR)

Salehi and Tayebi (2012) [15] reported the use of binary Logistic regression (LR) procedure as the most popular for detecting DIF in dichotomous test items. In addition, LR is a useful technique for detecting both kinds of DIF, uniform and non-uniform DIF in dichotomously scored items. LR, in a predictive context, uses regression of the external criterion on test score. The effect level can be determined by the standardized regression parameters. Jodoin and Gierl (2001) [10] classified the effect levels of DIF that are determined with LR in the following ways:

A Level: If R < .035, a negligible level of DIF is present.

B Level: If .036 < R < .070, a medium level of DIF is present.

C Level: If R > .071, a magnitude level of DIF is present.

Methodology

The data used for the study consisted of 2010 multiple-choice English language students' responses in Dawakin kudu Education Zone, Kano State, Nigeria. This examination consists of 100 items grouped into six sections. Section 1 contains 1-10 items, section 2: 11-20 items, section 3: 21-30 items, section 4: 31-40 items, section 5: 41-73 items, and section 6: 74-100 items respectively. An ex-post facto design was adopted which seeks to find out the factors that are associated with certain occurrences of already existing condition or state of affairs and searching back in time for plausible causal factors retrospectively (Cohen, Lawrence & Morrison, 2007; Cooper & Schindler, 2001). The population of the study consisted comprised 3,076 SS III students and a sample of 346 students were used out of which 249 are day students while 97 are boarding students. The item responses were scored in a binary format of "1" correct and "0" incorrect. The data for the study was collected by the researchers person-by-item matrix format. The data was analyzed using Binary Logistic Regression in determining the indices of DIF for each item. Descriptive statistics using frequency cross tabulation was used in identifying the percentage of correct response for each item that functioned differently by school type.

Results

Research Question One: Which items exhibits large magnitude level of DIF with regards to school type in NECO/SSCE English language multiple-choice examinations of 2010?

Table 1: Summary of Logistic Regression Analysis for GSP2202 with Regards to Gender

Effect Level	Item Numbers	Percentages
R < .035, a negligible level of DIF	1,4,5,6,9,11,12,15,16,22,23,24,26,27,31,32,34,36,44, 46,48,50,52,56,58,61,63,68,71,72,74,75,82,83,84,85, 86,87,88,89,90,95	38.00
.036 < R < .070, a medium level of DIF	2,13,19,29,30,33,41,51,55,57,59,60,66,67,70,73,93,94 96,98	20.00
R > .071, a magnitude level of DIF	3,7,8,10,14,17,18,20,21,25,28,35,37,38,39,40,42,43,4547,49,53,54,62,64, 65,69,76,77,78,79,80,81,91,92,97, 99,100	38.00
	Total	100.00

In the above table, it can be seen that 42.00% of the items have negligible level of DIF, 20.00% were found to have moderate DIF and 38 items (38.00%) were found to have a magnitude level of DIF.

Research Question Two: Which of the items in NECO/SSCE English language multiple-choice examinations in 2010 is/are in favour of each group?

Table 2: Summary of Cross Tabulation Analysis by Location indicating items that favour day and boarding school students in each section of English language multiple-choices examination set and administered by NECO/SSCE 2010

Content source	No of items in each source	No of items that favour day students	No of items that favour boarding students
SECTION 1	10	3	1
SECTION 2	10	4	--
SECTION 3	10	2	1
SECTION 4	10	3	2
SECTION 5	33	11	--
SECTION 6	27	10	1
Total	100	33	5

The above table depicts the number of items (38items) identified as showing school type bias by each content source. Among which item number:(3, 7, 10, 14, 17, 18, 20, 25, 28, 35, 39, 40, 42, 43, 45, 47, 49, 53, 54, 62, 64, 65, 69, 76, 77, 78, 79, 80, 81, 91, 92, 99 and100)are in favour of day school students while item:(8, 21, 37, 38 and 97) favoured boarding students.

Summary and Discussions

Items in English language multiple-choice items set by NECO/SSCE 2010 functioned differently for students by school type in which thirty-eight (38) out of the one hundred (100) items were flagged as having significant school type bias, among which thirty-three (33) were in favour of day school students and five (5) were in favour of boarding school students.

The findings from study shows that items in English Language multiple-choices examination set and administered by NECO/SSCE 2010, differentially functioned for day and boarding school students in which day school students showed a statistically and consistent advantage over boarding school students. This can be attributed to the fact that most of the boarding students may likely be preoccupied with other school activities such as cleaning; games etc. in addition to other challenges associated with the environment because most of the boarding schools are located in the rural areas. As such cannot have access to social infrastructural facilities such as radio, television, mobile phones that will aid in their learning of English language. Day students on the other hand can have those facilities at home or school since most of the schools are located in the urban areas.

Conclusion

It can be concluded that items in English Language multiple-choices examination by NECO/SSCE 2010 were flagged as having school type bias which significantly and consistently favoured day students than their boarding student counterparts.

Recommendations

For any test to be judged good; it should be subjected to the process of item analysis, it is therefore recommended that differential item functioning procedure for detecting item bias should be carried out on all items in various subjects of public examinations. Examination bodies, experts in Education sector and those that are saddled with the responsibility of developing, validation and administration of test(s) need to carry out measurement bias on at all times.

References

1. Adedoyin OO. Identifying Location Biased Items in the 2010 Botswana Junior Certificate Examination Mathematics Paper One Using the Item Response Characteristics Curves. Journal of International Review of Social Sciences and Humanities. 2014; 7(2):63-82.
2. Dodeen H, Johanson GA. An Analysis of Sex-Related Differential Item Functioning in Attitude Assessment. Assessment & Evaluation in Higher Education. 2003; 28(2):129-134.
3. Embretson SE, Reise SP. Item response theory for psychologists. Mahwah, NJ: Lawrence Erlbaum, 2000.
4. Fakeye DO. ICT-Assisted Instructure and Students Vocabulary Achievement in Selected Senior Secondary Schools in Ibadan. Journal of Humanities in Education. 2006; 1(1):14-21.
5. Federal Republic of Nigeria FRN. National Policy on Education (6th Ed.). Lagos: NERDC, 2014.
6. Ferne T, Rupp A. A Synthesis of 15 Years of Research on DIF in Language Testing: Methodological Advances, Challenges, and Recommendations. An International Journal Language Assessment Quarterly. 2007; 4:113-148.
7. Garba A. Nigerian educational system and the sustainable development goals: problems and prospects. Being a lead paper presented at the first international conference organized by faculty of education, north west university, Kano. 26-27th April, 2017.
8. Greer TG. Detection of Differential item Functioning (DIF) on the SATV: A Comparison of Four Methods: Mantel-Haensrel, Logistic regression, Simultaneous Item Bias and Likelihood Ratio Test (Unpublished doctoral dissertation). Retrieved, 2004. from www.readperiodicals.comon 28-06-2012.
9. Gregory RJ. Psychological testing: History, principles and applications. 4th edition. India: Pearson Education Inc, 2006.
10. Jodoin GM, Gierl MJ. Evaluating type I Error and Power Rates using an Effect Size Measure with Logistic Regression Procedure for DIF Detection. Applied measurement in Education. 2001; 14:329-349.
11. Kamata A, Vaughn B. An Introduction to Differential Item Functioning Analysis. Learning Disabilities: A Contemporary Journal. 2004; 2(2):49-69.
12. Lai JS, Teresi J, Gershon R. Procedures for the analysis of differential item functioning (DIF) for small sample sizes. Evaluation & the Health Professions. 2005; 28(3):283-294.

13. McNamara T, Roever C. Language testing: The social dimension. New York: Blackwell publishing, 2006.
14. Nworgu BG, Odili JN. Analysis of DIF in WAEC SSCE Biology Multiple Choice Test. Review of Education. 2005; 16(2):140-152.
15. Salehi M, Tayebi A. Differential Item Functioning: Implications for Test Validation. Journal of Language Teaching and Research. 2012; 3(1):84-92.
16. Suen HK, McClellan S. Item construction principles and techniques. Retrieved, 2017, from, http://suen.ed.psu.edu
17. Zumbo BD. Three generations of differential item functioning (DIF) analyses: Considering where it has been, where it is now, and where it is going. Language Assessment Quarterly. 2007; (4):223-233.