

ASSESSMENT OF PSYCHOMETRIC PROPERTIES OF USE OF ENGLISH MULTIPLE CHOICE ITEMS USING IRT AMONG UNDERGRADUATE STUDENTS IN BAYERO UNIVERSITY KANO, NIGERIA

SOFEME R. JEBSON¹, PH.D & ABUBAKAR R. UBA²

¹Department of Science Education, Federal University of Kashere, Gombe State.

²Department of Education, Sule Lamido University, Kafun Hausa, Jigawa State.

¹sofemejebson@yahoo.com; ²abbkarbabura@gmail.com

Abstract

Nigeria University System requires a student to pass all the courses of General Studies Programme (GSP) of which Use of English is one of them. This paper sought to assess psychometric properties of Use of English multiple choice items using IRT among undergraduate students in Bayero University Kano, Nigeria. Ex-post facto research design and multi-phase cluster sampling technique were employed to select the samples. The sampled size for each of Form A and Form B was 1000 undergraduate students. Based on the thrust of the study three null hypotheses were formulated and tested at 0.05 level of significance. The analyses were done using SPSS and IRTPRO. It was found that GSP 1201 Form A and Form B Use of English multiple choice items for 2021/2022 session did not differ significantly with respect to one- and two- based parameters. It can be concluded that Form A and Form B can be said to be equivalent of each other. Form A and Form B differs significantly with respect to three-parameter model. It was recommended that item writers of such examinations should use IRT in the design, analysis and scoring so as to enhance the information function.

Keywords: Assessment, Psychometric Properties, Use of English, General Studies Programme, IRT

Introduction

Educational measurement focuses either on measures of aptitude or measures of achievements and the data obtained from the measurements assist the teacher, the learner and even parents or guardians to make decisions about the progress of the learner. Therefore, for an individual to function effectively in the furtherance of national development and sustainability, he must have to possess an impressive command of the official language of communication which is English language. Therefore decision emanating from modern theory of measurement such as item response theory (IRT) becomes imperative in this regard.

IRT also known as the latent response theory which refers to a family of mathematical models that attempts to explain the relationship between latent traits (unobservable characteristics or attribute) and their manifestations (i.e. observed outcomes, responses or performance). As people's abilities vary their position on the latent construct's continuum changes and is determined by the sample of respondents and item parameters. An item must be sensitive enough to rate the respondents within the suggested unobservable continuum.

In psychometrics, the a-parameter, b-parameter, and the c-parameter are terms used to describe characteristics of items in IRT. The a-parameter (Discrimination parameter) measures an item's ability to distinguish between test-takers with different levels the latent trait (e.g. ability, attitude). High a-values indicate an item is highly discriminating, meaning it effectively differentiates between individuals with varying levels of the trait. The b-parameter (Difficulty parameter) represents the item's

difficulty or threshold. It is the point on the latent trait continuum where a test-taker has a 50% chance of answering the item correctly. High b-values indicate harder items. While the c-parameter (Guessing, parameter or Pseudo-guessing parameter) estimates the probability of a correct response due to guessing. It represents the minimum probability of getting the item correct, even for test-taker with low levels of the latent trait. Assessment of test items is paramount in the teaching and learning process in the tertiary in institutions for purpose of effectiveness in an evaluation data.

In Bayero University Kano, all the efforts invested in teaching and learning of the various concepts of the Use of English would have been wasted if there is nothing like the process of testing and measurement of various attributes of learners. One of the principal roles of a teacher in the education enterprise is to effectively instruct learners in order to bring about desirable changes in their behaviors. How does the teacher discover whether or not the desired changes in the learners' behaviors have actually taken place and in what direction? The qualities of test items develop by teachers would be of interest to any tester especially of Use of English. A way of ascertaining whether desirable changes in the learners' behaviors have been achieved is through a process that involves testing, measuring, assessing and evaluating (Jebson, 2017). The items that constitute the evaluation instruments vary in their psychometric properties. Other factors could be related to the undergraduate students' performance in GSP.

The items on the GSP 1201 examination in Bayero University Kano were organized into two forms which varied based on programs and these can conveniently be called GSP form A and form B. The items for students in the science-based departments which are being offered in the first semester were different from items in the arts-based departments in the second semester. This difference might likely cause a program factor in the behavior of the items.

The Use of English is among the compulsory courses that are offered under General Studies Programme in the Nigerian universities and other tertiary institutions due to incessant complaints by lecturers and employers of labour that many undergraduates and even graduates themselves lacked the ability to express themselves competently in the English language. It was rather shameful and actually disheartening that Nigerian graduates could not even write common application letters for employment in a language considered a national language (Nweye & Nwoye, 2016).

Cursory look at GSP 1201 Use of English examination results indicates high failure rate in Bayero University Kano. A number of questions need to be asked as to the possible causes of the failure in these examinations; are the items too difficult? These questions need to be answered to find out solution to the problem. Therefore, the paper assessed the psychometric properties of the GSP 1201 Use of English multiple choice items among undergraduate students in Bayero University Kano.

Based on the thrust this paper the following null hypotheses were tested at 0.05 level of significance.

H01: There is no significant difference in the mean of a-parameter on items in GSP 1201 Form A and Form B of 2021/2022 Academic Session in Bayero University Kano.

H02: There is no significant difference in the mean of b-parameter on items in GSP 1201 Form A and Form B of 2021/2022 Academic Session in Bayero University Kano.

H03: There is no significant difference in the mean of c-parameter on items in GSP 1201 Form A and Form B of 2021/2022 Academic Session in Bayero University Kano.

Methodology

The ex-post facto research design was employed since the focus of this study was to assess the differential psychometric properties of the Use of English Examination in 2021/2022 academic session in Bayero University Kano using IRT models. The design that was used to examine whether items in GSP Use of English examination set and administered by School of General and Entrepreneurship

Studies in Bayero University Kano possess different a-parameter, b-parameter and c-parameter by GSP Form A and Form B.

All undergraduate students who registered for compulsory course GSP 1201 Use of English of 2021/2022 academic session in Bayero University Kano, constituted the target population for the study. The total number of students who sat for the Use of English Examination of 2021/2022 academic session was nine thousand and fifty (9050) which comprises both male and female undergraduate students and their average age’s ranges from 18 to 22 years. The sample size employed was 1000. The size was deemed appropriate because the parameter model that has been used in this study requires large sample size in order to obtain accurate parameter estimates and this was achieved via the use of IRTPRO software (students version), in this regard, Research Adviser (2019) recommends the use of 500, 1000, 1500 and 2000 sample sizes for a study under Item Response Theory.

The multi-phase cluster sampling technique was employed to select the samples that adequately represented the population of this study. In phase- one, Hat and draw method without replacement was used to sample eight (8) from which eight faculties were sampled from the sixteen (16) Faculties to adequately represent the faculties in GSP form A and form B. In phase-two, the same Hat and draw method was used to sample thirty-two (32) Departments from the selected Faculties to adequately represent 80 Departments. In phase- three, one thousands (1000) undergraduates were sampled using systematic sampling method to select samples from the selected departments in the phase two to adequately represent the population in this study.

The researchers used IRTPRO software to determined indices of a-parameter, b-parameter and c-parameter from the data. While mean, standard deviation and t-test (independent samples) was used to test the hypotheses at 0.05 level of significance using SPSS version 24.

Results

Hypothesis One: There is no significant difference in the mean of a-parameter of 2021/2022 academic session GSP 1201 Form A and Form B in Bayero University Kano.

Table 1: Result of t-test for a Parameter Indices by GSP 1201 Form A and B in 2021/2022 Academic Session in Bayero University Kano.

Form	N	Mean	S.D	Std. Error Mean	t-value	Df	P value	Remark
Form A	1000	1.0183	1.32418	.13242	-.599	198	.550	Not Sig
Form B	1000	1.1237	1.15828	.11583				

Table 1 presents the summary of t-test computed with SPSS. The outcome from the analysis shows that the mean of a-parameter indices did not differ significantly between GSP 1201 Form A and Form B. The t- value was found to be -.599 with significant value (p-value) of 0.550. The null hypothesis stated above is thereby accepted, meaning there is no significant difference with regards to GSP form A and B in their respective mean a-parameter indices.

Hypothesis Two: There is no significant difference in the mean of b-parameter of 2021/2022 academic session GSP 1201 Form A and Form B multiple choice items in Bayero University Kano.

Table 2: Result of t-test for b Parameter Indices by GSP 1201 Form A and B in 2021/2022 Academic Session in Bayero University Kano.

Form	N	Mean	S.D	Std. Error Mean	t-value	Df	P value	Remark
Form A	1000	1.0183	1.32418	.29156	-1.291	198	.199	Not Sig
Form B	1000	1.1237	1.15828	.11583				

Form B	1000	1.1237	1.15828	.16745
--------	------	--------	---------	--------

Table 2 presents the summary of t-test computed with SPSS. The outcome from the analysis shows that the mean of b-parameter indices did not differ significantly between GSP 1201 Form A and Form B. The t value was found to be -1.291 with significant value (p-value) of 0.199. The null hypothesis stated above is thereby accepted, meaning there is no significant difference with regards to GSP form A and B in their respective mean b-parameter indices.

Hypothesis Three: There is no significant difference in the mean of c-parameter of 2021/2022 academic session items GSP 1201 Form A and Form B multiple-choice items in Bayero University Kano.

Table 3: Result of t-test for c Parameter Indices by GSP 1201 Form A and Form B in 2021/2022 Academic Session in Bayero University Kano.

Form	N	Mean	S.D	Std. Error		t-value	Df	P value	Remark
				Mean					
Form A	1000	.0637	1.91152	.19115		3.214	198	.002	Significant
Form B	1000	-.6741	1.27136	.12714					

Table 3 presents the summary of t-test computed with SPSS. The outcome of the analysis showed that the mean of c-parameter indices differ significantly between GSP 1201 form A and Form B. The t-value was found to be 3.214 with significant value (p-value) of .002. The null hypothesis stated above is thereby rejected, meaning there is significant difference with regards to GSP form A and B in their respective mean c-parameter indices

Discussion of Findings

The data collected for the study was analyzed according to hypotheses formulated for the study, the findings of this study was in agreement with the theoretical postulation of IRT which makes its assumption about a person’s behavior when responding to test items. It posits a single unobservable proficiency variable. This unobservable variable is referred to as ‘trait’ or ability which is defined as an innate inherited characteristic of the test takers that causes consistent performance on the test and influences response to test items (Bashir, 2018). It is therefore theoretically rational that each test item has one or more characteristics such as difficulty level and tendency to separate test takers based on their mastery level. IRT is basically concerned with the inter play of four aspects; the ability of the individual on the variable being assessed, the extent to which a test item discriminates between high and low-scoring groups, the difficulty of the item, and the probability that a person of low ability on that variable makes the correct response (Mustapha, 2018).

The finding from hypothesis one indicated that GSP 1201 Use of English multiple choice items of 2021/2022 academic session of Bayero University Kano showed no significant difference in the mean difficulty levels by GSP 1201 Form A and Form B. This finding contradicts the work of Mustapha (2018) who examined the difficulty levels of multiple choice items of the 2014 Jigawa State Mathematics Qualifying Examination, reporting that there was a significant difference in the mean difficulty levels of items by location. The finding from this study supports the work of Bashir (2018) who investigated the difficulty levels of items on English Language multiple choice examination of WAEC/SSCE in 2014, reporting that there is no significant difference in the mean difficulty levels by location.

The finding from hypothesis two indicated that GSP 1201 Use of English multiple choice items of 2021/2022 academic session of Bayero University Kano showed no significant difference in the mean

discrimination indices of GSP 1201 Form A and B. This finding contradicts the work of Abubakar (2018) who examined difficulty and discrimination indices of NECO multiple choice Mathematics examination in Malumfashi Education Zone. The finding of his study revealed that there is a significant difference in the discrimination indices by school location and school type. The finding from this study also supports the work of Bashir (2018) who investigated the difficulty levels of items on English Language multiple choice examination of WAEC/SSCE in 2014, reporting that there is no significant difference in the mean discrimination indices by location.

The finding of this study with respect to hypothesis three showed that GSP 1201 Use of English multiple choice items of 2021/2022 academic session differs significantly for Form A and Form B. This implies that the two forms are not equivalent of each other. This finding contradicts the work of Mustapha (2018) who examined the guessing indices (c-parameter) of multiple choice items of the 2014 Jigawa State Mathematics Qualifying Examination, reporting that there was no significant difference in the guessing indices of the items by location.

Conclusion

From the findings of this study, it was concluded that the GSP 1201 Use of English multiple choice items for Form A and Form B of 2021/2022 academic session of Bayero University, Kano did not significantly differ in terms of their psychometric properties with respect to one- and two- parameter models. While GSP 1201 Form A and Form B differ significantly with respect to three-parameter. The item difficulties and discriminating power of the items for the two examination Forms are similar or the same. The multiple choice items differentiates in terms guessing based on three- parameter model.

Recommendations

Based on the findings of this study the following recommendations are made:

1. The test constructors in the School of General and Entrepreneurship Studies of Bayero University, Kano to carry out appropriate item analysis from the beginning.
2. The School of General and Entrepreneurship Studies (SGES) should encourage pilot testing of the Use of English multiple choice items during test construction stage prior to administering the test to ensure that the item analysis is properly undertaken to judge the quality of the items in this examination.
3. There is the need to ensure equating of the Forms through the use of IRT models in design, analysis and scoring of the examination.

References

- Abubakar, L. (2018). *Analysis of Psychometric Properties of 2011 NECO Mathematics Examination in Malumfashi Education Zone, Katsina State*. An unpublished M.Ed. Dissertation, Bayero University Kano. Kano Nigeria.
- Bashir, A. (2018). *Analysis of the Psychometric Properties of 2014 WAEC English Examination in Minjibir Education Zone, Kano State*. An Unpublished M.Ed. Dissertation, Bayero University Kano, Nigeria.
- Ibrahim, A. (2017). An Empirical Analysis of One, Two and Three Parametric Logistic Models of Item Response Theory and Differential Item Functioning in Dichotomous and Ordinal Tests. *Journal of Educational Foundations (JEF)*.
- Jebson, S.R. (2017). Validity of Unified Tertiary Matriculation Examination: Implication of Test Theories Application to Stakeholders' Educational Assessment Literacy. *African Journal of Theory and Practice of Education Assessment (AJTPEA)*, Volume 5 November 2017, Pg 1-13. A Publication of Educational Assessment & Research Network in Africa (EARNiA).
- Mustapha, S. H. (2018). *Analysis of Items On 2014 Jigawa State Mathematics Qualifying Examination in Hadejah Education Zone*. An Unpublished M.Ed. Dissertation, Bayero University Kano State.
- Nweke, C. & Nwoye, C. (2016). Higher Education and General Studies in Nigeria: A Philosophical

Investigation.

Research, A.V. (2019). *Determining Sample Size for Research Activities*. Educational and Psychological Measurement, 30, 607-610.