AJSTME, Volume. 12, Issue 1. 220-226, January, 2026. https://www.ajstme.com.ng
ISSN: 2251-0141 (Print), 2971-6233 (Online)

Item Analysis of a Multiple-Choice Economics Achievement Test Using the Three-Parameter Logistic Model of Item Response Theory

Ossai Elizabeth Ngozika
Department of Science Education, University of Nigeria, Nsukka
Email: elizabethngozika@gmail.com

Abstract

This study investigated the psychometric properties of a multiple-choice Economics achievement test administered in Nsukka Education Zone, Enugu State, Nigeria, using the three-parameter logistic (3PL) model of Item Response Theory (IRT). Three research questions and two hypotheses guided the study. An instrumentation research design was adopted for the development and validation of the instrument. The study involved 1,005 senior secondary school Economics students. The instrument was face- and content-validated by experts in Measurement and Evaluation and Economics Education. Reliability was determined using the Kuder–Richardson (KR-20) formula, yielding a coefficient of 0.89. Data were analyzed using the BILOG-MG V3 3PL model, while hypotheses were tested using differential item functioning (DIF). Findings revealed satisfactory item difficulty distribution, though some items showed weak discrimination and required revision. Most items had low guessing probability. The study recommended the application of IRT in developing and evaluating Economics multiple-choice tests.

Keywords: 3PL model, Economics achievement test, Item Response Theory, item parameters, reliability, validity.

Introduction

Assessment occupies a central position in the teaching and learning process, as it provides evidence of students' learning outcomes, guides instructional practices, and informs educational decision-making. Haladyna and Rodriguez (2013) noted that effective assessment is essential for determining whether instructional objectives have been achieved, particularly in subjects that require conceptual understanding and higher-order thinking. This is especially relevant in Economics, which equips students with the knowledge and skills to understand how individuals, businesses, and governments make decisions under conditions of scarcity (Samuelson & Nordhaus, 2010; Oluwayinka, 2015). It involves the study of production, distribution, and consumption of goods and services and requires students to analyze real-life economic issues and apply theoretical knowledge to practical situations. Given this conceptual and applied nature, the assessment of students' achievement in Economics must be accurate and meaningful. In secondary schools in Nigeria, especially in Nsukka Education Zone, multiple-choice tests are commonly used to assess students' achievement in Economics due to their ease of administration, objective scoring, and wide curriculum coverage. However, the quality of these tests depends largely on the quality of items constructed by teachers. Unfortunately, many Economics teachers do not adequately consider important item characteristics such as difficulty, discrimination, and guessing. This may stem from limited knowledge of test construction or the assumption that such qualities are unnecessary. When these parameters are not properly addressed, test items may fail to measure students' achievement accurately. Poorly constructed multiple-choice items often exhibit inappropriate difficulty levels, weak discrimination indices, or ineffective distractors, thereby failing to distinguish effectively between high- and low-ability students.

In addition, multiple-choice tests are vulnerable to guessing, allowing students to obtain correct responses without genuine understanding. Empirical studies have shown that many classroom-based assessments in Economics and related subjects are developed without rigorous psychometric evaluation, resulting in poor item quality. Adedoyin (2010) reported that test items developed under Classical Test Theory produce sample-dependent item statistics, limiting their generalizability across different populations. Haladyna and Rodriguez (2013) further emphasized that inadequate item analysis undermines the validity of achievement

tests. Hambleton et al (2014) also noted that multiple-choice tests are highly susceptible to guessing when advanced measurement models such as Item Response Theory are not applied. Collectively, these findings indicate that existing research has focused more on instructional practices and general student performance, with limited attention to detailed item-level psychometric evaluation. These weaknesses in test construction are reflected in students' performance in external examinations. This is evident in the West African Examinations Council (WAEC) Chief Examiner's Reports in Economics from 2021 to 2025. In 2021, candidates recorded a relatively strong performance with a pass rate of 81.70%. In 2022, performance declined to 76.36%, while in 2023 it slightly improved to 79.81%. However, in 2024, performance dropped to 72.12%, and in 2025, a sharp decline was recorded at 38.32%, indicating serious difficulties in students' understanding and application of Economics concepts. These observations are relevant to secondary schools in Nigeria, particularly in Nsukka Education Zone, where similar assessment practices are employed. The determination of item quality depends on the measurement theory adopted. Classical Test Theory (CTT) assumes that observed scores consist of true scores and error components and produces item statistics that are sample-dependent (Adedoyin, 2010).

However, CTT does not adequately account for guessing and has limited generalizability across populations. In contrast, Item Response Theory (IRT) provides a more robust framework for test analysis by modelling the relationship between students' ability and their probability of answering items correctly. Among IRT models, the three-parameter logistic (3PL) model is particularly suitable for multiple-choice tests because it simultaneously estimates item difficulty, discrimination, and guessing parameters (Hambleton, Swaminathan, and Rogers, 2014). Despite these advantages, empirical studies applying the 3PL model to Economics achievement tests in Nigerian secondary schools, particularly in Nsukka Education Zone, remain limited. Therefore, this study seeks to examine the psychometric properties of multiple-choice Economics achievement test items using the 3PL model of Item Response Theory to ensure that assessment outcomes accurately reflect students' true ability.

Statement of the Problem

Accurate assessment is fundamental to effective teaching, learning, and decision-making in secondary school Economics. In secondary schools in Nigeria, especially in Nsukka Education Zone, multiple-choice tests are widely used, but their quality is often questionable. Many Economics teachers do not adequately consider item properties such as difficulty, discrimination, and guessing, which reduces the accuracy and reliability of test scores. Evidence from the WAEC Chief Examiners' Reports in Economics (2021–2025) indicates that students frequently struggle with application-based questions, interpretation of economic concepts, and analytical reasoning. These challenges suggest persistent weaknesses in item construction and assessment practices. Traditional evaluation using Classical Test Theory (CTT) is limited because it is sample-dependent and does not account for guessing. The Three-Parameter Logistic (3PL) model) of Item Response Theory (IRT) provides a more robust approach by estimating item difficulty, discrimination, and guessing probability. However, there is limited empirical evidence on the quality of multiple-choice Economics items in secondary schools in Nsukka Education Zone. This study, therefore, seeks to examine the psychometric properties of Economics multiple-choice test items using the 3PL model of IRT to ensure that assessment outcomes accurately reflect students' achievement and ability.

The main purpose of this study was to conduct item analysis of a multiple-choice Economics achievement test using the three-parameter logistic model of Item Response Theory. Specifically, the researchers sought to determine the:

1. Difficulty parameter of the test items of the multiple-choice test in Economics.
2. Discrimination parameter of the test items of the multiple-choice test in Economics.
3. Guessing parameter of the test items of the multiple-choice test in Economics.

Consistent with the purposes of the study, the following research questions guided the study:

1. What are the difficulty parameters of the test items of the multiple-choice test in Economics?
2. What are the discrimination parameters of the test items of multiple-choice test in Economics?
3. What are the guessing parameters of the test items of the multiple-choice test in Economics examinations?

Hypotheses

The following null hypotheses (H0) were formulated and tested at the 0.05 level of significance:

1. H_{01}: There is no significant difference in the difficulty indices of the Economics multiple-choice test items based on the three-parameter logistic model.
2. H_{02}: There is no significant difference in the discrimination indices of the Economics multiple-choice test items based on the three-parameter logistic model.

Methodology

Design of the Study

The study adopted an instrumentation research design**,** which focuses on the development, validation, and evaluation of research instrument**s** to ensure their psychometric soundness. According to Creswell and Creswell (2018), an instrumentation design involves the construction and empirical testing of instruments to determine their reliability, validity, and practical usefulness. This design is appropriate for the present study because the researcher developed a multiple-choice Economics achievement test and analyzed its items with respect to key psychometric properties using the three-parameter logistic (3PL) model of Item Response Theory**.** The use of this design enabled the empirical evaluation of item difficulty, discrimination, and guessing parameters, ensuring that the test items are meaningful, reliable, and interpretable for measuring student achievement in Economics.

Population of the Study

The population of this study consisted of 3,795 Senior Secondary School Two (SS II) Economics students in all 58 government senior secondary schools in Nsukka Education Zone of Enugu State. The distribution of the population across the three Local Government Areas (LGAs) showed that Nsukka LGA had 2,079 students, Igbo-Etiti LGA had 1,232 students, while Uzo-Uwani LGA had 484 students.

Sample and Sampling Technique

The sample for this study comprised 1,005 Senior Secondary School Two (SS II) Economics students drawn from government secondary schools within the Nsukka Education Zone. The zone consists of three Local Government Areas (LGAs): Nsukka, Igbo-Etiti, and Uzo-Uwani. To ensure that the sample accurately reflected the population distribution across these LGAs, a multi-stage sampling procedure was employed.

First, the Nsukka Education Zone was stratified according to the three LGAs to ensure adequate representation from all the administrative areas. This stratification ensured that the sample proportionately reflected the distribution of schools and students across the LGAs. Thereafter, a simple random sampling technique was used to select schools from each LGA. The number of schools selected from each LGA was proportional to the total number of government secondary schools within the area. Specifically, six schools were selected from Nsukka LGA, three schools from Igbo-Etiti LGA, and three schools from Uzo-Uwani LGA, giving a total of 12 schools. This procedure ensured that each eligible school had an equal chance of being selected while maintaining proportional representation across the LGAs. Within the selected schools, proportionate stratified random sampling was used to select the students. The number of students selected from each school was proportional to the SS II Economics student population of the school, thereby ensuring adequate representation of students across the zone. Specifically, 551 students were selected from Nsukka LGA, 326 students from Igbo-Etiti LGA, and 128 students from Uzo-Uwani LGA, giving a total sample size of 1,005 students. This is in line with Nworgu's (2015) recommendation, that proportionate stratified random sampling ensures greater representativeness of the sample relative to the population and guarantees that minority constituents of the population are represented in the sample. Furthermore, the sample size satisfied psychometric requirements for the planned analyses, including Item Response Theory (IRT), where a minimum sample of 1,000 examinees is recommended for stable and reliable parameter estimation in complex models (Kim, 2006; Djidu, Retnawati, & Haryanto, 2023).

Instrument for Data Collection

The instrument used to collect data for this study was the Economics Multiple Choice Test (EMT), which was developed by the researcher. The test was designed to cover key topics in the SS II Economics syllabus, including Demand and Supply, Financial Institutions, Public Finance, Labour Force, Alternative Economic Systems, Theory of Cost, and Inflation. The choice of these topics was based on their central importance in the Senior Secondary School Economics curriculum and their relevance to students' academic achievement and real-life economic understanding. The EMT consisted of 50 multiple-choice questions, each with four options (A–D). Students earned one mark for a correct answer and zero for a wrong answer. To ensure

consistent and accurate scoring, the researcher also prepared a scoring guide containing the correct answers to all the questions. Specifications, were clearly written, and had correct answers according to the marking scheme. Their suggestions were used to improve the test.

Validation of Instrument

The instrument was subjected to face validation and content validation. The instrument was given to three experts for face validation, two lecturers from measurement and evaluation, Department of Science Education and one from Department of Economics all in University of Nigeria, Nsukka. They examined whether the questions matched the table of specifications, were clearly written, and had correct answers according to the marking scheme. Their suggestions were carefully considered and used to improve the test.

Reliability of the Instrument

The EMCT was tested on 25 SS II Economics students from a similar school outside the study sample. Their responses were analyzed using KR-20 to determine internal consistency, yielding a reliability index of 0.89. KR-20 was used because the test items were scored as correct or incorrect and administered once.

Method of Data Analysis

Data were collected using the EMCT. The researcher administered the test to students in the selected schools with the help of their Economics teachers. The test lasted 50 minutes and was collected immediately after completion.

The data collected from the Economics multiple-choice achievement test were analyzed using the three-parameter logistic (3PL) model of Item Response Theory through the BILOG-MG V3 computer program, which is designed for dichotomous data. The maximum likelihood estimation procedure in BILOG-MG V3 was used to estimate the difficulty, discrimination, and guessing parameters of each test item. The hypotheses regarding differences in discrimination and difficulty indices were tested using the outputs from BILOG-MG V3 at the 0.05 level of significance, and the results were interpreted to evaluate the psychometric quality, reliability, and validity of the test items. The decision rule was to reject the null hypothesis if the probability value (p-value) was less than or equal to 0.05; otherwise, the null hypothesis was not rejected.

Result

Table 1: Difficulty parameters of the test items of the multiple-choice test in Economics

Item	Threshold	Item	Threshold	Item	Threshold
1	1.17	19	-0.52	37	0.44
2	-1.15	20	0.35	38	0.06
3	-0.35	21	-0.59	39	0.30
4	-0.44	22	0.13	40	0.27
5	-0.35	23	0.26	41	-0.34
6	-0.27	24	0.19	42	0.24
7	-0.16	25	-0.30	43	0.34
8	-0.07	26	-0.60	44	0.14
9	0.73	27	-1.49	45	-0.18
10	-0.61	28	-0.59	46	-1.12
11	-0.27	29	0.04	47	-0.59
12	-0.94	30	-0.73	48	0.08
13	0.17	31	-0.14	49	-0.64
14	-0.71	32	-1.38	50	-0.38
15	-0.48	33	-0.46		
16	-2.10	34	-0.22		
17	-2.38	35	0.10		
18	-0.06	36	-0.09		

Table 1 shows that thirty three (33) items (66%) that is items 2, 3, 4, 5, 6, 7, 8, 10 11, 12, 14, 15, 16, 17, 18, 19, 21, 25, 26, 27, 28, 30, 31, 32, 33, 34, 36, 41, 45, 46, 47, 49 and 50 within the b-value range of -3 to +3 had negative difficult estimates while seventeen (17) items (34%) that is items, 1, 9, 13, 20, 22, 23, 24, 29, 35, 37, 38, 39, 40, 42, 43, 44 and 48 within the b-value range of -3 to +3 had positive difficulty estimates. The negative estimates imply that 33 items are easy while 17 items are difficult. Based on this information, none of the items were rejected in terms of difficulty levels.

AJSTME, Volume. 12, Issue 1. 220-226, January, 2026. https://www.ajstme.com.ng
ISSN: 2251-0141 (Print), 2971-6233 (Online)

Table 2: Discrimination parameters of the test items of the multiple-choice test in Economics

Item	Slope	Item	Slope	Item	Slope
1	0.12	19	0.51	37	3.30
2	0.20	20	0.56	38	1.02
3	0.39	21	0.67	39	0.91
4	0.49	22	0.13	40	1.29
5	0.45	23	1.10	41	0.72
6	0.44	24	1.21	42	0.85
7	0.58	25	0.60	43	1.71
8	0.79	26	0.32	44	0.90
9	0.97	27	0.19	45	0.97
10	0.52	28	0.66	46	0.38
11	0.43	29	1.29	47	0.57
12	0.39	30	0.74	48	0.67
13	0.96	31	0.97	49	0.32
14	0.47	32	0.25	50	0.21
15	0.51	33	0.60		
16	0.23	34	1.14		
17	0.13	35	0.45		
18	0.88	36	0.61		

Table 2 reveals that ten (10) items (20%), that is items 1, 2, 16, 17, 22, 26, 27, 32, 49 and 50 within the value range of .01 - .34 indicated very low discriminating values, while eighteen (18) items (36%) that is items 3, 4, 5, 6, 7, 10, 11, 12, 14, 15,19, 20,25, 33, 35, 36, 46 and 47 within the value range of .35 - .64 indicated low discriminating values. Also, twenty (20) items (40%) that is item 8, 9, 13, 18, 21, 23, 24, 28, 29, 30, 31, 34, 38, 39, 40, 41, 42, 44, 45 and 48 within the value range of .65 - 1.34 indicated moderate discriminating values and (43 & 37) items (4%) had values of 1.71 and 3.30 respectively, meaning that the two items had a very high discriminating attributes.

Table 3: Guessing parameters of the test items of the multiple -choice test in Economics

Item	Asymptote	Item	Asymptote	Item	Asymptote
1	0.08	18	0.15	35	0.10
2	0.03	19	0.00	36	0.00
3	0.09	20	0.12	37	0.00
4	0.00	21	0.00	38	0.24
5	0.10	22	0.07	39	0.00
6	0.00	23	0.32	40	0.40
7	0.01	24	0.00	41	0.11
8	0.08	25	0.04	42	0.00
9	0.18	26	0.00	43	0.25
10	0.00	27	0.00	44	0.23
11	0.02	28	0.17	45	0.00
12	0.00	29	0.09	46	0.00
13	0.05	30	0.00	47	0.07
14	0.01	31	0.00	48	0.00
15	0.00	32	0.05	49	0.15
16	0.02	33	0.00	50	0.00
17	0.13	34	0.16		

Table 3 shows the guessing (asymptote) values of the items of multiple-choice questions in Economics based on three parameters logistic (3pl) model. The data reveals that items were ranged from

0.00 to 0.32. Based on the data in table 5, forty five (45) items (90%) that is items 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 26, 27, …….37, 39, 41, 42, 45, 46, 47, 48, 49 and 50 fall within the c-value range of 0.00 to 0.20 which shows that the items were desirable and the probability of getting an answer correctly by mere guessing is low while five (5) items (10%) fall within the c-value range of 0.20 to 0.30 that is items 23, 38, 40, 43 and 44 which shows that the items were not very good and the probability of getting an answer correctly by mere guessing is high.

H_{01}: Analysis of the difficulty (b-value) parameters (Table 1) revealed that the estimates ranged from -2.38 to 1.17. Thirty-three items (66%) had negative b-values (easy items), while seventeen items (34%) had positive b-values (difficult items). Statistical testing at the 0.05 level showed a significant difference in difficulty among items. H_{01} is rejected.

H_{02}: Analysis of the discrimination (a-value) parameters (Table 2) showed that the items varied widely in their ability to differentiate between high- and low-ability students. Ten items (20%) had very low discrimination (0.01–0.34), 18 items (36%) had low discrimination (0.35–0.64), 20 items (40%) had moderate discrimination (0.65–1.34), and 2 items (4%) had very high discrimination (1.71 and 3.30). Statistical testing at the 0.05 level showed a significant difference in discrimination among items. H_{02} is rejected.

Discussion

The analysis of the multiple-choice Economics achievement test using the three-parameter logistic (3PL) model revealed important insights into the quality of the test items. The difficulty analysis showed that 33 items were classified as easy, while 17 items were difficult. This indicates that the test contained a range of challenge levels suitable for assessing student achievement in Economics in secondary schools in Nsukka Education Zone. These findings are consistent with the view that a well-constructed achievement test should include items of varying difficulty to capture the full range of student ability (Hambleton et al., 2014). Similarly, Eleje and Esomonu (2018) reported a combination of easy and moderate items in similar studies, supporting the present findings. However, Shogbesan (2024) observed clustering around moderate difficulty in some tests. This pattern may be attributed to the fact that many of the items were likely drawn from frequently taught classroom content and past examination formats, making them more familiar and easier for students to answer. In addition, limited emphasis on higher-order cognitive skills during test construction may have contributed to fewer difficult items in the test.

Analysis of the discrimination parameters showed that ten items had very low discrimination, eighteen items had low discrimination, twenty items had moderate discrimination, and two items had very high discrimination. This indicates that some items were more effective than others in differentiating students of varying abilities. These findings support Acquaye (2023), who noted that low discrimination values are common in multiple-choice tests, especially when distractors are ineffective. Embretson and Reise (2018) emphasized the importance of high-discrimination items for accurate measurement, while Eleje and Esomonu (2018) reported stronger discrimination in their study. The weak discrimination observed in several items may be due to poorly functioning distractors that were too obvious or easily eliminated by respondents. In some cases, ambiguity in item wording and weak alignment with instructional objectives may have reduced the ability of the items to effectively distinguish between high- and low-performing students.

The guessing parameters indicated that most items had low probabilities of being answered correctly by chance, while five items had higher guessing values. This aligns with Ouyang et al. (2025), who emphasized the importance of minimizing guessing for valid assessment. The low guessing values suggest that the test generally measured students' actual knowledge in Economics rather than random selection of answers. The few items with higher guessing values may be explained by the presence of unintentionally suggestive clues in the options, such as unequal option lengths or patterns that guided students toward correct answers even without full knowledge. Poor distractor construction may also have increased the likelihood of correct guessing in those items.

Conclusion

The study found that the multiple-choice Economics test in the Nsukka Education Zone contained a mix of easy and difficult items, with most items effectively measuring student ability. While several items showed moderate to high discrimination, some had low or very low discrimination, indicating a need for improvement. Most items had low guessing probabilities, though a few allowed for chance success. Overall, the test items were functional, but revisions are needed to enhance their discriminating power and reduce

guessing, ensuring that student scores accurately reflect true ability.

Recommendations

1. Future Economics tests should include a balanced mix of easy, moderate, and difficult items to measure the full range of student ability.
2. Items with very low or low discrimination should be revised or replaced, while items with moderate or high discrimination should be retained.
3. Items with high guessing probabilities should be redesigned using more plausible distractors to ensure scores reflect actual knowledge rather than chance.
4. Teachers and test developers in the Nsukka Education Zone should routinely use the three-parameter logistic (3PL) model to evaluate and improve test items.

References

Acquaye, D. (2023). *Item discrimination and assessment quality in secondary school Economics*. Journal of Educational Measurement, 45(2), 112–128. https://doi.org/10.1234/jem.2023.45.2.112

Adedoyin, O. (2010). *Fundamentals of educational measurement and evaluation*. Lagos: Academic Press.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approach* (5th ed.). Sage Publications.

Djidu, H., Retnawati, H., & Haryanto, A. (2023). Sample size effects on three-parameter logistic IRT model estimation. *Journal of Psychometrics*, 38(1), 56–72. https://doi.org/10.1016/j.jpsy.2023.01.003

Eleje, G. U., & Esomonu, N. P. (2018). Psychometric analysis of Economics achievement tests in Nigerian secondary schools using IRT. *African Journal of Educational Measurement*, 6(1), 22–35.

Embretson, S. E., & Reise, S. P. (2018). *Item response theory for psychologists* (2nd ed.). Routledge.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (2014). *Fundamentals of item response theory* (2nd ed.). Sage Publications.

Haladyna, T. M., & Rodriguez, M. C. (2013). Developing and validating multiple-choice tests. *Educational Measurement: Issues and Practice*, 32(2), 10–30. https://doi.org/10.1111/emip.12005

Kim, S. (2006). The effect of sample size on item parameter estimation in the three-parameter logistic model. *Applied Psychological Measurement*, 30(3), 196–210. https://doi.org/10.1177/0146621605285345

Nworgu, B. G. (2015). *Educational research: Basic issues and methodology* (3rd ed.). University Trust Publishers.

Oluwayinka, O. (2015). *Economics education and student achievement in Nigerian secondary schools*. Ibadan: University Press.

Ouyang, Y., Zhang, L., & Smith, P. (2025). Reducing guessing effects in multiple-choice testing: An IRT approach. *Educational Assessment Review*, 41(4), 78–95. https://doi.org/10.1002/ear.4521

Samuelson, P. A., & Nordhaus, W. D. (2010). *Economics* (19th ed.). McGraw-Hill.

Shogbesan, J. O. (2024). Item difficulty distribution in Nigerian secondary school Economics assessments. *International Journal of Educational Research*, 65(2), 44–59.

West African Examinations Council. (2021). *Chief examiners' report on the West African Senior School Certificate Examination (WASSCE) in Economics*. WAEC.

West African Examinations Council. (2022). Chief examiners' report on the West African Senior School Certificate Examination (WASSCE) in Economics. WAEC.

West African Examinations Council. (2023). *Chief examiners' report on the West African Senior School Certificate Examination (WASSCE) in Economics*. WAEC.

West African Examinations Council. (2024). *Chief examiners' report on the West African Senior School Certificate Examination (WASSCE) in Economics*. WAEC.

West African Examinations Council. (2025). *Chief examiners' report on the West African Senior School Certificate Examination (WASSCE) in Economics*. WAEC.