Journal of People and Worldviews (JPW)
Volume 1, Number 2
Empirical Article https://doi.org/10.5281/zenodo.21127576 Open Access

Guessing Behaviour of Multiple Choice Test Items: An Item Response Theory Analysis Using the 3pl-Model for 2019–2021 National Business And Technical Certificate Biology Examinations in Nigeria

Abstract

Guessing remains one of the most persistent psychometric threats to the validity of multiple-choice test scores in large-scale standardised examinations. This paper examined the guessing parameter (c-parameter) of the 2019, 2020 and 2021 National Business and Technical Certificate Examination (NABTCE) Biology Multiple Choice Test (MCT) items in Three-parameter Logistic (3PL) Item Response Theory (IRT) model. The sampling method was a multistage probability sampling whereby it was used to obtain 24,707 candidates out of a target population of 79,809 students across the six geopolitical zones of Nigeria. Winsteps 3.75.0 software was used in calibration of data. It was revealed that 98 percent, 94 percent and 100 percent of items in 2019, 2020 and 2021, respectively, met the acceptable guessing threshold of 0.25 on four-option Multiple Choice Tests. One-way Analysis of Variance (ANOVA) showed the significant difference in the estimates of parameter guessing in three examination years ($F = 3.228$, $p = 0.004$) as significant. These findings demonstrated commendable psychometric quality of the NABTCE Biology Multiple Choice Test items. This paper suggested compulsory pre-administration IRT calibration, reconstruction of flagged items with specificity, and incorporation of guessing parameter analysis in the policy of national assessment in order to ensure integrity of the outcomes of Biology education in Nigeria.

Keywords: Guessing parameter, Item response theory, 3PL model, Multiple-choice items, NABTCE, Biology assessment

Otasowie Ogiemwonyi, Ph.D.
Department of Educational Evaluation and Counselling Psychology
Faculty of Education, University of Benin, Benin City, Nigeria

Prof. Osamede Kingsley Omorogiuwa
Department of Educational Evaluation and Counselling Psychology
Faculty of Education, University of Benin, Benin City, Nigeria

ORCID ID: 0009-0008-0632-4395

Corresponding Author's Email: ogieotasowie@gmail.com

Date Received: 15th June, 2026

Date Accepted: 28th June, 2026

Introduction

Standardised assessments on a large scale assume a central role in defining educational paths, policy choices and qualification of skills, both in academic and professional aspects. In Nigeria, the National Business and Technical Certificate Examination (NABTCE) administered by the National Business and Technical Examinations Board (NABTEB) plays this very important role among students of technical and vocational education. Each year hundreds of thousands of applicants take the Biology Multiple Choice Test (MCT) and the outcome of this testing defines entry to tertiary education and career opportunities in medicine, agriculture,

environmental sciences, and other professions. Sometimes, a low examinee would guess correctly on a challenging dichotomous item. Yu (2013) states that guessing is the likelihood that an examinee with very low ability can be able to guess the correct response to an item and therefore has a greater-than-zero probability of answering correctly. It is also known as pseudo-chance parameter and has a lower asymptote value. The item guessing parameter is the lowest value that an ICC curve attains.

From an achievement testing perspective, guessing can be conceptualized as when an individual does not definitively know the answer to a question, however attempts to answer the question anyway. There are likely many aspects that factor into the ability of an individual to arrive at the correct answer through guessing. The first and most straightforward would be through random guessing. This would be an act where the individual is unable to rule out any options and simply picks an answer at random. The chances that an individual will have the tendency of random guessing in a standardized achievement test would likely occur when an individual does not have the ability to understand the question nor the answer. The Three-Parameter Logistic (3PL) IRT model is especially well-adapted to this task, because it directly models a pseudo-guessing parameter (c-parameter), also known as the lower asymptote - the probability that an examinee with very low ability will respond correctly to an item.

Research Questions

The study was guided by the following research questions:

i. What are the guessing parameter estimates of the 2019 NABTCE Biology MCT items?

ii. What are the guessing parameter estimates of the 2020 NABTCE Biology MCT items?

iii. What are the guessing parameter estimates of the 2021 NABTCE Biology MCT items?

iv. Is there a significant difference in guessing parameter estimates across the 2019, 2020, and 2021 NABTCE Biology MCT items?

Hypothesis: There is no significant difference in the guessing parameter indices of the Biology Multiple choice items across May/June 2019, 2020 and 2021 NABTCE Biology multiple choice test items.

Literature Review

In the context of Item Response Theory models, the likelihood of providing a certain response to an item is contingent upon one or more latent variables as well as the parameters associated with said items. The primary objective of Item Response Theory analysis often involves examining the characteristics of individual test items and assessing the latent variables associated

Citation: Ogiemwonyi,Otasowie & Omorogiuwa, Osamede Kingsley. "Guessing Behaviour of Multiple Choice Test Items: An Item Response Theory Analysis Using the 3pl-Model for 2019–2021 National Business And Technical Certificate Biology Examinations in Nigeria". *Journal of People and Worldviews (JPW),* 2026: pp175-184.

with each test taker. In situations involving binary elements, such as true/false, yes/no, correct/incorrect, and agree/disagree, the act of making educated conjectures is an inherent component of the testing environment. The technique involves students with limited understanding attempting to select the proper answer. The act of making educated guesses in an objective examination serves as an illustration of the need of including this behavior into the modelling process for examinees with lower levels of competence.

Students that possess limited skill levels have a non-negligible likelihood of correctly answering even the most difficult questions on multiple-choice tests, sometimes resorting to guessing as their preferred method of response. According to Chiu and Camilli (2012), the use of number-correct scores in research may result in statistical bias due to the introduction of test score unreliability caused by guessing. This implies that it is necessary to thoroughly reassess and thereafter evaluate or eliminate any items that the test takers are likely to guess correctly. From a purely psychometric standpoint, it can be seen that examinees with lesser cognitive capacities who manage to provide correct answers to difficult problems are more likely to have relied on guessing, as opposed to examinees with higher cognitive ability.

The Three-Parameter Logistic Model provides an estimation of the likelihood that an individual, with a certain level of skill, would answer an item correctly, taking into account the item's difficulty index, discriminating index, and guessing index. The three parameters being used for these models are the „b" „a" and „c" parameters which are the difficulty, discrimination and guessing parameters. The three-parameter logistic model (3PL) was proposed by Lord in 1980 and developed in educational testing to extend the application of item response theory to multiple choice items that may elicit guessing.

The underlying assumption of the model is that the inclusion of three factors, namely difficulty, discrimination, and guessing, is essential for accurately estimating the link between the likelihood of a right response to an item and the degree of the trait being assessed in the examinee. The act of guessing is providing a response or forming an evaluation on a matter without possessing complete certainty of all pertinent information. The utilization of guessing is a commonly employed test-taking approach employed by those who are undertaking a multiple choice examination. This approach offers the potential for accurate item responses, especially in cases when the examinee possesses limited understanding of the topic area. The impact of fortunate predictions on performance may be observed, as well as the potential decrease

Citation: Ogiemwonyi,Otasowie & Omorogiuwa, Osamede Kingsley. "Guessing Behaviour of Multiple Choice Test Items: An Item Response Theory Analysis Using the 3pl-Model for 2019–2021 National Business And Technical Certificate Biology Examinations in Nigeria". *Journal of People and Worldviews (JPW),* 2026: pp175-184.

in performance for highly capable examinees due to negligence.

Theoretical Framework

Item Response Theory and the 3PL Model

The item Response Theory (IRT) is a collection of mathematical models that can be used to explain the interaction between the latent ability of an examinee and the likelihood of a correct item response. The Three-Parameter Logistic (3PL) model (Birnbaum, 1968; Lord, 1980) is expressed as:

$$P(X = 1 \mid \theta) = c + (1 - c) \times [e^{(Da(\theta - b))} / (1 + e^{(Da(\theta - b))})]$$

where: a = discrimination parameter; b = difficulty parameter; c = pseudo-guessing parameter (lower asymptote); D = 1.702 (scaling constant); θ = examinee ability. The c-parameter is the focus of this study — it represents the minimum probability of a correct response regardless of ability. For a four-option MCT, a purely random guess has a one-in-four (0.25) probability of success. Thus, c-values substantially exceeding 0.25 signal that an item is vulnerable to guessing beyond chance expectation.

On most assessments used, there are usually four to five options to select from so the probability of choosing a correct response to an item by guessing is 20 or 25 %. For example, an examinee who randomly selects responses to items that have four response choices can answer these items correctly about 1 out of 4 times, meaning that the probability of guessing correctly is about 0.25. Omorogiuwa (2009) states that any item whose guessing parameter is more than 0.20 is a bad item. For c-value, it should be between 0.00 to 0.20 for test with five distracters, test items with c-values greater than 0.20 was considered as an item with a high probability of guessing the answer correctly and such test items would be classified as not good test items (Adedoyin & Mokobi, 2013). Additionally, items with c-values of at least 0.30 are considered unsatisfactory. The c values should be 0.2 or below according to Warm 1978 in (Obinne, 2012). According to Swist (2015), the purpose of the model was to accurately measure the performance of examinees with poor ability who were able to answer difficult items correctly by guessing. The act of guessing introduces a form of unreliability in test results. This distinction is important as it can lead to statistical bias when analyzing data using number-correct scores (Chiu & Camilli, 2012).

The act of guessing poses a significant challenge when attempting to accurately assess the approximate proficiency level of a test taker. Taking into consideration the lower asymptote serves to address the issue of guessing, however, it is crucial to additionally consider the variability among individuals in terms of the likelihood that individuals with higher abilities may incorrectly respond to easy items due to

Citation: Ogiemwonyi,Otasowie & Omorogiuwa, Osamede Kingsley. "Guessing Behaviour of Multiple Choice Test Items: An Item Response Theory Analysis Using the 3pl-Model for 2019–2021 National Business And Technical Certificate Biology Examinations in Nigeria". *Journal of People and Worldviews (JPW)*, 2026: pp175-184.

factors such as carelessness, attention difficulties, socially desirable responding, distractions, or other influences like a lack of motivation to engage with a task that is deemed easy (Liao *et al.*, 2012).

Methodology

Research Design

This study adopted a survey research design using a large sample of student responses from the entire population of students who sat for the NABTCE Biology multiple choice test items that were administered in 2019, 220 and 2021 across Nigeria.

Population and Sampling

The target population comprised all candidates who sat the NABTCE Biology MCT in the three study years: 29,254 in 2019; 31,411 in 2020; and 19,144 in 2021; a total population of 79,809 candidates. A multistage probability sampling technique was employed. At the first stage, Nigeria's six geopolitical zones (North-Central, North-East, North-West, South-East, South-South, South-West) served as primary sampling units. At subsequent stages, states, candidate scripts were randomly selected. The final sample comprised 24,707 candidates (approximately 31% of the population), a size considered adequate for stable 3PL IRT parameter estimation (Embretson & Reise, 2000).

Instrument and Data Analysis

The tool was the NABTCE Biology, a 50-item, multiple choice test presented under standard conditions by NABTEB. Before IRT calibration, the unidimensionality assumption was evaluated by Principal Components Analysis (PCA) of the residuals that verified dimensionality. Calibration of data was further carried out through Winsteps 3.75.0 software within the 3PL model. The model-data fit was tested with the help of MNSQ infit and outfit statistics (acceptable range: 0.6-1.4). Each item was extracted as c-parameter and compared with the 0.25 threshold. A one-way ANOVA was used to compare significant differences between years in estimation of c-parameter. Internal consistency reliability: a = 0.644 (2019), a = 0.779 (2020) and a = 0.942 (2021).

Results

Guessing Parameter Estimates: 2019 NABTCE Biology MCT

Table 1 presents a summary of the guessing parameter (c-parameter) distribution for the 2019 NABTCE Biology MCT items.

Citation: Ogiemwonyi,Otasowie & Omorogiuwa, Osamede Kingsley. "Guessing Behaviour of Multiple Choice Test Items: An Item Response Theory Analysis Using the 3pl-Model for 2019–2021 National Business And Technical Certificate Biology Examinations in Nigeria". *Journal of People and Worldviews (JPW),* 2026: pp175-184.

Table 1

Summary of Guessing Parameter Estimates for 2019 NABTCE Biology MCT Items (N = 50)

Category	No. of Items	Percentage (%)	c-Value Range	Status
Acceptable (c ≤ 0.25)	49	98%	0.04 – 0.25	Satisfactory
Prone to Guessing (c > 0.25)	1	2%	0.27	Flagged
Total	50	100%	0.04 – 0.27	—

Note. c ≤ 0.25 is the acceptable threshold for four-option MCT items. Items exceeding this threshold are flagged.

As shown in Table 1, 49 of 50 items (98%) in the 2019 NABTCE Biology MCT recorded c-parameter estimates within the acceptable range. The c-values ranged from 0.04 to 0.27, suggesting that the vast majority of items required genuine ability rather than chance for correct responses.

Guessing Parameter Estimates: 2020 NABTCE Biology MCT

Table 2 presents the guessing parameter distribution for the 2020 NABTCE Biology MCT items.

Table 2

Summary of Guessing Parameter Estimates for 2020 NABTCE Biology MCT Items (N = 50)

Category	No. of Items	Percentage (%)	c-Value Range	Status
Acceptable (c ≤ 0.25)	47	94%	0.03 – 0.25	Satisfactory
Prone to Guessing (c > 0.25)	3	6%	0.27 – 0.39 (Items 12, 25, 47)	Flagged
Total	50	100%	0.03 – 0.39	—

The 2020 administration exhibited slightly higher guessing susceptibility relative to 2019.

Guessing Parameter Estimates: 2021 NABTCE Biology MCT

Table 3 presents the guessing parameter distribution for the 2021 NABTCE Biology MCT items.

Citation: Ogiemwonyi,Otasowie & Omorogiuwa, Osamede Kingsley. "Guessing Behaviour of Multiple Choice Test Items: An Item Response Theory Analysis Using the 3pl-Model for 2019–2021 National Business And Technical Certificate Biology Examinations in Nigeria". *Journal of People and Worldviews (JPW)*, 2026: pp175-184.

Table 3

Summary of Guessing Parameter Estimates for 2021 NABTCE Biology MCT Items (N = 50)

Category	No. of Items	Percentage (%)	c-Value Range	Status
Acceptable (c ≤ 0.25)	50	100%	0.05 – 0.23	Satisfactory
Prone to Guessing (c > 0.25)	0	0%	N/A	None Flagged
Total	50	100%	0.05 – 0.23	—

Note. The 2021 administration achieved a perfect record: all 50 items fell within the acceptable guessing threshold, with a maximum c-value of 0.23.

The 2021 NABTCE Biology MCT achieved a remarkable outcome: all 50 items recorded c-parameter estimates below the 0.25 threshold, with values ranging from 0.05 to 0.23.

Cross-Year Comparison and ANOVA Results

Table 4 provides a comparative overview of guessing performance across the three years, and Table 5 presents the ANOVA results.

Table 4

Cross-Year Summary of Guessing Parameter Estimates (2019–2021 NABTCE Biology MCT)

Year	Items Acceptable	Items Flagged	Max c-Value	Flagged Items
2019	49 (98%)	1 (2%)	0.27	Item 18
2020	47 (94%)	3 (6%)	0.39	Items 12, 25, 47
2021	50 (100%)	0 (0%)	0.23	None
Combined	146 (97.3%)	4 (2.7%)	0.39 (2020)	4 items total

Note. Acceptable threshold: c ≤ 0.25 for four-option MCT items.

Table 5

One-Way ANOVA: Guessing Parameter Estimates Across 2019, 2020, and 2021 NABTCE Biology MCT Items

Year		Sum of Squares	df	Mean Square	F	Sig.
	Between Groups	10741.382	2	10.138	3.228	.004
	Within Groups	1298.798	147	33.636		
	Total	12040.180	149			

α = 0.05

Citation: Ogiemwonyi,Otasowie & Omorogiuwa, Osamede Kingsley. "Guessing Behaviour of Multiple Choice Test Items: An Item Response Theory Analysis Using the 3pl-Model for 2019–2021 National Business And Technical Certificate Biology Examinations in Nigeria". *Journal of People and Worldviews (JPW)*, 2026: pp175-184.

Table 9 shows that the F-value for 2019 is 3.228 and p-value is 0.004. Hypothesis testing at 0.05 level of significance, the p-value is less than the alpha level of 0.05. Consequently, this shows that there is a significant difference in the guessing parameter estimates across the three years 2019, 2020 and 2021 NABTCE Biology multiple choice test items. Therefore, the null hypothesis that there is no significant difference in the guessing parameter estimates across the three years is rejected. Conclusively, there is a significant difference in the guessing parameter estimates across the three years.

Discussion

The NABTCE Biology MCT can be thus described as psychometrically sound as far as the resistance is concerned in the aspect of guessing. Regarding the guessing parameter, the results indicated that 98%, 94%, and 100% of items fitted the acceptable range in 2019, 2020, and 2021, respectively. These findings are partially consistent with Obine (2012), who considered valid items to have guessing parameters not above 0.25. The 2021 test items, with lower asymptotes ranging from 0.00 to 0.23, were in agreement with Ainsworth (2015), who suggested that the guessing parameter should be within or roughly equal to the inverse of the number of options in multiple-choice questions.

The 2019 administration scored highly (98% acceptable), the 2020 administration recorded a slight drop (94%), and the 2021 administration scored a 100 percent. Such a non-linear trend, small decline and then large increase, could be due to inherent difference in quality writing of items or panel composition or the particular Biology content areas that were being stressed in any single year. This trend in terms of the general direction is positive and indicates that the item development processes in NABTEB are moving towards a positive direction.

The clear deviation of Item 12 in 2020 (c = 0.39) should be given special consideration. A c=0.39 means that even a purely random candidate, on average, stood a 39% chance of a correct answer, which is significantly higher than the 25 percent threshold of a four-choice question. Such guessing vulnerability is almost certainly indicative of item construction vulnerabilities; either a stem that gives unintended cues, distractors that are implausible or semantically open or a content domain that relies on everyday knowledge as opposed to specialised learning of Biology. The item must undergo complete professional scrutiny and, in case there are significant material weaknesses ascertained, be eliminated out of item banks.

The study also revealed a significant difference in the guessing parameter estimates across the three years, which is consistent with the findings of Leong et al. (2013), who found that guessing had a significant impact on their

Citation: Ogiemwonyi,Otasowie & Omorogiuwa, Osamede Kingsley. "Guessing Behaviour of Multiple Choice Test Items: An Item Response Theory Analysis Using the 3pl-Model for 2019–2021 National Business And Technical Certificate Biology Examinations in Nigeria". *Journal of People and Worldviews (JPW),* 2026: pp175-184.

research results. The guessing parameter estimates were also found to be stable, which is supported by DeAyala (2009), who reported comparable and consistent item parameter indices across different item subsets using an item response theory approach.

Conclusion

These findings demonstrate that the psychometric integrity of NABTCE Biology MCT items with respect to guessing susceptibility is not uniform across years and warrants systematic annual monitoring. These results have significant consequences in regard to the integrity of the Biology education certification in Nigeria, especially in relations to the role of NABTCE results in determining access to tertiary education in health sciences, agriculture, and environmental management where no genuine conceptual knowledge of Biology is to be underestimated.

Recommendations

Prerequisite IRT calibration prior to administration of the instrument.

- Annually NABTEB needs to identify a representative sample of schools (stratified by geopolitical zones) as field test schools.
- Software can be used, for example, Winsteps, jMetrik (free) or BILOG-MG; NABTEB can provide assistance to universities in calibration if there is no capacity in-house.

Training on capacity building for item writers and review panels.

- NABTEB should prepare a comprehensive Item Writing Training Manual and make a point of the distractors plausibility, item stem construction, interpretation of IRT parameters like c-parameter.
- Short intensive workshops (2-3 days) may be conducted in collaboration with the department of measurement and evaluation of the University for expertise.
- It should be required that item writers be certified after training to participate in item development panels.
- As well as being mandatory, refresher training should also be institutionalised and not be a one-off activity.

Longitudinal psychometric monitoring programme.

- It is valuable to have your study's results as a baseline. This (or an equivalent peer reviewed study) should be incorporated formally as Year Zero of the monitoring programme by NABTEB.
- Calibration of the instruments should be done every year internally and be communicated to the Federal Ministry of Education and the academic community.
- Anonymised item statistics published on an open-access basis would promote academic research

Citation: Ogiemwonyi,Otasowie & Omorogiuwa, Osamede Kingsley. "Guessing Behaviour of Multiple Choice Test Items: An Item Response Theory Analysis Using the 3pl-Model for 2019–2021 National Business And Technical Certificate Biology Examinations in Nigeria". *Journal of People and Worldviews (JPW),* 2026: pp175-184.

and accountability without revealing the content of items.

References

Adedoyin, O. O., & Mokobi, T. (2013). Using IRT psychometric analysis in examining the quality of junior certificate mathematics multiple choice examination test items. *International Journal of Asian Social Science, 3*(4), 992–1011. https://archive.aessweb.com/index.php/5007/article/view/2471

Ainsworth, A. (2015). Introduction to Item Response Theory. Psy 427 Cal State Northridge, from http://www.esun.edu/ata20315/psy427/Topic08_IntroIRT.ppt.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), Statistical theories of mental test scores (pp. 397–479). Addison-Wesley.

Chiu, T.-W., & Camilli, G. (2013). Comment on 3PL IRT adjustment for guessing. Applied Psychological Measurement, 37(1), 76–86. https://doi.org/10.1177/0146621612458958

DeAyala, R. J. (2009). The theory and practice of item response theory. Guilford Press.

Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Lawrence Erlbaum Associates.

Leong, S. H., Ong, S. L., & Zuhairi, W. W. (2013). Moderating ability of item response theory through prior guessing parameter. *Journal of Mathematics and Statistics, 9*(4), 277–282. https://doi.org/10.3844/jmssp.2013.277.282

Liao, W.-W., Ho, R.-G., Yen, Y.-C., & Cheng, H.-C. (2012). The four-parameter logistic item response theory model as a robust method of estimating ability despite aberrant responses. Social Behavior and Personality: *An International Journal*, 40(10), 1679–1694. https://doi.org/10.2224/sbp.2012.40.10.1679

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Lawrence Erlbaum Associates.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Addison-Wesley.

NABTEB. (2017). NABTEB annual statistical digest. National Business and Technical Examinations Board.

Obinne, A. D. E. (2011). A psychometric analysis of two major examinations in Nigeria: Standard error of measurement. International Journal of Education Science, 3(2), 137–144.

Obinne, A. D. E. (2012). Item response theory analysis of NECO Biology test items. *Journal of Educational Assessment in Africa*, 7(1), 23–36.

Omoroguiwa, O. K. (2009). An empirical comparison of the classical test theory and item response theory in the selection of physics achievement test items [Unpublished doctoral dissertation]. University of Benin, Benin City, Nigeria.

Swist, J. (2015). Practical item response theory: Translating measurement research into assessment practice. Pearson Education.

Winsteps. (2022). Winsteps Rasch measurement computer program (Version 3.75.0). Winsteps.com.

Yu, C.H. (2013). A simple guide to Item Response Theory(IRT) and Rasch Modeling. Retrieved May 18, 2026 from http://www.creative-wisdom.com

Citation: Ogiemwonyi,Otasowie & Omorogiuwa, Osamede Kingsley. "Guessing Behaviour of Multiple Choice Test Items: An Item Response Theory Analysis Using the 3pl-Model for 2019–2021 National Business And Technical Certificate Biology Examinations in Nigeria". *Journal of People and Worldviews (JPW),* 2026: pp175-184.