

**DIFFICULTY ITEM PARAMETER DRIFT AND ITS THREATS TO THE
VALIDITY OF ECONOMICS MULTIPLE CHOICE TEST ITEMS OF
NATIONAL BUSINESS CERTIFICATE EXAMINATION IN NORTH
CENTRAL NIGERIA**

BY

¹IORNIENGE, Moses Tarnum

²Prof. A.D.E. Obinne

³Dr. M.O. Aduloju

&

⁴Dr. C.I. Agi

¹Department of Vocational and Technical Education, Faculty of Education, Benue State University, Makurdi Nigeria.

^{2, 3 & 4} Department of Educational Foundations, College of Agricultural and Science Education, Federal University of Agriculture, Makurdi Nigeria.

Abstract

The study investigated the item parameter of difficulty of Economics multiple-choice test items of National Business Certificate Examination in Nigeria from 2016-2018. One research question and one hypothesis guided the study. The *ex-post facto* research design was adopted. The population of the study was all the 58,957 SS3 students in the North-Central States of Nigeria and the Federal Capital Territory, Abuja, who registered and sat for Economics of National Business Certificate Examinations (NBCE) organized by the National Business and Technical Examinations Board (NABTEB) in Nigeria for the years 2016, 2017 and 2018. A sample of 31,775 students was selected in multi-stages. The instrument used for data collection was a Pro-Forma named Economics Multiple-Choice Test Items Matrix Response Proforma (ECOMUCTIMRP) with 53 columns and 31,775 rows which corresponds to the sampled students. The research question was answered using Item parameter of difficulty, computed using jMetrik software. One Way Analysis of Variance (ANOVA) was used to test the hypothesis at 0.05 level of confidence. The study was hinged on Item Response Theory (IRT). Findings of the study revealed that a greater percentage (54%, 58% & 44% for 2016, 2017 & 2018 respectively) of the items of Economics Multiple Choice Test were of acceptable difficulty index except for 2018 where most of the items were either too easy or hard. Again, the study revealed that there was drift in the difficulty parameter, between 2016 and 2018 recording the highest mean drift of 0.0697 followed by that between 2016 and 2017 which was 0.0518 and the least mean drift was between the years 2017 and 2018 which was 0.0179. The study also established that, though there was evidence of drift in all the three study years, this was not statistically significant to threaten the validity of the test within any of the study years. Based on the findings of this study, it was concluded that the difficulty item parameter of Economics

multiple-choice test of National Business Certificate Examination have been consistent in terms of difficulty for the purpose of scoring, grading and interpretation of performance over the years of study. The difficulty item parameter does not constitute a threat to the validity of the examination.

Key words: Difficulty, Item parameter drift, Validity, Threat.

Introduction

The Nigerian government is continually making frantic efforts at ensuring that high academic standards are maintained in her education system. Psychometricians are saddled with the challenge of ensuring accurate measurement of educational growth and high examination standards that could be comparable globally.

In Nigeria, there are two different levels of post-primary certificate examinations that students are expected to write for the purpose of certification. These are Basic Education Certificate Examination (BECE) which is administered after the completion of three years of Basic Education, and Senior Secondary Certificate Examination (SSCE) which is administered at the end of senior secondary or technical schools. The examination bodies vested with the responsibility of administering examinations for senior secondary/technical certification are the West African Examination Council (WAEC), National Examinations Council (NECO) and National Business and Technical Examinations Board (NABTEB). These examination bodies are mandated to assess, evaluate and certify students who are exposed to the curriculum of Senior Secondary and Technical Schools. Students are examined every year for the purpose of certification. The quality of the certificates issued to students who are examined by these examination bodies are expected to be comparable in worth and value over the years. However, parents, teachers, employers of labour and other stakeholders complain of the falling quality of education. Corroborating this statement, Aworanti, Onuka and Taiwo (2013) declare that, since examinations have been identified as a tool for assessing educational achievement, it is surrounded with the danger of examination fraud and concomitant credibility problems. There is every need therefore, to put a mechanism in place to ascertain, with empirical evidence that there has been consistency in the quality of the certificates issued over the years by these examination bodies, with particular reference to National Business and Technical Examinations Board (NABTEB).

The National Business and Technical Examinations Board (NABTEB) was established under decree 70 of 1993 and mandated to conduct examinations leading to award of the National Business Certificate (NBC) and Advanced National Business Certificate (ANBC). The Board is again mandated to conduct research, publish statistics and other information in order to develop appropriate examinations, tests and syllabi in Technical and Business Studies. These examinations are standardized examinations that are written by students in Nigeria who wish to pursue Vocational and Technical Education career. Students are examined in many academic and vocational subjects, including Economics, which is a requirement especially for students who want to read Business and Management courses such as Accounting, Marketing, and

Business Management, Business Education, Economics, Banking and Finance. The Economics examination is divided into two papers, which are made up of Essay type questions and multiple-choice objective test questions respectively. However, the focus of this study is on the multiple-choice objective test items. The reason for limiting the study to multiple-choice objective test items is because of its wide adaptability to different content areas and the ease to use it in measuring factual recall, as well as complex skills, exercise of judgement and critical thinking. It is imperative that decisions on the performance of students arising from their scores and grades in this subject area should have empirical evidence emanating from a measurement framework that guarantees the minimization of error in test score interpretations.

One way of determining whether the quality of items in an examination is consistent is to investigate the item parameters of those items that constitute the test. The invariant property of the items, which is a concept that explains how stable and consistent the item parameters of test items are over several administrations, is one way of checking consistency of quality of examinations. Failure to fulfill this assumption leads to what is called Item Parameter Drift (IPD) which might constitute a threat to the validity of results that are obtained from the administration of the test items on the basis of which the interpretation of the scores and grades were based.

The phenomenon in which the parameter values of the same test items change systematically over multiple testing occasions is referred to as Item Parameter Drift (IPD). It is described by Orheruata, Omorogiuwa and Osunde (2017) as “the shift of item parameter estimates from the acceptable theoretical scale”. Item Parameter Drift in measurement context holds when there is a violation in the stability of the parameter scale. This situation threatens test results. It can have impact on examinees’ classification accuracy and could complicate the comparison being made of examinees’ performance over time, if unchecked. Han in Orheruata *et al* (2017) posited that, the changes in the interaction between a test item and an examinee essentially results in changes in the item parameter values from the initially calibrated parameter estimates in the item pool.

Item Parameter Drift is sometimes regarded as a special case of Differential Item Functioning (DIF). Differential Item Functioning occurs when examinees from different groups (gender or location, as the case may be) show differing probabilities of success on endorsing an item after matching on the underlying ability that the item is intended to measure. In studies of DIF, a focal group is defined and often that group is thought of as being potentially disadvantaged by the assessment (example, female, in case of gender), while the reference group is the set of examinees that the focal group is being compared to (example, male, in the case of gender). However, in IPD studies, the reference and focal groups can be defined as ‘First Administration’ and ‘Second Administration’ examinees, respectively where the first administration may be delivered one year and the second in another. Parameter estimates for the same items that vary across test administrations are therefore said to possess IPD, and thus those items function differentially between testing occasions. The difference between DIF and IPD concepts is in terms of number of administrations. Differential Item Functioning is concerned with single test administration, while IPD is concerned with several administrations of the test.

There are a variety of sources that could cause the parameters of item difficulty: discrimination, pseudo guessing and carelessness to drift upwards or downwards when several administrations are made. Whatever the cause of IPD, several studies have shown that its presence can lead to biased estimates of ability and ultimately improper classification of examinees (Han, 2008; Wells, Subkoviak, & Serlin, 2002). It is therefore important to investigate the presence of IPD in the Economics multiple-choice test items of National Business Certificate Examination conducted by NABTEB to ensure the examination is devoid of these threats to its validity.

Validity is defined by Ogoamaka, Onah and Okwa (2017) as the extent to which a test is free from both chance and systematic errors. On the one hand, a test is free from chance errors if it measures whatever it measures consistently without any interference from unpredicted sources. On the other hand, a test that is free from systemic errors is an indication that it is relevant to the objective it was designed to serve. In scientific inquiry, validity of statements refers to the degree to which there is empirical evidence to support the adequacy and appropriateness of these statements. Validity is therefore, how accurate a method measures what it claims to measure and the results closely corresponds to real-world values (Middleton, 2019).

Basically, validity is always concerned with the specific use to be made of the results. It is the extent to which results of an evaluation procedure serve the particular uses for which they are intended. It is concerned with the result of test instrument and not with the instrument itself. The validity of a measurement instrument depends on the quality of the items of the instrument, as well as the specific use to be made of the test results. Again, Anikweze (2014) posits that validity of a test is when it measures accurately and consistently what it is designed to measure and nothing else. Enunwah and Agi (2019) summarize the definition of validity as the trustworthiness of scoring, meaning and its interpretation. Validity of a test is therefore, the degree or extent to which a test measures what it claims to measure. The basic concern is with the usefulness of test results which is usually achieved through a process.

Validation, according to Gruijter and Leo, (2008) is the process through which the validity of the proposed interpretation of scores is investigated. What this implies is that, the process of validation amounts to collecting empirical evidence to provide stable and generally accepted theoretical basis for the interpretations of test scores and other modes of assessment. When the validity of test scores is threatened through successive administrations, leading to drift in item parameters, it is an obvious indication that harm might be caused through the decisions arising from the interpretations of such scores.

Since the commencement of National Business Certificate examination (NBCE), several editions of the Economics multiple-choice tests have been administered to students over the years with the aim of quantifying the extent of mastery of the content of the subject matter by these students in the subject. Conducting Item Parameter drift check is therefore essential for the examination that continually produces new editions of tests and for which the expectation is that, scores from these editions should have the same meaning over time (Dorans, Moses & Eignor, 2010).

It is in line with this professional best practice that the researcher investigated the difficulty item parameter estimates of NABTEB's National Business Certificate Examination (NBCE) Economics Multiple Choice Test Items to ascertain whether there has been the presence of drift in the difficulty parameter of the items; and to what magnitude given that much evidence is required to establish empirically whether IPD might be a threat to the validity of the Economics Multiple-Choice Test Items of National Business Certificate Examination over the years. Moreover, evidences are needed to establish that the inferences made are based on appropriate test results for the purpose of certification.

Research Question

The research was guided by this research question:

- (i) What is the difficulty index of Economics Multiple-Choice Test Items of National Business Certificate Examination from 2016 - 2018?

Statement of Hypothesis

The following null hypothesis was formulated and tested at 0.05 level of significance.

- (i) The Drift in mean Item Difficulty Parameter of Economics Multiple-Choice Test Items of National Business Certificate Examination from 2016 – 2018 do not differ significantly to threaten its validity.

Methodology

The *ex-post facto* research design was adopted. The population of the study was all the 58,957 SS3 students in the North-Central States of Nigeria and the Federal Capital Territory, Abuja who registered and sat for Economics of National Business Certificate Examinations (NBCE) organized by the National Business and Technical Examinations Board (NABTEB) in Nigeria for the years 2016, 2017 and 2018. A sample of 31,775 students were selected in multi-stages. The instrument used for data collection was a Pro-Forma named Economics Multiple-Choice Test Items Matrix Response Proforma (ECOMUCTIMRP) with 53 columns and 31,775 rows which corresponds to the sampled students. The research question was answered using Item parameter of difficulty, computed using jMetrik software. One Way Analysis of Variance (ANOVA) was used to test the hypothesis at 0.05 level of confidence.

The matrix data responses of the students to each of the 50 multiple-choice items for the three years received from NABTEB were formatted into numeric data and saved into excel as Coma Delimited format (CSV) files for the respective years. They were then imported into jMetrik (version 3.1) for item parameter calibration. jMetrik is a windows application software computer program for implementing classical and modern psychometric methods. It is designed to score and perform estimation of IRT parameters with ease (Meyer, 2014).

The data were analysed using descriptive statistics of mean, percentage and common odds ratio in answering the research question. One Way Analysis of Variance (ANOVA) was used to test for significant drift at 0.05 level of significance. The Cochran-Mantel Haenszel Statistic (CMH) and practical significance were used in testing the hypothesis on DIF. According to Abegunde and Adeyemo (2019), the conventional guide for interpretation of item difficulty index is as follows:

- (i) Difficulty
- | | | |
|------------|---------------|----------------|
| 75% - 100% | (0.75 - 1.00) | Easy items |
| 26% - 74% | (0.26 - 0.74) | Moderate items |
| 0% - 25% | (0.00 - 0.25) | Hard items |

RESULTS AND DISCUSSION

The results obtained from the analysis of data collected based on the research question asked and in line with the hypothesis formulated are hereby presented.

Research Question One: What is the difficulty index of Economics Multiple Choice Test Items of National Business Certificate Examination from 2016 - 2018?

To answer Research Question one, Tables 1 and 2 were used. While Table 1 showed the difficulty parameter of the items for each of the three years, Table 2 showed the percentage summary of the difficulty parameter.

Table 1: Difficulty Parameters of Economics Multiple-Choice Test Items from 2016-2018 Examination Years

ITEM	Difficulty Parameter for Each of the examination 2016-2018 Years					
	2016	REMARKS	2017	REMARKS	2018	REMARKS
Q1	0.8564**	Very Easy	0.6624	Moderate	0.6454	Moderate
Q2	0.8931**	Very Easy	0.9197**	Very Easy	0.6153	Moderate
Q3	0.927**	Very Easy	0.911**	Very Easy	0.9671**	Very Easy
Q4	0.6012	Moderate	0.5846	Moderate	0.832**	Very Easy
Q5	0.8305**	Very Easy	0.872**	Very Easy	0.8808**	Very Easy
Q6	0.6466	Moderate	0.2577**	Hard	0.7791**	Very Easy
Q7	0.0371**	Hard	0.8782**	Very Easy	0.9249**	Very Easy
Q8	0.9283**	Very Easy	0.3646	Moderate	0.6098	Moderate
Q9	0.9019**	Very Easy	0.3657	Moderate	0.8444**	Very Easy
Q10	0.6788	Moderate	0.3841	Moderate	0.8115**	Very Easy
Q11	0.1466**	Hard	0.7793**	Very Easy	0.6758	Moderate
Q12	0.0712**	Hard	0.3531	Moderate	0.531	Moderate
Q13	0.4754	Moderate	0.0995**	Hard	0.0184**	Hard
Q14	0.8774**	Very Easy	0.8427**	Very Easy	0.4446	Moderate
Q15	0.3721	Moderate	0.4329	Moderate	0.9487**	Very Easy
Q16	0.7134	Moderate	0.9162**	Very Easy	0.6675	Moderate
Q17	0.7395	Moderate	0.509	Moderate	0.8446**	Very Easy
Q18	0.4505	Moderate	0.7397	Moderate	0.4292	Moderate
Q19	0.7104	Moderate	0.5408	Moderate	0.9694**	Very Easy
Q20	0.7662**	Very Easy	0.7032	Moderate	0.8256**	Very Easy
Q21	0.1964**	Hard	0.5614	Moderate	0.7353	Moderate
Q22	0.8231**	Very Easy	0.1155**	Hard	0.1421**	Hard
Q23	0.9266**	Very Easy	0.5741	Moderate	0.8457**	Very Easy

Q24	0.5941	Moderate	0.6332	Moderate	0.6072	Moderate
Q25	0.3956	Moderate	0.9136**	Very Easy	0.5602	Moderate
Q26	0.7397	Moderate	0.3399	Moderate	0.6034	Moderate
Q27	0.5969	Moderate	0.4687	Moderate	0.2422**	Hard
Q28	0.8078**	Very Easy	0.5516	Moderate	0.5606	Moderate
Q29	0.934**	Very Easy	0.8592**	Very Easy	0.073**	Hard
Q30	0.8978**	Very Easy	0.5154	Moderate	0.8252**	Very Easy
Q31	0.6885	Moderate	0.5958	Moderate	0.8626**	Very Easy
Q32	0.2045**	Hard	0.8755**	Very Easy	0.6771	Moderate
Q33	0.931**	Very Easy	0.7273	Moderate	0.9093**	Very Easy
Q34	0.8564**	Very Easy	0.6759	Moderate	0.9146**	Very Easy
Q35	0.7353	Moderate	0.0639	Moderate	0.9227**	Very Easy
Q36	0.7157	Moderate	0.9073**	Very Easy	0.3146	Moderate
Q37	0.8529**	Very Easy	0.6253	Moderate	0.6592	Moderate
Q38	0.3036	Moderate	0.7809**	Very Easy	0.4877	Moderate
Q39	0.7644**	Very Easy	0.7746**	Very Easy	0.4198	Moderate
Q40	0.5737	Moderate	0.7415	Moderate	0.41	Moderate
Q41	0.6183	Moderate	0.682	Moderate	0.94**	Very Easy
Q42	0.5878	Moderate	0.4324	Moderate	0.4245	Moderate
Q43	0.7464	Moderate	0.7406	Moderate	0.0805**	Hard
Q44	0.6106	Moderate	0.9017**	Very Easy	0.6757	Moderate
Q45	0.5657	Moderate	0.8378**	Very Easy	0.7775**	Very Easy
Q46	0.687	Moderate	0.879**	Very Easy	0.4716	Moderate
Q47	0.4412	Moderate	0.6511	Moderate	0.111**	Hard
Q48	0.5411	Moderate	0.9024**	Very Easy	0.8079**	Very Easy
Q49	0.2784	Moderate	0.608	Moderate	0.879**	Very Easy
Q50	0.1072**	Hard	0.8795**	Very Easy	0.7824**	Very Easy

Table 1 presents the item difficulty of the 50 items for each of the three study years from 2016 to 2018. The items are classified depending on the values of the difficulty parameter as: Very Easy, Moderate and Hard. The criteria used for the classification was based on the conventional guidelines stipulated by Abegunde and Adeyemo (2019), Meryer (2014) and Anikweze (2014); Items with difficulty index 0.75 to 1.00 are very easy, those with difficulty index range of 0.26 to 0.74 were considered as moderate, while items whose difficulty index was between 0.00 to 0.25 were considered as hard items.

In 2016, seventeen (17) items (items Q1, Q2, Q3, Q5, Q8, Q9, Q14, Q20, Q22, Q23, Q28, Q29, Q30, Q33, Q34, Q37, and Q39) were very easy, twenty-seven (27) items (Q4, Q6, Q10, Q13, Q15, Q16, Q17, Q18, Q19, Q25, Q26, Q27, Q31, Q35, Q36, Q38, Q40, Q41, Q42, Q43, Q44, Q45, Q46, Q47, Q48, and Q49) were of moderate difficulty; while six (6) items (Q7, Q11, Q12, Q21, Q32 and Q50) were very hard.

In 2017, eighteen (18) items (Q2, Q3, Q5, Q7, Q11, Q14, Q16, Q25, Q29, Q32, Q36, Q38, Q39, Q44, Q45, Q46, Q48 and Q50) were very easy; twenty-nine items (Q1, Q4, Q8, Q9, Q10, Q12, Q15, Q17, Q18, Q19, Q20, Q21, Q23, Q24, Q26, Q27, Q28, Q30, Q31, Q33, Q34, Q35, Q37, Q40, Q41, Q42, Q43, Q47 and Q49) were moderate in difficulty; while three items (Q6, Q13 and Q22) were hard.

In 2018, twenty-two (22) items (Q3, Q4, Q5, Q6, Q7, Q9, Q10, Q15, Q17, Q19, Q20, Q23, Q30, Q31, Q33, Q34, Q35, Q41, Q45, Q48, Q49 and Q50) were very easy; twenty-two (22) items (Q1, Q2, Q8, Q11, Q12, Q14, Q16, Q18, Q21, Q24, Q25, Q26, Q28, Q32, Q36, Q37, Q38, Q39, Q40, Q42, Q44 and Q46) were of moderate difficulty; while six (6) items (Q13, Q22, Q27, Q29, Q43 and Q47) were hard.

Table 2: Summary of Item Difficulty for Economics Multiple-Choice Test from the 2016-2018 examination years

YEAR	HARD		MODERATE		VERY EASY		TOTAL	
	N	%	N	%	N	%	N	%
2016	6	12	27	54	17	34	50	100
2017	3	6	29	58	18	36	50	100
2018	6	12	22	44	22	44	50	100

Table 2 shows the percentage distribution of the difficulty parameter of Economics Multiple-Choice Test Items of National Business Certificate Examination during the 2016-2018. The results indicate that, in 2016, 12% of the items were hard, 54% of the items were of acceptable moderate difficulty; while 34% of the test items were very easy for the students. Again, the Table shows that, in 2017, 6% of the items were hard, 58% of the items were found moderately difficult; while 36% of the items were very easy. Also in 2018, 12% of the items were hard, 44% were moderate; while 44% of the items were very easy.

Research Question Two: How does the mean difficulty index of Economics Multiple-Choice Test Items of National Business Certificate Examination exhibit drift from 2016 – 2018 to threaten its validity?

Research Question two is answered with data in Table 5 which shows the mean differences in the difficulty parameter of Economics Multiple-Choice Test Items from 2016 – 2018.

Table 3: Mean Difference in the Difficulty Parameter of Economics Multiple choice Test from 2016-2018 examination years

Year	Mean	Std. Deviation	\bar{x} difference
2016	0.2080	0.17603	0.0518
2017	0.1562	0.16649	
2016	0.2080	0.17603	0.0697
2018	0.1383	0.15087	
2017	0.1562	0.16649	0.0179
2018	0.1383	0.15087	

From Table 5, the result of the analysis shows that, Economics Multiple-Choice Test Items of 2016 had a mean difficulty index of 0.2080; that of 2017 was 0.1562 and 2018 had a mean difficulty index of 0.1383. This shows that, the mean difference of the difficulty index of Economics Multiple-Choice Test Items for the examination years of 2016 and 2017 was 0.0518, that of 2016 and 2018 was 0.0697 and the mean difference between 2017 and 2018 was 0.0179. This indicates that, the difficulty parameter drifted within the study period.

Hypothesis: The drift in mean item difficulty parameter of Economics Multiple-Choice Test Items of National Business Certificate Examination from 2016 – 2018 is not significant to threaten its validity.

The hypothesis was tested using Analysis of Variance (ANOVA).

Table 4: ANOVA Result of Drift in Difficulty Index of Economics Multiple-Choice Test of National Business Certificate Examination of NABTEB between 2016-2018 examination years

Source of Variation	Sum of Squares	Df	Mean Square	F	P-Value
Between Groups	5.005	2	2.503	.005	.995
Within Groups	79244.806	147	539.080		
Total	79249.812	149			

$$F_{(2,147)} = 0.005; p = 0.995; p > 0.05$$

Table 11 shows that, $F(2,147) = 0.005$, $p = 0.995$ ($p > 0.05$) for the three examination years. Since $p > 0.05$, the test statistics is not significant. This implies that, the null hypothesis is not rejected. Therefore, the null hypothesis, which states that the drift in mean item difficulty

parameter of Economics Multiple-Choice Test Items of National Business Certificate Examination from 2016 – 2018 is not significant to threaten its validity is not rejected.

Discussion of Findings

Based on the findings of the study on Item Parameter Drift and its Threats to the Validity of Economics Multiple-Choice Test Items of National Business Certificate Examination conducted by NABTEB in North Central Nigeria, the following discussions were made:

Findings from research question one shows that, 54% of the items were generally of moderate difficulty in 2016. The difficulty parameter values for the examination year ranged from 0.0712 to 0.9340. Six items (12%) were hard, seventeen items (34%) were very easy. None of the items showed negative difficulty index which indicates that the items had no misinformation to mislead the examinees in their responses. This finding is consistent with the submission of Okozor, Onah and Ukah (2017) as they aver that, IRT is a tool for accurate use for development of instrument with high validity.

The Difficulty parameter for 2017 indicated that, 58% of the items were of acceptable difficulty. The values of the Difficulty parameter for the year ranged from 0.0639 to 0.9162. Details of the results indicated that, twenty-nine items (58%) were of moderate difficulty, three (3) items (6%) were very hard while eighteen (18) items, representing 36% were very easy. None of the 2017 items had negative difficulty values, which strongly indicated that they were not mis-keyed or with other major flaw that needed review. Based on the recommendations of Hambleton and Swaminathan (2013), the items do not need to be reviewed or discarded in future examinations.

Again, findings from research question one further revealed that, difficulty parameters of the 2018 Economics Multiple-Choice Test items ranged from 0.0184 to 0.9694. Details of the difficulty index indicated that, 22 items (44%) were of moderate difficulty, 6 items (12%) were very difficult, while 22 items (44%) were very easy. None of the items exhibited negative difficulty. It is therefore remarked that the difficulty parameter for the three examination years were of acceptable difficulty levels. This finding was consistent with a study carried out by Ethe and Odjegba (2019) as they aver that the items satisfied IRT statistical conditions.

The findings of hypothesis one showed that, though there were differences in the mean difficulty of the items in the three years, these differences were not statistically significant. Thus, the null hypothesis which said the drift in item difficulty of Economics multiple-choice test items of National Business Certificate Examination from 2016 – 2018 was not significant to threaten its validity was not rejected. This finding was consistent with the findings of Wells, Subkoviak and Serlin (2002) that, it is only when there is 20% IPD that could have a significant impact on test scores. The differences here were not up to 10% in the years of study.

Conclusion and Recommendations

Based on the findings of this study, it was concluded that, the item Parameters of Economics Multiple-Choice Test of National Business Certificate Examination has been consistent in terms

of Difficulty and Discrimination for the purpose of scoring, grading and interpretation of performance over the years of study. The item parameters do not constitute a threat to the validity of the examination. Based on the findings of this study, the following recommendations were made:

1. The items that had negative parameter values was a clear indication that they had some misinformation and were favouring the weak examinees at the expense of the intelligent ones. These items should be reviewed or discarded in subsequent usage.
2. Content experts should be involved in the review of the flagged items to investigate whether the negative parameter values obtained were due to misinformation, thereby placing the weak students at an advantage over the intelligent ones.
3. The examination body should make use of competent psychometricians to calibrate their items properly to ensure credibility of item parameters to improve the quality of the examination.

References

- Abedalaziz, N. (2010). A gender- related differential item functioning of mathematics test items. *Time Taylor Academic Journals: The International Journal of Educational and Psychological Assessment*. 5, 125–132
- Abegunde, A.T., & Adeyemo, E.O. (2019). Psychometric analysis of English and Mathematics items of the senior school qualifying examination. *Journal of Evaluation*. 4(1), 37–44
- Akande, S. A., & Azike, J. A. (2002). *Fundamental principles of Economics for senior secondary schools*. Evans Brothers (Nigeria Publishers) Limited, Ibadan-Nigeria.
- Aliyu, R.T., & Akuche, U. E. (2019). Assessment of differential item functioning in Mathematics multiple choice test in Basic education certificate examination in Oyo State. *Journal of Evaluation*. 4(1),125–139.
- Anikweze, C.M. (2014). *Measurement and Evaluation for Teacher Education*. (3rd Edition). Constellation (NIG) Publishers. Ibadan.
- Ary, D., Jacobs, L.C., & Sorensen, C. (2010). *Introduction to Research in Education*. (8th Edition). USA: Belmont.
- Aworanti, A.O., Onuka, A.O.U. & Taiwo, M.B. (2013). Securing examination integrity: The NABTEB approach. *Nigerian Journal of Educational Research and Evaluation*. 12, 125–132
- Aworanti, O.A. (2014). Security breach: A threat to the quality of examination and assessment. *African Journal of Theory and Practice of Educational Assessment (AJTPEA)*. Yaounde Cameroun. 1, 19–33
- Aworanti, A. O. (2015). Predictive validity of school-based assessment on NBC/NTC candidates' performance in English language and Basic Electricity. *Nigerian Journal of Educational Research and Evaluation*. (14) 2, 86–97.

- Cohen, J., Jiang, T., & Yu, P. (2008). *Efficient Linking and Equating of Rasch Scales*. American Institutes for Research (Unpublished manuscript).
- Crocker, L. (2003). Teaching for the test: Validity, fairness and moral action. *Educational Measurement: Issues and Practice*. 22(3), 5–11. Retrieved from <https://doi.org/10.1111/j.1745-3992.2003.tb00132.x>
- Crooks, T.J., Kane, M.T., & Cohen, A.S. (2010). Threats to the valid use of assessments. *Assessments in Education: Principles, Policy and Practice*. Retrieved from <https://tandfonline.com/doi/abs/>
- DeMars, C. E., Lau, A. (2011). Differential item functioning detection with latent classes: how accurately detect who is responding differentially? *Educational Psychology Measures*. 71, 597–616.
- Dorans, N.J., & Holland, P.W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P.W. Holland, & H. Wainer (Eds.), *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum. 35–66
- Dorans, N.J., Moses, T.P., & Eignor, D.R. (2010). *Principles and Practices of Score Equating*. Educational Testing Service (ETS). Princeton, New Jersey. <http://www.ets.org/research/contact.html>. Retrieved on 20/10/2018.
- Emaikwu, S.O. (2011). *Fundamentals of Test, Measurement and Evaluation with Psychometric Theories*. Selfers Academic Press. Makurdi. 89–92
- Emaikwu, S.O. (2013). *Fundamental of Research Methods and Statistics*. Makurdi: Selfers Academic Press Limited..
- Enunwa, C.I., & Agi, C.I. (2019). Assessment of construct validity under- representation threat in senior secondary school mathematics achievement test: A recipe to functional education. *Journal of Evaluation*. 4(1), 149–158
- Ethe, N., & Odjegba, O.G. (2019). Comparison of WAEC and NECO SSCE mathematics multiple – choice test item parameters across school location: Application of item response theory. *Journal of Evaluation*. 4(1), 194–205
- Federal Republic of Nigeria (2014). *National Policy on Education*. (6th Edition). NERDC Press. Lagos.
- Gruijter, D.N.M., & Leo, J.T.K. (2008). *Statistical test theory for the behavioural sciences*. Taylor & Francis Group, Boca Raton. Retrieved from <http://www.crcpress.com>.
- Hambleton, R. K. & Swaminathan, H. (2013). *Item response theory: Principles and applications*. Netherlands: Springer.
- Han, K.T. (2008). *Impact of item parameter drift on test equating and proficiency estimates*. Electronic Doctoral Dissertations for UMass Amherst. Paper AAI3325324. <http://scholarworks.umass.edu/dissertations/AAI3325324>
- Han, K.T., & Guo, F. (2011). Potential impact of item parameter drift due to practice and curriculum change on item calibration in computerized adaptive testing. *Graduate Management Admission Council*. 11(2).

- Harbour-Peters, V.F.A. (1999). *Noteworthy points on measurement and evaluation*. Enugu: Snaap Press Limited.
- Huggins-Manley, A. C. (2016). Psychometric consequences of subpopulation item parameter drift. *Educational and Psychological Measurement*. 77(1), 143–164. journals.sagepub.com/home/epm. Retrieved 22/5/2019.
- Kim, D.H., & Huynh, H. (2007). Comparability of computer and paper-and-pencil versions of algebra and biology assessments. *Journal of Technology, Learning, and Assessment*. 6(4). Retrieved 6/15/09 from <http://www.jtla.org>.
- Koutsoyiannis, A. (1977). *Theory of Econometrics*. (Second Edition). London: Macmillan Press Ltd.
- Krause, J. (2012). *Assessment of item parameter drift of known items in a university placement exam*. An unpublished Master's thesis. Arizona State University.
- Kyauve, T.A. (2019). *Development and validation of mathematics achievement test for senior secondary school students in Benue state Nigeria*. Master's thesis. Department of educational foundations and general studies; Federal university of agriculture, Makurdi.
- Mehrens, W.A., & Lehmann, J.I. (1991). *Measurement and evaluation in education and psychology*. Ted Buchholz. Wadsworth/Thomson Learning, 10 Davis Drive, Belmont CA USA. Retrieved from <http://www.wadsworth.com>.
- Meyer, J. P. (2014). *Applied measurement with jMetrik*. New York: Routledge Taylor and Francis Group.
- Middleton, F. (2019). *The four types of validity*. <https://www.scribber.com/methodology>. Retrieved on 7/11/2019.
- Mustapha, B., & Abba, B. (2019). Differential item functioning of 2010 senior school certificate English language examination scores of students in Dawakin Kudu education zone of Kano state Nigeria. *African Journal of Theory and Practice of Educational Assessment (AJTPEA)*. 7, 62–72
- Nworgwu, B.G. (2015). *Educational measurement and evaluation: Theory and practice*. (2nd Edition). Enugu: University Trust Publishers.
- Nworgwu, B.G. (2015). *Educational research: Basic issues and methodology*. (3rd Edition). University Trust Publishers. Enugu.
- Odili, J. N. (2004). Effect of language manipulation on differential item functioning of Biology multiple choice test. *University of Nigeria Research Publications*.
- Ogbebor, U., & Onuka, A. (2013). Differential item functioning method as an item bias indicator. *Educational Research*. 4(4), 367–373. <http://www.interestjournals.org/ER>.
- Ogomaka, P.M.C., Onah, F.E., & Okwa, C.O. (2017). Concurrent validity of the three instruments for identifying giftedness in primary school pupils. *Nigerian Journal of Educational Research and Evaluation*. 16(1), 53–63.
- Okoh, O. (2020). *Analysis of item discrimination and reliability indices as a function of change in mathematics objective tests of basic education certificate examination (BECE) in*

- Benue State*. Master's thesis. Department of educational foundations and general studies; Federal University of Agriculture, Makurdi.
- Orheruata, M.U. (2015). Item parameter drift in certificate examinations and its implication on decision making. *African Journal of Theory and Practice of Educational Assessment (AJTPEA)*. 2, 98–105
- Orheruata, M.U., Omorogiuwa, O.K., & Osunde, A.U. (2017). Assessing scale drift of WAEC and NECO SSCE Agricultural science multiple choice items with item response theory. *ASSEREN Journal of Education*. 2(1), 15–24.
- Osadebe, P.U., & Agbure, B. (2018). Assessment of differential item functioning in social studies multiple choice questions in basic education certificate examination. *European Journal of Education Studies*. 4(9), 236–257. www.oapub.org/edu. Retrieved on 18/06/2019.
- Oshima, T.C., Raju, N.S., & Nanda, A.O. (2006). A new method for assessing the statistical significance in the differential functioning of items and tests (DFIT) framework. *Journal of Educational Measurement*. 43(1), 1–17.
- Paek, I., & Young, M.J. (2005). Investigation of student growth recovery in a fixed – item linking procedure with a fixed – person prior distribution for mixed-format test data. *Applied Measurement in Education*. 18, 199–215.
- Roever, C. (2005). “That’s not fair!” *Fairness, bias, and differential item functioning in language testing*. Retrieved from the University of Hawai’I System Web site: <http://www2.hawaii.edu/~roever/brownbag.pdf>.
- Tawiah, P.K. (2006). *Basic Economics for West Africa*. (Fourth Edition). Idodo Umeh Publishers Ltd. Benin City.
- Thompson, N. (2017). *Identifying and Dealing with Threats to Test Security*. *Assessment Systems for Good Measure*. Retrieved from <http://www.assess.com/identifying-dealing-threats-test-security/>.
- Toit, M.D. (ed). (2003). *IRT from SSI: Bilog-MG, Multilog, Parscale, Testfact*. *Scientific Software International, Inc.* United State of America. Retrieved from <http://www.ssicentral.com>.
- Ukozor, F.F.I., Onah, F. E. (2017). Application of three parameter latent trait model in development of secondary school Economics achievement test. *Nigerian Journal of Educational Research and Evaluation*. 16(1), 79-99.
- Ukozor, F.F.I., Onah, F.E., & Ukah, F.O. (2017). Application of three parameter latent trait model in development of secondary school Economics achievement test. *Nigerian Journal of Educational Research and Evaluation*. 16(1), 79–93.
- Wells, C. S., Subkoviak, M. J., & Serlin, R. C. (2002). The effect of item parameter drift on examinee ability estimates. *Applied Psychological Measurement*, 26, 77–87.