East African Journal of Education and Social Sciences
EAJESS January –February 2024, Vol. 5, No. 1, pp. 19-26.
ISSN: 2714-2132 (Online), 2714-2183 (Print). Published by G-Card
DOI: https://doi.org/10.46606/eajess2024v05i01.0345.

Psychometric Assessment of Botswana's 2018 Junior Certificate Agriculture Science Items through Item Response Theory

***Tapela Bulala**
ORCiD: https://orcid.org/0000-0003-4084-1501
Agricultural Education and Extension, Botswana University of Agriculture and Natural Resources
Email: tabulala@buan.ac.bw

Prof. Eric Anane
ORCiD: https://orcid.org/0000-0001-8505-8903
Department of Educational Foundations University of Cape Coast, Ghana
Email: eanane@ucc.edu.gh

Andrews Cobbinah, PhD
ORCiD: https://orcid.org/0000-0002-5315-659X
Department of Educational Foundations, University of Cape Coast, Ghana
Email: andrews.cobbinah@ucc.edu.gh

s

***Corresponding Author:** tabulala@buan.ac.bw

Abstract: This study aimed to assess the psychometric properties of the Junior Certificate Examination (JCE) agricultural science multiple-choice items in Botswana, using the 3-parameter logistic model (3PLM). Adopting a quantitative descriptive research design, the study analyzed responses from 40,343 examinees who took the 40 multiple-choice items in the 2018 JCE agricultural science examination. The study utilized secondary data from Botswana Examination Council (BEC) records. They also obtained the ethical clearance from the Institutional Review Board of the University Cape Coast. The analysis, conducted using 3PLM within the R-studio environment revealed that 15% of items exhibited poor discrimination indices, 10% had difficulty and 12.5% were prone to guessing. There was an identification of Key anomalies in specific items, warranting careful consideration. Approximately 90% of the multiple-choice items fell within the good difficulty level range but 10% appeared to be poor difficulty. For item discrimination, up to 85% were within the acceptable range while 15% were outside. The study recommends a thorough revision or removal of items with poor parameter estimates, emphasizing continuous monitoring and collaboration between the Ministry of Basic Education and test constructors for ongoing improvement. Additionally, the study recommends stringent review of items with extreme difficulty indices to maintain a balanced assessment.

Keywords: Multiple-choice; item response theory; item difficulty; item discrimination; pseudo guessing.

How to cite: Bulala, T., Anane, E. and Cobbinah, A. (2024). Psychometric Assessment of Botswana's 2018 Junior Certificate Agriculture Science Items through Item Response Theory. East African Journal of Education and Social Sciences 5(1)19-26. **Doi:** https://doi.org/10.46606/eajess2024v05i01.0345.

Introduction

Testing is one critical component through which a society views the effectiveness and quality of the educational system (Medina et al., 2019). Not surprisingly, national examinations serve as an important milestone in the education journey of students. Thus, in basic education, national high-stakes examinations are at the center of measuring the quality of education. Scores emanating from these national public examinations form the basis for decision-making such as placement, repeating a

class, progression and even determining a career path (Osei-Mensah, 2012).

It is against this backdrop that countries and at times regions have set up examination bodies such as Botswana Examination Council (BEC) and West African Examinations Council (WAEC), respectively. In Botswana, BEC was established by an act of parliament in 2002 with a mandate is to provide a credible and responsive assessment of learners. To achieve its mandate, BEC develop and deliver examinations that measure learners' abilities.

According to Botswana Statistics (2018), schools systems still record low pass rates of below 50% from Junior Secondary School to Senior Secondary against the national pass rate set at 70%. This is an indication that Junior Secondary Schools have not been effective in preparing learners who complete to qualify for Senior Secondary School and vocational schools as intended.

A study by Enwereji et al. (2017) in Botswana's Southern Region revealed that teachers cited government challenges, including insufficient support, as demotivating factors that affect their commitment to learners. Additionally, a noteworthy finding was that some teachers tended to focus solely on high-performing learners, neglecting low performing ones. This study unlike the current study focused on secondary factors at the exclusion of the quality of instrument used to measure student outcomes, which the test items.

Suping (2022) reported that the decline in performance is attributed to poor teaching-learning conditions and that is kept hidden by the political motivated education reforms. A study by Rasetimela (2016) reported various factors that contribute to academic performance. These factors include medium of instruction, availability of teaching and learning materials, adequacy of infrastructure, learner-teacher ratio, quality of school libraries, teacher motivation and qualifications of both teachers and learners. Additionally, the study recognized learner discipline as a crucial factor that influences educational outcomes.

There are few studies (Akanwa et al., 2020; Orheruata & Uyigue (2018) on how construction of test items contributes to the performance of students in agriculture science. The studies have reported high level of items with flawed item parameters. Notwithstanding, these studies were carried out in West African countries, which may be different from Botswana. Since agriculture is among the compulsory subjects for students to pass at the junior secondary education in Botswana, there is a need to assess the quality of instruments used to determine the progression of students. This will help to determine if the instruments accurately measures students' abilities.

Nenty (2004) reported that responsible persons can judge the quality of test items based either Classical Test Theory (CTT) or Item Response Theory (IRT). Item response theory makes use of the pattern of replies to all the test items to simulate a test taker's skill and the likelihood that they would answer items correctly. Rather than the raw test results, the focus of IRT is on an examinee's accuracy on a given item (Ayanwale, 2019). Based on the pattern of item responses, the item-pattern scoring method yields a maximum probability trait estimate (Eaton, et al., 2019; Bichi & Talib, 2018). More crucially, the test's item specifications must be considered to accurately gauge the examinee's ability based on his or her reaction to a specific test item. The theory is based on the assumption that examinees with high latent traits have high probability of accurately answering items (Zanon et al., 2016).

Conversely, according to Hambleton et al. (1991), in (CTT), the examinee's test score would be the total of the scores they obtained on all the test questions. As per Ayanwale, et al. (2018), this is known as number-correct scoring. Based on raw scores, the total number of correctly answered items generates maximum likelihood trait estimations. Examinees using this method receive the same scale score regardless of the number of correctly answered items and the degree of difficulty or discrimination. Consequently, the parameter estimates are unstable and change with the ability of examinees. Lord (1980) reported that even though there are three well-known models, the two-parameter model and the one-parameter model are both included in the three-parameter model. The model is most suitable for items that are prone to guessing (Effiom, 2021). This makes the three-parameter model most suitable for the current study that deals with multiple-choice items, which are vulnerable to guessing.

The consistently low pass rate in Agriculture Science in the BEC examination makes it harder to reach the Sustainable Development Goal 4 that addresses quality education. Botswana Examination Council (2020) reported that candidates who obtained

credit passes in agriculture during the years 2018, 2019 and 2020 were 42.10%, 46.05% and 45.15%, respectively. Therefore, investigation is necessary into the role that multiple-choice items, which account for 35% of the overall scores, play in the decline. Thus, it is crucial, to carry out this research to determine whether the JCE agriculture multiple-choice's item parameters (difficulty, discrimination, and guessing) lie within acceptable range.

Obinne (2012) argued that guessing involves offering an answer or making judgment without perfect certainty of all the facts. Guessing is a standard test-taking strategy for individuals taking multiple-choice assessments. This approach allows examinees to receive credit for an item, even when their knowledge of the subject matter is limited. When test scores rely solely on the number of correctly answered questions, random guessing can increase the likelihood of achieving a higher overall score. The guessing index measures the likelihood of an examinee answering an item correctly by chance. A typical cut-off point for the guessing index is often set at 0.25 for four option items as examinees have a one in four probability (1/*k*) of getting the item correct. If the guessing index for an item is below this threshold, the item is not likely to be guessed correctly and it contributes meaningfully to the quality of the assessment.

Item difficulty, often termed the index of difficulty, indicates the ease or difficulty of a specific test item (Jumran, 2015). Baker's (2001) suggested range for the difficulty index is from negative 3 to positive 3. In this context, if an item has a difficulty index falling within this range, it implies that the item's difficulty level is well distributed, encompassing a wide range of examinee abilities. Items with extreme difficulty indices, either close to negative 3 or positive 3, may be considered very challenging or very easy, respectively. Indices beyond this threshold is problematic.

Koçdar et al. (2017) added that the discrimination index of an item reflects its ability to differentiate between high and low-scoring learners' ranges. The analysis of each item, considering both difficulty and discrimination indices, offers valuable insights into the learning outcomes of students. This approach enables instructors to identify and rectify problematic items, contributing to effective assessment and instructional improvement. A discrimination index, according to Baker (2001), is good between 0.5 and 2.0 but runs from $-\infty$ to $+\infty$. If an item exhibits a discrimination index meeting or exceeding this threshold, the item differentiates between high and low performers effectively. High discrimination indices are desirable as they contribute to the precision of the test in distinguishing between students with varying levels of ability.

Methodology

Design

This study used the descriptive design. This type of research seeks to explain events, symptoms or specific groups of people based on numerical data from tests. In this study, the researchers explained characteristics of test items according to the nature of the agricultural questions used in the 2018 Botswana Junior Certificate Examination. This design is relevant to this study as it allows the analysis of available data to describe the JCE agriculture science multiple-choice items.

Population and Sampling

The population of this study involved the scores of all 40 items attempted by all examinees who wrote the Junior Certificate Examination in 2018 in Botswana. The gender distribution is as follows: 20,096 male students, constituting 49.8% of the total and 20,247 female students, making up 50.2%. In total, there were 40,343 students in the dataset. Those students had completed the three-year agriculture curriculum at the time of data collection. The study dealt with all responses to individual multiple-choice items.

Sources of Data

The researchers obtained ethical clearance from the Institutional Review Board (IRB) of the University of Cape Coast (UCC). The Department of Education and Psychology at UCC also issued a letter of introduction to the field. The researchers then attached the letter to the IRB ethical clearance, to apply for permission to access the BEC data. The researchers collected student-by-item responses as text data files on the Microsoft Excel platform from Botswana Examination Council office.

Ethical Considerations

With BEC being an institutional body, the researchers awaited board approval before initiating data gathering. Acknowledging BEC's confidentiality policy, the researchers emphasized the necessity for de-identifying data to exclude student and school names. This cautious approach aimed to safeguard the privacy of student scores, reflecting a

commitment to upholding the dignity of the subjects involved. The researchers considered this stringent privacy protection sufficient to uphold human rights. Given the use of secondary data and the absence of direct participation or contact with examinees, consent was retrospective. The obtained permissions served as a gateway for the researcher to access Ministry of Basic Education and BEC officials for the release of data.

Statistical Treatment of the Data

The researchers subjected the 40 items to IRT psychometric analysis. They processed the data in the R-studio environment and analysed it using the 3 Parameter Logistic Model (3PLM The researchers reported the characteristics of the JCE agriculture multiple-choice items based on item difficulty, discrimination and guessing parameters.

Results and Discussion

This section presents the results of the psychometric properties for the 40-item agricultural test. The section presents the guessing parameters, the difficulty index and the discrimination index. Discrimination, known in IRT as location, is represented by *a;* difficulty or slope is denoted by *b* whiles guessing is denoted as *c* as reflected in Table 1.

For the guessing parameters, the cut off points considered ideal is less than or equal to 0.25. If the guessing index for an item is below this threshold, candidates are likely not to be guess the item correctly. If it is above this threshold, then it is more susceptible to guessing (Baker, 2001). As presented in Table 1, the guessing parameter ranged from 0 to .423, indicating that some of the items were susceptible to guessing. Comparing these guessing parameters with the acceptable random guessing factor of .25 for four option multiple-choice items, five items (Q14, Q16, Q24, Q31, Q38) were found to be susceptible to guessing with guessing parameter values ranging from .272 to .423. When it comes to the guessing factor, up to 87.5 percent of the multiple-choice questions were less likely to be guessed and 12.5 % was prone to guessing.

Items with extreme difficulty indices, either close to negative 3 or positive 3, may be considered very challenging or very easy, respectively. Items closer to +3 are very difficult and can only be passed by individuals with higher abilities while items with indices closer to -3 are very easy and can be scored by both individuals with high ability and those with low ability. Most of the items were found to have indices closer to -3. This suggests that the items were less challenging and could be passed by individuals with very low abilities. Indices beyond this threshold are problematic. Items Q2 (*b*=-4.449), Q7 (*b*=18.339), Q25 (*b*=14.076) and Q32 (*b*=-3.888) had difficult threshold values beyond the acceptable -3 to +3 range (Baker, 2001). This anomaly suggests that these four items were problematic. The same need to be checked, deleted, or revised. Based on the 3-parameter model, up to 90% of the questions were in the good difficulty level range, and 10% are in the poor difficulty level range.

s

Items with extreme difficulty indices, either close to negative 3 or positive 3, may be considered very challenging or very easy, respectively. Items closer to +3 are very difficult and can only be passed by individuals with higher abilities while items with indices closer to -3 are very easy and can be scored by both individuals with high ability and those with low ability. Most of the items were found to have indices closer to -3. This suggests that the items were less challenging and could be passed by individuals with very low abilities. Indices beyond this threshold are problematic. Items Q2 (*b*=-4.449), Q7 (*b*=18.339), Q25 (*b*=14.076) and Q32 (*b*=-3.888) had difficult threshold values beyond the acceptable -3 to +3 range (Baker, 2001). This anomaly suggests that these four items were problematic. The same need to be checked, deleted, or revised. Based on the 3-parameter model, up to 90% of the questions were in the good difficulty level range, and 10% are in the poor difficulty level range.

Discussion of Findings

The findings show that although there are items that were functional and they accurately measured the construct of student ability, there were challenges with some of the items. Between 12.5% and 15% of the items were prone to guessing. Expectedly, the tendency of students to guess on a multiple test item is very high since multiple-choice items offer learners a chance to obtain a correct mark for items they have little knowledge about. This phenomenon is more likely to occur in high stakes testing environments where failure to pass the examination has dire consequences. Thus, it is not surprising that this study reported some level of guessing, although relatively minimal. Similarly, previous studies demonstrated the presence of guessing factor on multiple-choice tests in some high-stakes examinations such as Senior School Certificate Examinations (SSCE), National

Examination Council (NECO) and West African Examination Council (WAEC) (Asuquo et al., 2022; Jimoh et al., 2020; Ogunbamowo et al., 2019). Whereas some of the earlier studies reported high guessing levels, others reported minimal factors. Jimoh et al. (2020), for example, reported 36% guessing on the 2016 WASSCE Economics multiple-choice test and 35% on the NECO Economics test. The discrepancies in the findings may stem from variations in the subjects/course (e.g., Economics, Mathematics) studied.

Table 1: Item Parameters of 3PL for the Year 2018

No.	Guessing (c)	Difficulty (b)	Discrimination (a)	P(x=1|z=0)
Q1	0.002	-0.213	0.751	0.541
Q2	0.018	-4.449	-0.097	0.405
Q3	0.010	-0.638	0.612	0.600
Q4	0.000	-0.797	1.224	0.726
Q5	0.000	-0.115	0.648	0.519
Q6	0.009	-0.077	0.959	0.523
Q7	0.078	18.339	0.054	0.327
Q8	0.125	2.419	0.990	0.199
Q9	0.207	1.709	2.292	0.222
Q10	0.147	-0.041	0.932	0.582
Q11	0.000	-1.338	0.489	0.658
Q12	0.211	1.058	1.841	0.310
Q13	0.000	-1.193	1.950	0.911
Q14	0.312	1.129	2.233	0.363
Q15	0.000	-0.560	0.994	0.636
Q16	0.282	2.488	1.683	0.293
Q17	0.113	1.018	1.517	0.269
Q18	0.235	0.311	2.063	0.499
Q19	0.188	-0.314	1.684	0.699
Q20	0.015	0.107	0.096	0.505
Q21	0.010	1.528	0.458	0.338
Q22	0.099	1.948	2.472	0.107
Q23	0.059	-1.286	1.151	0.826
Q24	0.423	-0.332	1.709	0.791
Q25	0.239	14.076	0.112	0.369
Q26	0.200	0.671	1.891	0.376
Q27	0.000	-1.649	2.041	0.967
Q28	0.214	2.769	2.073	0.216
Q29	0.137	-0.038	1.800	0.583
Q30	0.100	2.039	0.570	0.314
Q31	0.414	0.540	1.160	0.618
Q32	0.000	-3.888	0.265	0.737
Q33	0.002	-2.317	-0.134	0.424
Q34	0.081	-0.651	1.733	0.775
Q35	0.150	0.843	1.267	0.367
Q36	0.177	1.995	1.331	0.231
Q37	0.187	1.804	1.151	0.278
Q38	0.270	1.056	1.192	0.432
Q39	0.157	0.692	1.559	0.371
Q40	0.004	-0.544	0.762	0.604
Cut-off	$(c) \leq 0.25$	(b) -3 – +3	$(a) \geq 0.5$	

Another important finding is that the items had varied levels of difficulty. Whereas some items functioned at the lower ability level of the examinees, others functioned best at the higher ability level. However, four items showed a dysfunctional difficulty index, indicating that these items were faulty and should be removed/modified. It sounds good that the items had varying difficulty levels and thus, items that were dysfunctional were few. This understanding centers on the fact that item difficulty may contribute significantly to the variations in students' performance (Cobbinah et al., 2022). The findings appear to contradict observations of previous studies that examined the difficult indicators of national-wide exams revealed a moderate level of difficulty (Deborah et al., 2020; Ibrahim, 2023; Onah & Jiwueze, 2015). Unfortunately, these studies focused on the overall difficulty, ignoring the specific items and thus, there is little clarity regarding the difficulty parameter details of the items.

The results further reveal that 15% to 20% of the items exhibited poor discrimination. While the number of items with poor discrimination index was quite low, these items are still of concern since all items sum up to the conceptualization/measurement of the student's ability construct. Such items have limited functionality and utility in terms of scaling examinees into those with sufficient mastery over the content area and those without (Nitko, 2001; Quansah & Cobbinah, 2021). A notable concern from this study's results is that about 5% to 7.5% showed negative discrimination indices. A key factor that could explain this occurrence is the possibility that the agriculture tasks covered complex materials, such that it would be probable for examinees to choose the correct answer without having mastery of the task (Matlock-Hetzel, 2011). From a broader perspective, poor distractor functioning can contribute to a poor discrimination index or even negative discrimination. Even though this study did not analyze the effectiveness of the distractors for the items with poor discrimination, the presence of weak distractors could be a possible cause of the dysfunction of the items flagged as having poor discrimination.

Conclusions and Recommendations

Conclusions

In conclusion, there is a notable susceptibility to guessing in several items being investigated. This indicates that the test might not have accurately measured the abilities of students, and the distractors did not serve their intended purpose effectively. Consequently, the learners achieved the correct answers with relative ease. One plausible explanation for this phenomenon is the presence of multiple clues within the items, facilitating informed guessing by students lacking in-depth knowledge on the subject matter. The observed susceptibility to guessing raises concerns about the test's ability to provide a fair and precise assessment of students' capabilities.

Furthermore, approximately 90% of the questions fall within the good difficulty level range. However, about 10% of the questions appeared in the poor difficulty level range. This distribution implies that a majority of the questions were effectively challenging students at an appropriate level, allowing for a meaningful assessment of their abilities. Nonetheless, the presence of a notable percentage in the poor difficulty level range raises concerns about the potential impact on the test's ability to accurately gauge students' capabilities. It emphasizes the need for further examination and refinement to ensure a fair and precise assessment across all difficulty levels.

Finally, the test demonstrates a commendable test construction, with approximately 85% of multiple-choice questions exhibiting acceptable levels of item discrimination. This indicates the successful differentiation between students with varying proficiency levels, aligning well with the test's intended purpose. However, the identification of 15% of questions falling outside the acceptable discrimination range highlights a need for careful consideration from test constructors. These questions may not effectively fulfil their role in distinguishing performance, raising concerns about their validity and impact on overall assessment goals. Addressing and refining this subset is crucial for maintaining the test's integrity and ensuring accurate measurement of students' abilities.

Recommendations

The study recommends that the BEC conduct a thorough review of test items that exhibit susceptibility to guessing if such items are banked and are to be reused in future. Collaboration with experienced educators and subject matter experts is crucial in revising these items to align with the intended difficulty level and minimize the presence of clues facilitating informed guessing. Implementing a rigorous item-writing protocol is

essential, emphasizing the creation of distractors that effectively challenge students' knowledge. There is a need to provide clear guidelines to item writers to minimize cues and clues that may lead to guessing. Pilot testing revised items can validate their effectiveness.

For test constructors, maintaining a systematic approach to regularly reviewing and updating the test bank is essential. This process should include a careful examination of the difficulty distribution, ensuring representation of items that challenge students appropriately. The study recommends collaborating with educational experts to establish best practices in item development, with a focus on alignment with educational objectives.

The Ministry of Basic Education should actively monitor the ongoing development of test items to ensure that the majority maintain acceptable levels of item discrimination. Collaboration with psychometricians and educators is essential to identify and address issues in item construction. There is a need to provide training for item developers (comprises of few teachers selected from schools led by a subject specialist based at BEC) on effective item writing and analysis which can enhance discrimination levels

References

Akanwa, U. N., Ihechu, K. J. P., & Nkwocha, P. C. (2020). Effect of language manipulation on differential item functioning of agricultural science multiple choice test items. *Journal Of The Nigerian Academy Of Education*, *14*(1), 211-223.

Ayanwale, M. A. (2019). *Efficacy of item response theory in the validation and score ranking of dichotomous and polytomous response mathematics achievement tests in Osun State, Nigeria*. (Doctoral dissertation, University of Ibadan).

Ayanwale, M. A., Adeleke, J. O., & Mamadelo, T. I. (2018). An assessment of item statistics estimates of basic education certificate examination through classical test theory and item response theory approach. *International Journal of Educational Research Review*, *3*(4), 55-67.

Asuquo, U. M., Ekon, J. S., & Chigbu, N. N. (2022). Guessing parameters and multiple-choice test items of 2016 - 2017 National Examination Council (NECO) Mathematics Examination Instruments in Calabar Education Zone of Cross River State. *J Adv Educ Philos, 6*(9), 492-497.

Baker, F. B. (2001). *The basics of item response theory*. http://ericae. net/irt/baker.

Bichi, A. A., & Talib, R. (2018). Item Response Theory: An Introduction to Latent Trait Models to Test and Item Development. International *Journal of Evaluation and Research in Education, 7*(2), *142-151.*

Botswana Examination Council [BEC] (2020). *Summary of results report. Retrieved from: http://www.bec.co.bw/pastresults/2020/2020_bgcse_book_version.pdf/view.*

Botswana Statistics (2018). Botswana demographic survey report 2017. Retrieved from: https://statsbots.org.bw/.

Cobbinah, A., Annan-Brew, R. K., & Quansah, F. (2022). Item Difficulty as a Source of Variability in Student Achievement in the West African Secondary School Certificate Examination (WASSCE): Application of Generalizability Theory. *J Adv Educ Philos*, *6*(3), 136-142.

Deborah, A. C., Temitope, B., & Peter, A. O. (2020). Comparism of Item Difficulty and Discrimination of Pre And Post University Entrance Examinations in Nigeria. *The Universal Academic Research Journal*, *3*(1), 1-9.

Eaton, P., Johnson, K., Frank, B., & Willoughby, S. (2019). Classical test theory and item response theory comparison of the brief electricity and magnetism assessment and the conceptual survey of electricity and magnetism. *Physical Review Physics Education Research*, *15*(1), 010102.

Effiom, A. P. (2021). Test fairness and assessment of differential item functioning of mathematics achievement test for senior secondary students in Cross River state, Nigeria using item response theory. *Global Journal of Educational Research*, *20*(1), 55-62.

Enwereji, C., Mbukanma, I., & Chukwuere, J. (2017). Factors affecting low performance of junior learners in a developing country: Southern region of Botswana. *The New Educational Review*, *47*(1), 81-93.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.

Ibrahim, B. M. (2023). Psychometric properties analysis of mathematics West African Senior School Certificate Examination in Dala Education Zone, Kano State, Nigeria (2020-2022). *International Journal of Advanced Academic Research, 9*(1), 199-213.f

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

Jumran, M. (2015). Measuring Test Difficulty Level and Discriminating Power. Makassar State University.

Koçdar, S., Karadağ, N., & ŞAHİN, M. (2017). Analysis of the Difficulty and Discrimination Indices of Multiple-Choice Questions According to Cognitive Levels in an Open and Distance Learning Context. *Turkish Online Journal of Educational Technology-TOJET*, 15(4), 16-24.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.

Matlock-Hetzel, S. (2011). *Basic concepts in item and test analysis*. A paper presented at the annual meeting of the Southwest Educational Research Association, Austin. Retrieve November 21, 2020. Available online at http://ericae.net/ft/tamu.Espy.htm. 1997.

Medina, J. E. C., Poveda, D., & Ortíz, D. A. R. (2019). Education Quality: Reflections On Its Evaluation Through Standardized Testing. *Saber, ciencia y libertad*, *14*(2), 247-255.

Nenty, H. J. (2004). From CTT to IRT: An introduction to a desirable transition. In O. A.Afemikhe & J. C. Adewale (Eds.), *Issues in educational measurement and evaluation in Nigeria (in honour of Professor Wole Falayajo)* (pp.371 383). Ibadan: Institute of Education, University of Ibadan, Nigeria.

Nitko, A. J. (2001). Conceptual frameworks to accommodate the validation of rapidly changing requirements for assessments. *Curriculum and assessment*, *1*, 143-163.

Obinne, A. D. E. (2012). Using IRT in Determining Test Item Prone to Guessing. *World Journal of Education*, *2*(1), 91-95.

Ogunbamowo, A. O., Adediwura, A. A. & Diyan, R. O. (2019). Psychometric properties of 2017 West African Examination Council and National Examinations Council's Economic Senior School Certificate Examination Items. *Journal of Advances in Education and Philosophy, 3(1):* 9-18*.*

Onah, F. E., & jiwueze, c. c. (2015). Psychometric Properties of West African Examination Council (WAEC) O'level Multiple Chioce Past Question Paper on Economics for the year 2013. *Journal of Science & Computer Education*, *3*(3).

Orheruata, M. U., & Uyigue, A. V. (2018). Investigating flaw items of WASSCE Agricultural Science multiple choice items. *African Journal of Interdisciplinary Studies*, *11*(1), 1-7.

Osei-Mensah, F. (2012). *Factors that influence the performance in general knowledge in Art of senior high school students in Abura Asebu Kwamankese district in the Central Region* (Doctoral dissertation).

Quansah, F., & Cobbinah, A. (2021). Equivalence of Parallel Tests in a Basic Statistics Course in Higher Education Using Classical Measurement Theory. *Canadian Journal of Educational and Social Studies*, *1*(2), 13-28.

Rasetimela, M. (2016). Factors that influence the performance of students in botswana primary schools. *IOSR Journal Of Humanities And Social Science (IOSR-JHSS)*, *9*(5), 40-53.

Suping, K. (2022). Political spectacle and the decline of public education in Botswana. *Journal of Asian and African Studies*, 59(2), 443-455.

Zanon, C., Hutz, C. S., Yoo, H. H., & Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica*, *29*.