

Assessment of Items Prone to Guessing in SSCE Economics Multiple-Choice Tests among Students in Kwara State, Nigeria

Mohammed Idris Jimoh

PhD., University of Ilorin, Ilorin, Nigeria jimoh.mi@unilorin.edu.ng

Dorcus Sola Daramola

PhD., University of Ilorin, Ilorin, Nigeria, olatunji.ds@unilorin.edu.ng

Jumoke Iyabode Oladele

PhD., University of Ilorin, Ilorin, Nigeria, oladele.ji@unilorin.edu.ng

Adaramaja Lukman Sheu

PhD., Federal University Gusau, Zamfara State, Nigeria, adaramaja4real@yahoo.com

The study investigated items that were prone to guessing in Senior School Certificate Examinations (SSCE) Economics multiple-choice tests among students in Kwara State, Nigeria. The 2016 West African Senior Secondary Certificate Examinations (WASSCE) and National Examinations Council (NECO) Economics multiple-choice test items were subjected to guessing parameters. This study adopted a descriptive survey research design, carried out in 25 randomly selected public Senior Secondary Schools in Kwara State. A multi-stage sampling technique was used to sample 768 Senior School III students. The instruments used for data collection were 2016 WASSCE and NECO multiple-choice Economics items. Guessing parameter was computed using WinGen IRT software package used to compute and answer research question 2 (What are those 2016 WASSCE and NECO Economics multiple-choice test items prone to guessing?). The hypothesis was tested using the chi-square statistic at 0.05 level of significance. One of the findings revealed that items prone to guessing in WASSCE and NECO were of the same rate. The finding also revealed that no significant difference existed in items guessed in WASSCE and NECO. One of the recommendations revealed that the classroom teachers and professional developers should avoid clauses such as “except”, “all of the above”, “none of the above” and “not” to mention a few. This could reduce the guessing tendency in multiple-choice items.

Keywords: guessing, senior school certificate examination economics, multiple-choice tests, senior secondary students

INTRODUCTION

Test consists of a set of uniform questions or tasks to which a student is to respond to independently and the result can be treated to provide a quantitative comparison of the performance among different test takers Obinne (2011). It is one of the indispensable measurement tools to assess educational outcomes. It is a tool for gauging knowledge of an individual learner in order to determine the ability of that individuals. It is one of the assessment techniques used to elicit students' behaviour with a view of making some inferences about a specific attribute of the test taker (Abiri, 2007). Test can be categorized into internal (teacher made) and external (public) test. In this study emphasis is placed on

Citation: Jimoh, M. I., Daramola, D. S., Oladele, J. I., & Sheu, A. L. (2020). Assessment of Items Prone to Guessing in SSCE Economics Multiple-Choice Tests among Students in Kwara State, Nigeria. *Anatolian Journal of Education*, 5(1), 17-28. <https://doi.org/10.29333/aje.2020.512a>

public test in Nigeria such as Senior School Certificate Examination (SSCE). It places value judgment on the achievement of individual candidate at the end of their.

In Nigeria, at the Senior Secondary level, there are public examinations such as the West African Senior School Certificate Examinations (WASSCE), National Examinations Council (NECO) and National Business and Technical Examination Board (NABTEB) that are expected to conduct credible examinations, better test scores and award reputable certificate to their candidates. The stakeholders in education industry, such as teachers, parents, students and tertiary institutions usually feel happy, confident and proud when they are assured that the certificates they issue will be true reflections of the candidate's abilities. The fundamental aim of these examination bodies is to measure the extent to which learners have addressed educational objectives as stated in the SSCE national curriculum. The tests administered by these examination bodies are believed to have been standardized because they followed the principles of test construction. These examination bodies are tasked with maintaining a common standard in the development and administration of their various examinations. The psychometric properties (quality of test instrument) are always examined through responses of the examinees that have similar characteristics before being used for the purpose the test is meant for. Thus, items could be analysed through different models (Adedoyin & Mokobi, 2013).

In the conduct of SSCE, students are usually examined using essay, objective and practical tests in case of science and vocational subjects. However, multiple-choice test is one of the test format commonly used in objective tests to examine candidates irrespective of their discipline. A multiple-choice test requires a student to select a correct answer to a stem from a set of options. The stem presents the item as a problem to be solved or an incomplete statement to be completed and provide with plausible options in which an examinee can choose from, with only one correct answer called the key and the incorrect answers referred to as distractors (Morrison & Free, 2001; Kolawole, 2007). Multiple-choice items are usually scored dichotomously (either right or wrong).

In a multiple-choice test, examinees' responses could be classified into the following three types: (a) responses reflecting examinees' true ability; (b) correct responses made through lucky guesses; and (c) incorrect responses derived from insufficient knowledge of the concepts, carelessness or distraction (Yen, Ho, Laio, Chen, & Kuo, 2012). It is assumed that multiple-choice test item is the most reliable instrument that can be used to collect data on examinees' ability. This is because it is free from examiners' biased. For this test format to be more effective and serve the purpose it meant for therefore, test items must be structured in such a manner that will not give different interpretation to examinees of the same ability. In this case, the test experts must be very careful to construct multiple-choice items that will give a test of time for an average test taker. Hence, the way a multiple-choice item is constructed exposes that item to guessing. In the case of a typical multiple-choice test item, this probability will be small for examinees of low ability and large for examinees of high ability. This indicates that for an examinee with low ability level, the probability of correct response will be near zero while it increases until at the highest levels of ability which the probability of correct response approaches 1. It should be noted that the figures only show a range of ability from -3 to +3, the theoretical range of ability is from negative infinity to positive infinity (Baker, 2001).

There are two main frameworks by which a test and the items it contains can be analysed. It could either be through the Classical Test Theory (CTT) or Item Response Theory (IRT). The Classical Test Theory (CTT) has been in existence for long and being used in the education field and in test development (Bechger, Gunter, Huub & Bèguin, 2003). The fundamental feature of Classical Test Theory is the formulation of every observed score (X) as a function of the individual's True score (T) and some random measurement Error (E) which can be mathematically written as {i.e. $X = T + E$ }. In CTT, the raw score is the summation of responses of a person to a test (Haertel, 2006). Despite the popularity of CTT and its usefulness in test development, it is fraught with so many limitations. One of

the weaknesses of CTT is that the theory depend on the group of examinees being tested and, therefore, do not adequately reflect the measurement quality of the test items (Adedoyin & Mokobi, 2013). It must be noted that each item in a test has three parameters either in CTT which are item difficulty, discrimination and guessing. However, guessing could be computed using equation 1.

$$S = R - \left(\frac{W}{O-1}\right) \dots\dots\dots(1)$$

Where S = the actual score of an examinee:

R = number of questions answered rightly;

W = number of questions answered wrongly;

O = number of options for an item; and

1 = is a constant figure.

Item Response Theory (IRT) is an important innovation in the field of test psychometrics. The theory (IRT) explains an examinee’s response to test items via a mathematical function based on his/her ability (Al-A'ali, 2006). The theory wants to see the level of interaction of the examinees with the items in the test which is based on the idea of the probability of correct response to an items (Magno, 2009). It must be noted that in IRT too, each item in a test also has three parameters. The first model is one-parameter model which is named Rasch Model. It can be used to find out the difficulty level of an item in relation to the testee’s ability which is denoted by “b” in IRT equation. However, a typical ability range have the range $-3 \leq b \leq +3$. The second model is two-parameter model which finds out the discrimination index of an item after item difficulty parameter have been considered (Obinne, 2012). The discriminating index is denoted by “a” which expresses how well an item differentiate (discriminate) from one examinee to another with different ability levels (Adedoyin & Mokobi, 2013).The last is three-parameter model (guessing an item correctly) which tells the probability of an individual guessing a multiple-item correctly with known examinee ability level, after identifying difficulty and discrimination index which is denoted by “c” in the IRT equation as shown in equation 2 (Baker, 2001).

The equation for the three-parameter model is

$$p(\theta) = \{c + (1 - c) \left\{ \frac{1}{(1 + e)^{(1+a\theta) - a(b-c)}} \right\} \dots\dots\dots(2)$$

Where: “b” is the difficulty parameter

“a” is the discrimination parameter

“c” is the guessing parameter

“ θ ” is the ability level

“e” is the constant 2.718

The parameter “c” is the probability of getting the item correct by guessing alone. It is important to note that by definition, the value of “c” does not vary as a function of the ability level. Thus, the lowest and highest ability examinees have the same probability of getting the item correct by guessing. The parameter c has a theoretical range of $0 \leq c \leq 1.0$.

To ensure high quality test items, such items are expected to measure just one construct, but if an item does not measure one construct, the item reduces the validity of the test to assess the construct being measured. Hence, a well standardized test should therefore, not be too difficult or too simple. One characteristic of any good test item is that high-ability candidates should be able to answer an item

correctly more frequently than lower-ability test takers (Adedoyin & Mokobi, 2013). Items with high “b” values are hard items, that is, values of “b” greater than 1 indicate a very difficult item and low-ability examinees are unlikely to answer it incorrectly (Adedoyin & Mokobi, 2013). Lastly, the 3-parameter logistic model uses item difficulty, item discrimination and the extent to which examinees can guess the correct answer. It expresses the likelihood that an examinee with very low ability can be able to guess the correct response to an item (Adedoyin & Mokobi, 2013). It represents the possibility of low ability examinees answering an item correctly due to chance (Crocker & Algina, 1986).

The probability is half-way between the value of “c” and 1.0. Thus, the difficulty parameter “b” defines the point on the ability scale where the probability of correct response is halfway between this floor and 1.0. The discrimination parameter “a” can still be interpreted as being proportional to the slope of the item characteristic curve at the point $\theta = b$. However, under the three-parameter model, the slope of the item characteristic curve at $\theta = b$ is actually $a(1 - c)/4$. It is important to note that by definition, the value of “c” does not vary as a function of the ability level. Thus, the lowest and the highest ability examinees have the same probability of getting the item correct by guessing (Baker, 2001).

Guessing is a standard test-writing strategy used by the examinees in answering a multiple-choice test. This strategy provides an opportunity to have an item marked correct even when the examinee has insufficient knowledge of the subject matter (Obinne, 2012). Mehrens & Lehmann (1984) identified two types of guessing: 1). Blind guessing and 2). Informed guessing: Blind guessing is a type of guessing where an examinee chooses an answer at random from among the alternatives given while informed guessing is a type of guessing where the examinee chooses an answer based on his/her knowledge and abilities to pick the right answer from the alternatives.

In multiple-choice item, guessing tendency is very rampant. According to Alonge, (2003), the tendency of an examinee to guess is a function of the examinee’s ability and the test item difficulty level. For instance, if a four-option multiple-choice item is prepared, the probability of guessing an item correct is 0.25 (25%). One way of determining if guessing occurs is to examine how examinee with low ability, answer the most difficult items in the test. Ojerinde (2013) argued that in a multiple-choice items, the scores obtained on such an item should be corrected by a special expression which takes into consideration this probability of guessing the item correct.

In the earlier studies much emphasis was placed on variable that could influence academic performance, among them are nature and nature, but today, researchers have started redirecting their attention towards improving the fairness of tests across various subgroups of examinees (Adedoyin, 2010). A fair test is one that is comparably valid for all groups and individuals and that affords all examinees an equal opportunity to demonstrate the skills and knowledge which they have acquired (Roever, 2005).

Obinne, (2012) researched on using IRT in determining test item prone to guessing. He observed that more of the Biology test items constructed by the WAEC were found to be prone to guessing than those items of NECO. Bandele & Adewale (2013) conducted a study on comparative analysis of the item difficulty levels of WAEC, NECO and (NABTEB) Mathematics achievement examinations. They concluded in their findings that that the three Examinations (WASSCE, NECO and NABTEB are comparable in the candidates’ performance. They also found that NECO examinations are more difficult than that of WAEC and NABTEB. Contrarily, Olaunji & Owolabi (2009) found that students performed better in 2005 NECO Multiple Choice Economics examination than they did in WAEC Multiple Choice Economics examination for the same year; this performance was attributed to the number of options supplied to test item. WAEC supplies four options while NECO supplies five

options. It was found that some of the distracters supplied to NECO items for that year were not plausible. Hence, the students' good performance could be due to guess.

Brown & Abdulnabi (2017) researched on evaluating the quality of higher education instructor-constructed multiple-choice tests: impact on student grades. The study showed that the use of IRT item analysis has a potential beneficial impact on overall course grades and number of students passing. Tommy & Udo (2019) examined item difficulty and student ability parameters of national examinations council's Biology examinations using the Rasch measurement model in Nigeria. The finding revealed that item difficulty parameters were appropriate but were not arranged hierarchically from the least difficult to most difficult item and students' ability parameters were appropriately estimated by the examination items.

The rationale for carrying out this study is derived from the position held by Obinne, (2011) that there is public complaint and conception SSCE public examinations (WASSCE, NECO and NABTEB) in Nigeria that there is superiority of test items in one over the other. It is noted that this study therefore, investigated which items in WASSCE and NECO are vulnerable to guessing using IRT. The finding of the study might tell the integrity of Economics multiple-choice tests items constructed by the West African Examinations Council and the National Examinations Council that were administered to Senior Secondary School students in 2016. In order to carry out this study, two research questions and one hypothesis were raised, answered and tested respectively.

Research Questions

1. What are the ability levels of students that have attempted 2016 WASSCE and NECO multiple-choice Economics items?
2. What are those 2016 WASSCE and NECO Economics multiple-choice test items prone to guessing?

Research Hypothesis

H₀₁: There is no significant difference in the item guessed in 2016 WASSCE and NECO Economics multiple-choice test items.

METHOD

The study adopted descriptive survey research design. This research design was, therefore, considered suitable for this study because it enabled the researcher to obtain valid and reliable information on the variables under consideration and being able to describe actually the way they were observed. It allowed the researchers to determine and judge the extent to which the students engaged in guessing in 2016 WASSCE and NECO Economics multiple-choice test items.

The population of this study covered all public Senior Secondary School students of Economics in the 16 Local Government Areas (LGAs) in Kwara State. There are 30 Senior Secondary School in Asa LGA, 19 in Baruteen LGA, 17 in Edu LGA, 15 in Ekiti LGA, 50 in Ifelodun LGA, 38 in Ilorin East LGA, 28 in Ilorin South, 41 in Ilorin West LGA, 43 in Irepodun, 18 in Isin LGA, 12 in Kaima LGA, 28 in Moro LGA, 18 in Offa, 14 in Oke-Ero LGA, 23 in Oyun and 13 in Patigi LGA (Kwara State Ministry of Education and Human Development, 2016). The target population consisted of 25, 177 Senior Secondary School III students of Economics in public secondary schools in the 16 Local Government Areas of Kwara State (Kwara State Ministry of Education and Human Development, 2016). This was because of the SS III students were in their final stage of the Senior Secondary programme, where the researchers believed they might have covered the major parts of Economics curriculum as they were preparing for their Senior School Certificate Examinations.

Multi-stage sampling technique was adopted to sample respondents at different stages of sampling. At the first stage, fifty percent (50%) of the LGAs representing eight out of sixteen LGAs were sampled using simple random sampling. To select these LGAs, names of the 16 LGAs were written on pieces of paper, wrapped, put in a container and mixed thoroughly together. A student was asked to pick a paper one after the other until eight papers were picked. Those papers picked were opened and the names of LGAs found in them were used as sampled LGAs in this study. Therefore, the sampled LGAs are Asa LGA, Edu LGA, Ilorin East LGA, Irepodun LGA, Kaima LGA, and Offa LGA.

In each of the LGAs sampled, proportionally, 25% of the Senior Secondary Schools were sampled using simple random sampling technique, thus, making a total of 25 schools sampled. In the 25 sampled schools, 30 SS III students of Economics were randomly sampled in each school. Thus, 750 students of Economics participated in the study.

Unfortunately, in the course of marking students' scripts, the researchers discovered that 37 scripts were not properly completed and they were discarded. Therefore, 713 scripts were used in the analysis which represented the sample size for this study. The mortality rate of discarded scripts was calculated and put at 4.93%. The May/June 2016 WASSCE and June/July NECO multiple-choice Economics papers were used as the instruments for data collection which contained 50 and 60 items respectively.

It must be noted that the instruments were constructed by the West African Examinations Council (WAEC) and National Examinations Council (NECO) which are known to be organized examining bodies. The items were developed, moderated, validated and used on the candidates in 2016 Senior School Certificate Examinations. The items were already subjected to the processes of validation and standardization. The items were assumed to be at least moderately ok in terms of curriculum content coverage. The psychometric properties of these instruments could not be obtained from these bodies. The researchers, therefore, carried Content Validity (CV) of the instruments. Three and four teachers of Economics scrutinized the answers of Economics test items in WASSCE and NECO respectively. Item Level Content Validity (I-CV) was computed to derive Scale Content Validity (S-CV). Thus, Item Level Content Validity Scale (S-CVI) obtained in WASSCE was 0.95 and NECO was 0.89.

The researchers administered the instruments to sampled students. Data collected were scored dichotomously. Three parameter logistic model was adopted using WinGen IRT software package. The difficulty, discriminant and guessing parameters were computed. However, the guessing parameter of the two set of test items was used to answer research question two using three-parameter logistic model while the chi-square statistic was used to test the only hypothesis postulated at 0.05 level of significance. The difficulty and discriminating indices were calibrated using 0 and 1 as the mean and standard deviation respectively during the process of the analysis. In the output obtained, the results that best answer this research question 2 were extracted and reported in table 1.

FINDINGS

Research Question 1: *What are the ability levels of students sampled that sat for 2016 WASSCE and NECO multiple-choice Economics items?*

In order to answer this research question 1, ability levels of the examinees of the students of Economics sampled were computed using histogram via WinGen IRT software package. The report of the analysis is shown in Figures 1 and 2. The theta (ability) index of the students who responded to the set of items in the two tests was generated. The principle behind ability levels under IRT ranges from -3 to +3 as reported in Figures 1 and 2.

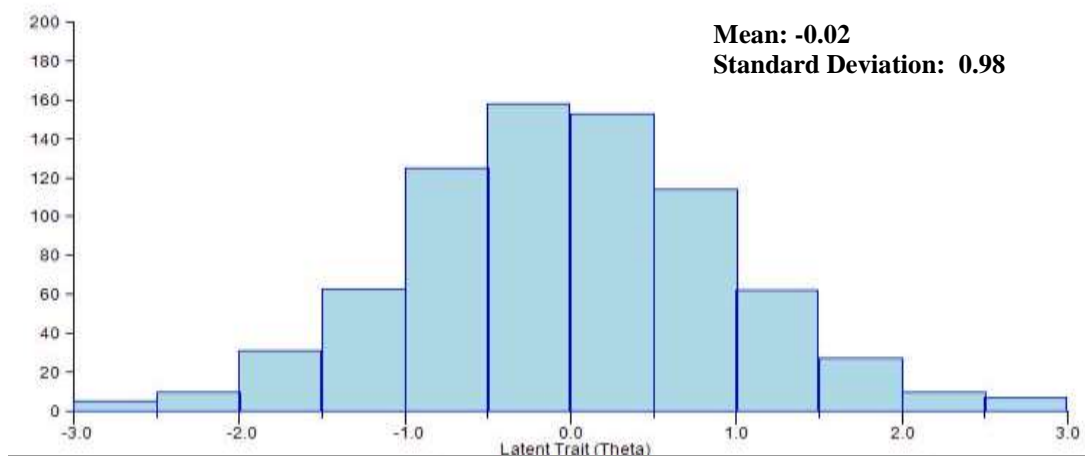


Figure 1
Examinees Abilities (Theta) of 2016 WASSCE Economics Test Items

The mean ability (theta) of the examinees in WASSCE was found to be between a mean of -0.02 and a standard deviation of 0.98, with a range from -3.16 to 3.29 as shown in figure 1. This implies that the examinees (SS III students of Economics) that participated in the study assumed a normal distribution of ability.

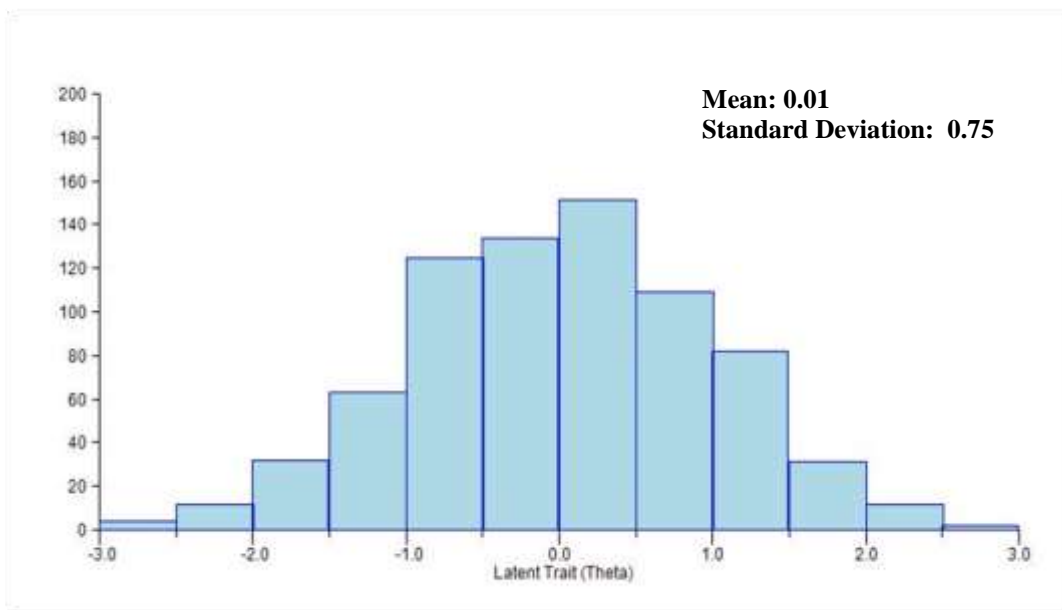


Figure 2
Examinees Abilities (Theta) of 2016 NECO Economics Test Items

As shown in figure 2, the mean ability (theta) of the examinees in NECO was found to have mean score of 0.01 and standard deviation of 0.75 with a range from -3.31 to 3.15. Based on reports in figure 1 and 2, it implies that the examinees (SS III students of Economics) that participated in the study also assumed normally distribution ability.

Research Question 2: *What are those 2016 WASSCE and NECO Economics multiple-choice test items are prone to guessing?*

In order to answer this research question 2, the response of the examinees on WASSCE and NECO multiple-choice items sampled were subjected, three-parameter logistic model using WinGen ITR software package. Guessing values “c” ranged from 0.00 to 0.40 was used in determining which item is prone to guessing. Warm, (1978) observed that item with the “c” value of 0.35 or greater indicates that the items are prone to guessing while items below 0.35 are not prone to guessing. The results of the analysis that best answer this research question were extracted from the output and reported in table 1.

Table 1

2016 WASSCE and NECO Economics Multiple-Choice Tests Items Guessed Index

2016 WASSCE Multiple-choice Economics Test				2016 NECO Multiple-choice Economics Test			
Items Guessing Index				Items Guessing Index			
Item1	0.024*	Item26	0.000*	Item1	0.000*	Item31	0.000*
Item2	0.000*	Item27	0.000*	Item2	0.000*	Item32	0.000*
Item3	0.672	Item28	0.153*	Item3	0.856	Item33	0.000*
Item4	1.813	Item29	0.000*	Item4	0.000*	Item34	0.605
Item5	0.459	Item30	0.246*	Item5	1.717	Item35	0.000*
Item6	0.000*	Item31	0.000*	Item6	0.000*	Item36	0.000*
Item7	0.000*	Item32	0.000*	Item7	0.000*	Item37	0.000*
Item8	0.000*	Item33	0.778	Item8	0.000*	Item38	0.000*
Item9	1.208	Item34	0.643	Item9	0.420	Item39	0.239*
Item10	0.000*	Item35	1.214	Item10	0.000*	Item40	0.709
Item11	0.000*	Item36	0.000*	Item11	0.000*	Item41	0.370
Item12	0.000*	Item37	0.000*	Item12	0.000*	Item42	0.000*
Item13	0.866	Item38	0.000*	Item13	1.461	Item43	0.000*
Item14	0.000*	Item39	0.000*	Item14	0.000*	Item44	1.112
Item15	0.000*	Item40	0.000*	Item15	0.218*	Item45	1.064
Item16	0.807	Item41	0.000*	Item16	0.298*	Item46	0.885
Item17	0.000*	Item42	0.392	Item17	0.234*	Item47	0.000*
Item18	0.806	Item43	0.000*	Item18	0.000*	Item48	0.463
Item19	0.000*	Item44	1.511	Item19	0.000*	Item49	0.973
Item20	0.383	Item45	0.217*	Item20	0.000*	Item50	1.027
Item21	1.052	Item46	0.422	Item21	0.715	Item51	0.000*
Item22	0.000*	Item47	0.328	Item22	0.000*	Item52	0.000*
Item23	1.149	Item48	0.000*	Item23	0.000*	Item53	0.000*
Item24	0.000*	Item49	0.000*	Item24	0.000*	Item54	0.000*
Item25	0.000*	Item50	0.384	Item25	0.000*	Item55	1.837
NB: * = Not prone to guessing				Item26	0.000*	Item56	0.000*
				Item27	0.000*	Item57	0.640
				Item28	0.000*	Item58	1.203
				Item29	0.572	Item59	0.429
				Item30	1.160	Item60	0.366
				NB: * = Not prone to guessing			

The table 1 reveals that 2016 WASSCE Economics multiple-choice test items prepared by the West African Examinations Council, were subjected to guessing indices, item 3 (0.672), item 4 (1.813), item

5 (0.450), item 9 (1.208), item 13 (0.866), item 16 (0.807), item 18 (0.806), item 20 (0.383), item 21 (1.052), item 23 (1.149), item 33 (0.778), item 34 (0.643), item 35 (1.214), item 44 (1.511), item 46 (0.422), item 47 (0.328) and item 50 (0.384) were prone to guessing while other items are not prone to guessing. This implies that 18(36%) of the 2016 WASSCE Economics multiple-choice test items were prone to guessing while 32 (64%) were not prone to guessing.

In the same vein, in 2016 NECO Economics multiple-choice test items prepared by the National Examinations Council, were subjected to guessing indices, This indicates that item 3 (0.856), item 5 (1.71), item 9 (0.420), item 13 (1.416), item 21 (0.715), item 29 (0.472), item 30 (1.160), item 34 (0.605), item 40 (0.709), item 41 (0.370), item 44 (1.112), item 45 (1.064), item 46 (0.885), item 48 (0.463), item 49 (0.973), item 50 (1.027), item 55 (1.837), item 57 (0.640), item 58 (1.203) item 59 (0.429) and item 60 (0.336) were prone to guessing while other items are not prone to guessing. This implies that 21 (35%) of the 2016 NECO Economics multiple-choice test items were prone to guessing while 39 (65%) were not prone to guessing.

Hypothesis Testing

H₀₁: There is no significant difference in the item guessed in 2016 WASSCE and NECO Economics multiple-choice test items.

The guessing indices of WASSCE and NECO multiple-choice Economics items were computed using chi-square statistic and reported in table 2.

Table 2

The X² Statistics Comparing Difference in Guessing Parameter of WASSCE and NECO Multiple-choice Economics Items

Examinations	Guessing Parameter		Total	Cal. X ²	Sig.	Remark
	Prone	Not Prone				
WASSCE	18 (17.7)	32 (32.3)	50 (50.0)	0.012	0.913	NS
NECO	21 (21.3)	39 (38.7)	60 (60.0)			
Total	39 (39.0)	71 (71.0)	110 (110.0)			

P > 0.05

As indicated in table 2, the chi-square value of 0.012 is obtained computed at 0.05 level of significance and obtained a p-value of 0.913. Since the p-value (0.913) is greater than the chosen 0.05 level of significance, hence, the null hypothesis is not rejected. This implies that there is no significant difference in the multiple-choice Economics items prone to guessed in 2016 WASSCE and NECO Examinations.

DISCUSSION

The first finding revealed that test takers (SS III students of Economics) that participated in the study assumed equal and normally distributed abilities. This implies that SS3 students in Kwara State seem they were well prepared for Senior School Certificate Examination conducted by WAEC and NECO which implies that they might have overcome some exam tensions associated with these examinations. Critical observation of the graphs shows that the values of the ability are symmetrically distributed where most observations clustered around the central peak and the probabilities for values inability further away from zero taper away equally in both directions. The extreme values in both tails of the distribution is similar.

The second finding revealed that 18(36%) and 21 (35%) of the 2016 WASSCE and NECO Economics multiple-choice test items respectively were prone to guessing while 32 (64%) and 39 (65%) of WASSCE and NECO respectively were not prone to guessing. This tells that the pattern of constructing multiple-choice Economics items in WASSCE and NECO is the same. This finding could

be attributed to the critical observations made by the researchers that some of the stems in those items guessed revealed that they were negatively worded. These observations probably allowed the test takers to misinterpret those items correctly and pick the right answers. This finding is not in line with Obinne (2012) observed that more of the Biology test items constructed by the WAEC were found to be prone to guessing than those items of NECO. This finding is also not in line with Bandele & Adewale's (2013) submission that NECO examinations are more difficult than that of WAEC and NABTEB.

The last finding revealed that there was no significant difference in the 2016 WASSCE and NECO Economics multiple-choice test items that were prone to guessing. The explanation for this pattern of the result might be attributed to the fact that the items in WASSCE and NECO were generated from the same Economics national curriculum published by the Nigerian Educational Research and Development Council. This might also be attributed to the fact that the test takers were exposed to the same learning environment, educational facilities, taught by teachers having similar educational qualifications with similar pedagogy and using the same national curriculum. This dictates that the test takers are expected to perform equally since the items were drawn from the same curriculum. This finding supports the finding of Bandele & Adewale (2013) when they observed that none of this examination is seen as being lower in standard than the other and there should be no discrimination in the recognition of WAEC, NECO, and NABTEB certificates.

CONCLUSION

Based on the findings obtained, it can be concluded that SS3 students in Kwara State have an asymmetrical distribution of ability clustered around the central peak and taper away in both directions. The extreme values in both tails of the distribution is similar. It is also concluded that 18(36%) and 21 (35%) of the 2016 WASSCE and NECO Economics multiple-choice test items respectively were prone to guessing because the items were negatively worded. It also concluded that there was no significant difference in the 2016 WASSCE and NECO Economics multiple-choice test items that were prone to guessing because the items in WASSCE and NECO were generated from the same Economics national curriculum published by the Nigerian Educational Research and Development Council.

RECOMMENDATIONS

It is could, therefore, be recommended that:

1. The classroom teachers and professional developers preparing multiple-choice Economics items should avoid clauses such as “except”, “all of the above”, “none of the above” and “not” to mention a few. This could reduce the guessing tendency in multiple-choice items.
2. Item guessing is believed to be unethical and immoral because it is a form of gambling. This act could be discouraged by means of instructions on the question paper to penalize those who guess incorrectly.
3. To avoid or reduce items that could be prone to guessing, classroom teachers and professional developers should be consistent in grammatical structure between stem and options; avoid one of the options that is significantly longer than the rest
4. Classroom teachers and professional developers should endeavour to distribute the key to each item fairly and evenly among the options in a test so that in two, three or four questions that follow each other should place their answers in the same position in the options.
5. When a test-taker is faced with difficult items, it is advice able to carry out what could be regarded as educated guessing which the process of is knocking out as many options as possible from the

plausible answers and therefore, make a guess from the last two options remaining. The more choices an examinee can cross off, the better the chances of picking the right answer.

REFERENCES

- Abiri, J. O. O. (2007). *Elements of evaluation measurement and statistical techniques in education*. Unilorin Press: University of Ilorin, Nigeria.
- Adedoyin, O. O., & Mokobi, T. (2013). Using irt psychometric analysis in examining the quality of junior certificate mathematics multiple choice examination test items. *International Journal of Asian Social Science*, 3(4), 992-1011.
- Al-A'ali, M. (2006). *IRT-item response theory assessment for an adaptive teaching assessment system*. Proceedings of the 10th WSEAS international conference on applied mathematics, Dallas, Texas, USA, 518–522.
- Alonge, M. F. (2003). Assessment and examination: the pathways to educational development. *Abstract conference*. Retrieved from <https://www.brookings.edu>.
- Baker, F. B. (2001). *The basics of item response theory*. United States of America: ERIC Clearinghouse on Assessment and Evaluation.
- Bande, S. O., & Adewale, A. E. (2013). Comparative analysis of the item difficulty levels of WAEC, NECO and NABTEB mathematics achievement examinations. *Mediterranean Journal of Social Sciences*, 4 (2), 761-765.
- Bechger, T., Gunter, M. H., & Bèguin, A. (2003). Using classical test theory in combination with item response theory. *Applied Psychological Measurement*, 27(5), 319 – 334.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- Gavin, T. L. B., & Hasan, H. A. A. (2017). Evaluating the quality of higher education instructor-constructed Multiple-choice Tests: impact on student grades. *Journal Frontiers in Education*, 2(24), 1-12.
- Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational measurement*. Washington, DC: American Council on Education and Praeger.
- Hambleton, R. K., Robin, F. & Xing, D. (2000). *Item response models for the analysis of educational and psychological test data*. In H. Tinsley, & S. Brown (Eds). *Handbook of applied multivariate statistics and modelling*. San Diego, CA: Academic Press.
- Kolawole E. B. (2007). A comparative analysis of psychometric properties of two Nigerian examining bodies for senior secondary schools in mathematics. *Research Journal of Applied Sciences*, 2(8), 913 - 915.
- Kwara State Ministry of Education and Human Capital Development (2016). Kwara State schools' census. Kwara State Ministry of Education and Human Capital Development.
- LeBeau, B. & Mcvay, A. (2017). Validity of the three parameter logistic item response theory model for field test data. Iowa Testing Program. The University of Iowa, 1-28.
- Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *The International Journal of Educational and Psychological Assessment*, 1(1), 1–11.

- Mehrens, W. A., & Lehmann, I. J. (1984). *Measurement and evaluation in education and psychology*. New York: Holt, Rinehart and Winston.
- Michael, P. K., Daniel, Y. F., & Steffi, N. L. L. (2011). Factors promoting pupils' academic performance in privately owned junior high schools in Accra, Ghana. *International Journal of Sociology and Anthropology*, 3(8), 280-289.
- Morrison, S. & Free, K. (2001). Writing multiple-choice test items that promote and measure critical thinking. *Journal of Nursing Education*, 40, 17-24.
- Nering, M. L., Ostini, R. eds. (2010). *Handbook of polytomous item response theory models*. Taylor & Francis.
- Nworgu, B. G. & Agah, J. J. (2012). Application of three- parameter logistic model in the calibration of a Mathematics achievement test. *Journal of Educational Assessment in Africa*, 7, 162 – 172.
- Obinne, A. D. E. (2011). A psychometric analysis of two major examinations in Nigeria: standard error of measurement. *International Journal of Education Science*, 3(2),137-144.
- Obinne, A. D. E. (2012). Using irt in determining test item prone to guessing. *World Journal of Education*. 2(1), 91-95.
- Oluwatimilehin, J. T. B., & Owoyele J. W. (2012). Study habits and academic achievement in core subjects among junior secondary school students in Ondo State, Nigeria. *Bulgarian Journal of Science and Education Policy (BJSEP)*, 6(1), 155–169.
- Ojerinde D. (2013). *Classical test theory vs item response theory: An evaluation of the comparability of item analysis results*. A guest lecture presented at the Institute of Education, University of Ibadan.
- Olatunji, D. S., & Owolabi, H. O. (2009). Difficulty and discrimination of senior school certificate economics objective test items with various option formats. *Ilorin Journal of Education*, 28, 50-64
- Ostini, R.; Nering, M. L. (2005). *Polytomous Item Response Theory Models*. Quantitative Applications in the Social Sciences. SAGE.
- Owolabi, H. O., Onuka, A. O. U., Ogunjimi, M. O., Mustapha, y., Adaramaja, L., & Daramola, D. (2017). Options functionality of English language tests for candidates' selection and placement into tertiary institutions in Nigeria: A pilot study. *Journal of Educational Assessment in Africa*, 12, 265-272.
- Tommy, U. E., & Udo, E. M. (2019). Examining item difficulty and student ability parameters of national examinations council's Biology examinations using the Rasch measurement model in Nigeria. *British Journal of Education*, 7(8), 66-80.
- Roever C (2005). "That's not fair!" fairness, bias, and differential item functioning in language testing. Retrieved from <http://www2.hawaii.edu/~roever/brownbag>.
- Yen, Y., Ho, R. Laio, W., Chen, L., & Kuo, C. (2012). An empirical evaluation of the slip correction in the four parameter logistic models with computerized adaptive testing. *Applied Psychological Measurement*, 36, 75–87.