

# ANALYSES OF THE STANDARD ERROR OF MEASUREMENT FOR SECONDARY SCHOOL CERTIFICATE QUALIFYING EXAMINATION AND SENIOR SCHOOL CERTIFICATE EXAMINATIONS ECONOMICS TEST ITEMS IN KANO, NIGERIA

Aminu Idris Harbau, Ibrahim Shu'aibu & Abubakar Abdullahi Yamusa

## Abstract

*The study is aimed at analysing the Standard Error of Measurement (SEM) of economics test items of 2015 SSCQE and 2016 SSCE economics test items employing the Item Response Theory (IRT). An expo facto research design was employed in the analysis. The design was adopted because the study used secondary data collected from the Kano State Education Resource Department (KERD) and National Examination Council (NECO). It aimed at establishing how standard the items were. The population of the study consists of the fifteen thousand four hundred and eighty one (15,481) students who sat for 2015 economics SSCQE and 2016 SSCE in Kano State owned senior secondary schools. It consisted of ten thousand nine hundred and fifty (10,950) male students and four thousand and thirty one (4,531) female students. A sample size of 1,000 students was selected using multistage sampling technique. The data collected was coded and run using IRTPRO package to determine the SEM indices of Kano State 2015 SSCQE. Findings from the study revealed that the SEM of 2015 SSCQE and 2016 SSCE economics test items based on two-parameter logistic (2PL) model was high. This indicates consistency in measuring student ability in economics. The study found that the massive failure in Kano State Certificate Qualifying Examinations over the years was not caused by the quality of the SSCQE items but by other variables which are not related with psychometric properties of the test items. It was therefore concluded that there was no superiority in terms of item quality between the items generated by KERD. As such performance in SSCQE may determine performance in NECO. Given the various advantages of IRT over the other popular measurement frameworks, the government should encourage KERD to adopt this measurement framework. IRT can enable KERD to place examinees on the correct ability level to describe the test items and the abilities of the examinees. This will enable KERD to meet the best practices of global examination standard and overcome the lingering superiority complex of one examining body over the others.*

**Keywords:** Standard error of measurement, IRT, KERD, Qualifying, NECO, Economics

## **Introduction**

The National Examinations Council (NECO), West African Examinations Council (WAEC), National Teachers Institute (NTI), National Business and Technical Education Board (NABTEB) as well as the Joint Admission and Matriculation Board (JAMB) are the major examining bodies saddled with organizing and administering standardized achievement tests in Nigeria. Each state has its own body that organizes and administers standardised examinations, entrance examinations, placement examinations and mock examinations. In Kano State, the body responsible for planning, organizing and administering standardized examinations is Kano State Education Resource Department (KERD). It is responsible for all examinations at the state upper and lower basic levels. It is also responsible for organizing and Senior School Certificate Qualifying Examination (SSCQE). The objective is to prepare Senior School II (SS II) students towards their final Senior School Certificate Examinations (SSCE) and motivate candidates for higher performance by paying their SSCE registration fees.

In examinations, such as the ones listed in the preceding discussion, reliability and content validity, according to Suruchi and Rana (2014), are the important features in determining psychometric quality. To ensure the validity and reliability of test items, a careful choice of test items in respect of subject contents and level of difficulty is essential as the quality and efficiency of a test is determined by the individual items. Item analysis from a Classical Test Theory (CTT) point of view is an attempt to measure the efficacy of each item in terms of its degree of discrimination ability and difficulty level. Item analysis helps in choosing the best items in a test by retaining the good and eliminating the poor test items (Suruchi & Rana, 2014). Item Response Theory (IRT) models use Standard Error of Measurement (SEM) in place of reliability. SEM is related to test reliability because reliability indices help us to find SEM. This makes SEM directly related to the reliability of a test - the larger the SEM, the lower the reliability of the test, and vice-versa.

Economics, which is one of the subjects students take in these examinations, is one of the major subjects taught in Nigerian schools. It came into the Nigerian school curriculum in 1967 as a late entrant compared with core subjects like mathematics and English and has become one of the most popular school subjects. The National Commission for Colleges of Education (NCCE) define the major objectives of teaching economics in Nigerian secondary schools as, to:

- a. prepare students for good citizenship,
- b. provide intellectual training which would create critical thinking,
- c. prepare recipients for vocations,
- d. acquire economic competence (NCCE, 2012).

The Kano state government, through the ministry of education, under the supervision of Kano State Education Resource Department (KERD) introduced Senior Secondary Certificate Qualifying Examination (SSCQE) in 1994 to

1. serve as mock examination,
2. encourage parents' involvement in the education of their children,
3. reduce government expenditure on external examinations for unserious students,
4. prepare schools for the task of re-addressing identified weaknesses in terms of syllabus coverage and training,
5. serve as a yardstick for Government sponsorship (KERD, 2013),

The qualifying examination (SSCQE) was based on the assumption that the subjects and the quality of the test items are similar to those of SSEC. Students with high scores in the qualifying examination will perform well in the final SSCE and secure admission in universities and other higher institutions of learning. This study involves the comparative analyses of SEM of 2015 SSCQE economics items in Kano State.

According to Cohen and Swerdlik (2009), SEM is directly related to the reliability of an item, because, the greater the SEM, the lower the reliability of the item and vice-versa. The three major concepts of SEM stated by Kubiszyn & Borich (2003) are:

- i. Students' test scores: These are the observed scores that students got on test administered.
- ii. Students' true scores: The true scores, in most cases, are unknown due to the fact that no measure can absolutely provide a perfect reflection of the true student's score. But it could be estimated by taking average student's scores over many administrations. Assume you can administer an item many times to a set of students and take the average of each student's scores. That would be the best estimate of that student's true ability in what is being tested.
- iii. Sampling errors: It is difficult, if not impossible, to test students many times while assuming that fatigue, testing effect and other variables remain constant. The best way is to accept each student's test score as the best estimate of the student's true score. It will be noted by the researcher that there are sampling errors in the estimate. The sampling errors are normally distributed with a standard deviation called the Standard Error of Measurement.

The difference between students' actual score and their higher or lower expected score is known as the Standard Error of Measurement. It is the standard deviation of all those scores averaged across persons and test administrations (Obinne, 2011). Hambleton & Swaminathan, (1985) had reported that the Classical Test Theory (CTT) was first conceived in early 1900s when Charles Spearman developed a theory based on the simple analysis that true score plus random 'error' is what makes a test score. The

characteristics of a test item depend on the particular sample of students to which the items were administered. An economics item may, for example, be difficult for secondary school students but possibly easy for university students (Domino & Domino, 2006). They lamented that the major shortcomings of the CTT is circular dependency, i.e. person statistic is item dependent; and item statistics are (examinee) sample dependent.

Embretson and Reise, (2000) opined that unlike the CTT techniques for equating tests and test scores, IRT applies mathematical model generated data in respect of the *item* from diverse and relatively large and diverse samples. The generated data from the items are used to calibrate items in line with one or more test taker variables and derive probability estimates of the total ability needed to attempt teach item in a definite way. The theory essentially, places both persons and items on a common scale. To this end the study adopted the two parameter model among the series of IRT models.

### IRT Assumptions

Wright, (1978) and Cohen (2009) highlighted the following assumptions necessary for analysing data using IRT framework in the fields of education and psychology

**Unidimensionality:** This assumption suggests that a single continuous latent construct is from the set of items called theta. It is denoted with the symbol ( $\theta$ ) and read like the Greek letter, *theta*. The student's theta level is what gives rise to a response to the items in the scale. For instance, a student's response to the question Are you okay? depends on the student's emotional stability. It assumes that one prevailing dimension is reasonable in explaining the underlying structure. However this does not mean that the unidimensionality assumption excludes the possibility that the set of items may have some minor dimensions (subscales).

**Local independence:** This assumption suggests that the probability of a test-taker responding correctly to a particular item must not depend on the previous responses made on the other items. It also implies that the response of a student to an item is not affected by the responses of other students to the same item. This means that the marginal differences in responses to items are functions of differences in the respondents' underlying abilities`

**Monotonicity:** Monotonicity assumption suggests that the probability of selecting an item indicates that higher levels of theta should increase as the underlying level of theta increases. The test taker's response probability increases with the probability of a positive response to the given item.

**Equality of discrimination:** This assumption means that the slope of the item characteristic curve (ICC) of each item is expected to be the same for all the items. If the slopes are not the same, the ICC for two items would eventually cross at some point. That does not mean that all items are supposed to have an equal point-biserial correlation indices with total test score.

Bichi, (2015) uses CTT in evaluating the quality of chemistry achievement multiple choice test items and found that 30% of the items didn't meet the set criteria of item standardization. This suggests the need to revise the items for the next administration. Similarly 60% of the items have been considered as 'good' based on the established standards. The study concluded that the items had significantly positive correlation between difficulty and discrimination indices.

Anigbo, (2015) investigated the psychometric properties of 2009-2011 SSCE economics multiple choice test items and discovered that 2009-2011 SSCE economics multiple choice test items were within the standard psychometric indices. The research recommended that examination bodies in Nigeria should make sure that adequate item analyses are done for all tests used in all national examinations in Nigeria.

Adonu, (2014) studied psychometric properties of practical physics set by WAEC and NECO using the Partial Credit Model. It was aimed at evaluating the SEM, the fit statistics and item difficulty estimates. It also tested the significant difference between SSCE and WAEC psychometric properties in different years. The major findings of the study indicated that the SEM of items of SSCE and WASSCE practical physics in 2011 and 2012 were below 0.18 for all items, which is low. Both SSCE and WASSCE items were valid and thus had adequately confirmed unidimensionality. The item parameter  $b$  for both examinations for the two years studied showed that all the items had difficulty estimates that ranged between -1.53 to +1.94 which show that their difficulty are moderate for all items. All the four different tests that constituted the instrument had very high proportion of their item fit to PCM with all the four parts having 0.92 proportion of fit. The similarity of the psychometric qualities of these examination bodies was shown by the findings of this study. It is therefore recommended that the confidence and recognition accorded to these two examination bodies by the members of the public and educational institutions should continue to be the same.

Obinne, (2011), focused on the analysis of the psychometric properties of the two school certificate examinations organised by SSCE and WAEC in Nigeria. The study compared the SEM of 2000 – 2002 biology examinations using the one-parameter model IRT. Findings from the study revealed significant differences in the SEM of 2000 – 2002 biology examinations organised by SSCE and WAEC. This indicated that the 2000 – 2002 biology examinations organised by NECO had a higher reliability than those of WAEC.

While there are available studies that analysed the psychometric properties of different tests as indicated in the foregoing review, there appear to be little or no empirical evidence backing the psychometric quality of the SSCQE. This calls for the attention of psychometricians to be drawn to the SSCQE. In response to this call, and also, considering public complaint on massive failure and inability of the SSCQE to predict student's performance in national examinations, the present study becomes necessary.

Over the years, experience shows that many students who failed the SSCQE were successful in national examinations like WASSCE and NECO SSCE. KERD record shows that less than 30% of students who sat for the SSCQE got five (5) credits including English and mathematics in 2013, 2014 and 2015. On September 11, 2014, the National Examinations Council (NECO) announced that 52.29 percent of the candidates that sat for June/July 2014 Senior School Certificate Examination (SSCE) passed at credit level in five subjects, including English Language and mathematics ([www.naijaloaded.com.ng](http://www.naijaloaded.com.ng)). Also the 'Premium Times' newspaper reported that the NECO registrar, Prof Abdulrashid Garba announced that 68.56 per cent of candidates got more than five credits, including mathematics and English Language (September 10, 2015 'Premium Times'). This suggests that SSCQE is more difficult than national examinations such as WASSCE and NECO SSCE.

The situation made parents, teachers, students and other educational stakeholders in the state to ask questions. Is the qualifying examination really a test for screening the best students or the minimum students to be sponsored by government? How valid and reliable are the SSCQE test items? What is the reliability indices of the SSCQE items? This study is designed to analyse the Standard Error of Measurement (SEM) of 2015 Kano State School Certificate Qualifying Examination (SSCQE). However, estimating the SEM of SSCQE will have less meaning without comparing it with a particular standard. This is what motivated the researchers to estimate the SEM of SSCE as a national and standard examination and compare the two examinations for decision making. This study therefore set out to estimate the SEM of Economics test items of 2015 SSCQE and 2016 SSCE using the Item Response Theory (IRT) measurement framework and comparing the two to show how standard the SSCQE items are. Specifically the objectives of the study were to:

- i. estimate the SEM indices of the Kano State 2015 SCQE economics test items.
- ii. estimate the SEM indices of the Kano State 2016 SSCE economics test items.
- iii. determine the differences of the standard error of measurement indices between the Kano State 2015 SCQE and 2016 SSCE economics tests.

### **Research Questions**

1. What are the SEM indices of the Kano State 2015 SCQE and SSCE economics test items?
2. What are the SEM indices of the 2016 SSCE economics tests?
3. To what extent do the Standard Error Measurement indices differ between the Kano State 2015 SCQE and 2016 SSCE economics tests?



### **Research Hypotheses**

H<sub>0</sub>1: There is no significant difference between the SEM indices of Kano State 2015 SSCQE and 2016 SSCE economics tests items.

### **Methods**

The study employed ex-post-facto research design. The population of the study consists of the fifteen thousand four hundred and eighty one (15,481) students who sat for the 2015 economics SSCQE and, subsequently, SSCE economics examination in 2016 in Kano State owned senior secondary schools. It consisted of ten thousand nine hundred and fifty (10,950) male students and four thousand and thirty one (4,531) female students. A sample size of 1000 students was selected using multistage sampling technique. In the first instance the percentage of 1,000 sample from 15,481 total population was determined to be (6.646%). This allowed the researcher to randomly select one local government from each zone and one school from each local government selected. The researcher finally drew a sample representing the zone from the school using proportionate stratified sampling.

The research is aimed at analysing the objective items of SSQCE economics 2015 and SSCE 2016 items. The SSQCE economics 2015 has 40 items which cover the SS I and SS II syllabi. The SSCE economics 2016 on the other hand contains 60 items which cover SS I, SS II and SS III economics syllabus. The study will not use any data collection instrument because the research deals with item analysis of existing, constructed and administered items by Kano State Educational Resource Department and National Examinations Council in Kano State senior secondary schools. The validity and reliability of the items have been established by these bodies. However the study intends to estimate the Standard Error of Measurement (reliability indices) of SSQCE economics 2015 and SSCE 2016 items using two-parameter IRT models. Therefore the study does not have any data collection instrument that will warrant the report of the validity and reliability indices of the instrument.

### **Results**

**Research Question One:** What are the SEM indices of the Kano State SSQCE 2015 economics test items?

**Table 1: SEM of the Kano State SSQCE 2015 Economics Test Items Based on Two - parameter Logistic (2PL) Model.**

ITEM	SSQCE SEM INDICES	ITEM	SSQCE SEM INDICES
1	.07	21	.04
2	.07	22	.17
3	.16	23	.09
4	115.56	24	.05
5	.15	25	.05
6	.12	26	.39
7	1.49	27	.06
8	.05	28	.16
9	.23	29	.09
10	.10	30	.07
11	.12	31	.06
12	.18	32	.53
13	.23	33	4.55
14	.37	34	.08
15	.10	35	2.09
16	.06	36	.46
17	.05	37	0.07
18	.08	38	3.57
19	.14	39	.12
20	.08	40	.07

Table 1 presents the standard errors of measurement of the Kano State SSQCE 2015 economics test items based on two-parameter logistic (2PL) model. Table 1 shows that 85% of the items were below the standard error of measurement of 0.5 while only items 4, 7, 32, 35, 33, 35, and 38 were above 0.5 level of the accepted standard errors of measurement which accounted for only 15% of the SSQCE 2015 economics items. This suggests that the reliability of the Kano State SSQCE 2015 economics test items, based on two-parameter logistic (2PL) model, is high. It indicates consistency in measuring student ability in economics.

**Research Question Two:** What are the SEM indices of the 2016 SSCE economics tests?



**Table 2 : SEM of the 2016 SSCE Economics TestItems Based on Two-parameter Logistic (2PL) Model.**

ITEM	SE	ITEM	SE	ITEM	SE
1	.11	21	.26	41	.31
2	.17	22	.10	42	.19
3	.12	23	.15	43	.12
4	.11	24	.08	44	.29
5	.09	25	.10	45	.25
6	.11	26	.19	46	.08
7	.11	27	.15	47	.18
8	.17	28	.21	48	.14
9	.15	29	.22	49	.17
10	.13	30	.30	50	.13
11	.11	31	.15	51	.09
12	.10	32	.21	52	.09
13	.17	33	.18	53	.10
14	.13	34	.15	54	.11
15	.18	35	.12	55	.13
16	.08	36	.20	56	.10
17	.08	37	.26	57	.10
18	.10	38	.14	58	.09
19	.14	39	.17	59	.08
20	.09	40	.22	60	.09

From Table 2 shows the SEM of the SSCE 2016 economics test items using two-parameter the IRT model. It could be seen from the table that 81.67% of the items were between the accepted levels of the standard error of measurement. Only 18.33 were above the standard error of measurement of 0.5. These items are 2, 5, 7, 17, 46, 49, 52, 57, 68, 59, and 60. This suggests that 81.67% of the 2016 SSCE economics items has low standard error of measurement. Since the range is below S.E of 0.5, the reliability of the test is high.

**Research Question Three:** To what extent do the Standard Error of Measurement indices differ between the Kano State 2015 SCQE and 2016 SSCE economics tests?

**Table 3 : Descriptive Statistics Showing the Standard Errors of Measurement of 2016 SSCE and 2015 SSCQE Economics Items Differences Based on Two-parameter Logistic (2PL) Model.**

Test	No of items	Mean	Stddev
NECO SEM	60	1.15	0.36
SSCQE SEM	40	1.13	0.33

**Table 3 presents the descriptive statistics of the** standard errors of measurement of 2016 SSCE and 2015 SSCQE economics items. It can be seen from the table that there were 60 items in 2016 SSCE economics test and 40 items from 2015 SSCQE economics test. The mean of the both 2016 SSCE and 2015 SSCQE economics items were at the same range of 1.15 and 1.13 respectively. Meanwhile the standard deviation of 2016 SSCE economics items stood at 0.36; that of 2015 SSCQE economics items was 0.33.

**Research Hypothesis H<sub>01</sub>:** There is no significant difference between the SEM indices of KanoState 2015 SCQE and 2016 SSCE economics tests items.

**Table 4 : Independent t-test Analysis Showing Differences in SME Indices of 2015 SCQE and 2016 NECO Economics Tests Items**

	Mean	Std. Deviation	Std. Error Mean	T	Df	p-value
S.E NECO - S.E-SSCQE	-2.750	18.537	2.931	-.938	39	.354

The result on Table 4 shows that the t-value obtained was – 0.938 with associated probability value of 0.354 which is greater than the 0.05 level of significance. Therefore, the null hypothesis which states that there is no significant difference in the SEM of Kano State 2015 SCQE and 2016 SSCE economics test items was retained. There was no significant difference in the SEM of Kano State 2015 SCQE and 2016 SSCE economics test items.

**Discussion**

The results presented in tables 1, 2, 3 and 4 reveal that based on the two-parameter logistic (2PL) model, standard errors of measurement of the 2015 SSCQE and 2016 SSCE economics test items were found with higher reliability levels and there was no significant difference between the SEM of Kano State 2015 SCQE and 2016 SSCE economics test items. Table 1 shows that 85% of the 2015 SSCQE items were found consistent in measuring the students' ability in economics with SME less than 0.5. The findings in Table 2 reveal that 81.67% of the 2016 SSCE economics items had low SMEs. These findings are consistent with findings of Obinne (2011), Bichi (2015) and Adonu (2014) that SSCQE and SSCE had a higher reliability. The findings also reveal that 2015 SSCQE and 2016 SSCE economics items were found to be of the same quality; therefore success in SSCQE may determine success in SSCE. Therefore the massive failure in SSCQE over the years was not caused by the quality of the SSCQE items, but by other variables which are not related with the quality of the test items. There is no item superiority between items generated by KERD and those of by NECO. It is, therefore, recommended that the confidence and recognition accorded these two examination bodies by the members of the public and educational institutions be maintained.

### **Conclusion**

The findings of this study justify the conclusion that the massive failure in Kano State Certificate Qualifying Examinations over the years was not caused by the quality of the SSCQE items, rather other variables which are not related with reliability indices of the test items are to blame. It also be concluded that there is no superiority in terms of item quality between the items generated by KERD and those by NECO. Performance in SSCQE, therefore, may determine performance in SSCE.

### **Summary**

The study analysed the 2015 SSCQE and 2016 SSCE economics items in which the SEM indices were estimated. The study employed an expo facto research design in analysing 40 economics items from 2015 SSCQE and 60 economics items from 2016 SSCE multiple choice test using two-parameter IRT model. The findings of the study revealed that 2015 SCQE and 2016 SSCE economics test items were reliable and there was no significant difference between the SEMs of Kano State 2015 SCQE and 2016 SSCE economics test items.

### **Recommendations**

From the findings of this study, it is recommended that:

- i. All examination bodies in Nigeria should adopt IRT analysis in conformity with best global practices.
- ii. KERD should employ the services of experts in test and measurement to update the skills of its staff on the latest IRT models.
- iii. There should be a collaboration between KERD, NGO's and members of the media team to enlighten members of the public that SSCQE is a standard examination meant to groom students for national examinations.
- iv. KERD should publish its examinations validation process and make it available in the state libraries and on its website for public consumption.

### **References**

- Adonu, I. I. (2014), Psychometric Analysis of WAEC and NECO Practical Physics Tests using Partial Credit Model. Unpublished PhD dissertation submitted to the Department of Science Education University of Nigeria, Nsukka
- Anigbo, L. C. (2015), Item Analysis of National Examination Council Senior School Certificate Examination Economics Objective Tests. International Journal of Innovative Education Research, 3(4), 23-30

- Bichi, A. A. (2015), 'Item Analysis using a Derived Science Achievement Test Data' International Journal of Science and Research (IJSR), 4(5), ISSN (Online): 2319-7064. [www.ijsr.net](http://www.ijsr.net)
- Cohen, R. J. & Swerdlik, M. (2009), Psychological Testing and Assessment: An Introduction to Tests and Measurement (7th Ed.) USA: McGraw-Hill X
- Domino, G. & Domino, M. L. (2006), Psychological Testing An Introduction (2<sup>nd</sup> ed) Cambridge, UK: Cambridge University Press
- Embretson, S. E. & Reise, S. P. (2000), Item Response Theory for Psychologists. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hambleton, R. K. & Swaminathan, H. (1985), Item Response Theory. Principles and Applications. Boston: Kluwer.
- Ibrahim, Y. A. (2012), "Controversy over Kano exam mass failure" in DAILY TRUST, Thursday January 27. P.45.
- Kubiszyn, T. & Borich G. (2003), Educational Testing and measurement: classroom Application and practice. (7<sup>th</sup> ed.) New York: John Wiley & Sons. Inc.
- NCCE (2012), Arts and social sciences curriculum. Retrieved on 13/11/2016 from [www.ncce.gov.ng](http://www.ncce.gov.ng).
- NECO (2014), Result Released NL Talk Talk, September 11, 2014 retrieved from <https://www.naijaloaded.com.ng/other/neco-2014-junejuly-result-released> on 21/07/2017
- NECO Result (2015), Premium Times, September 10, retrieved from <https://www.premiumtimesng.com/news/headlines/189794-neco-releases-2015-exams-results-declares-higher-pass-rate.html> On 21/07/2017
- Obinne, A. D. E. (2011), 'Psychometric Analysis of Two Major Examination in Nigeria: Standard Error of Measurement' International Journal of Educational Sciences, 3(2), 137-144.
- Suruchi & Rana, S. R. (2014), 'Test Item Analysis and Relationship Between Difficulty Level and Discrimination Index of Test Items in an Achievement Test in Biology' Paripex –Indian Journal of Research, 3(6), 56-58,
- WASSCE (2016), WAESSCE result. Retrieved from [www.wasscedirect.org](http://www.wasscedirect.org) on 21/07/2017
- Wright, D. B. (1978), 'The Rasch Model for Test Construction and Person Measurement Prepared for Fifth Annual Conference and Exhibition on Measurement and Evaluation. Office of the Los Angeles County Superintendent of Schools Division of Program Evaluation, Research and Pupil Services. Retrieved from [www.rasch.org](http://www.rasch.org) on 26/10/2015