ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)
The African Journal of Behavioural and Scale Development Research
AJB-SDR Vol. 6, No 2, 2024

Comparability of item parameter indices of 2019 senior school certificate Physics Multiple-choice examinations among Osun State Secondary school students

Oluwasola Damilola Owoeye Ph. D.[1], Prof. Alaba Adeyemi. Adediwura[2]
Department of Educational Foundations and Counselling, Obafemi Awolowo University, Ile-Ife, Nigeria
Department of Educational Foundations and Counselling, Obafemi Awolowo University, Ile-Ife, Nigeria

Abstract
The study assessed the difference in average item parameter indices (item difficulty and students' ability) in the 2019 Physics multiple-choice examination in the Senior School Certificate Examination (SSCE) among secondary school students in Osun State. It also examined the standard errors of measurement in both examinations. A survey research design was adopted to collect data from participants in an examination condition. The study population comprised 17,784 Senior Secondary School III students across 1,494 schools (1,100 private and 394 public) in the 2022/2023 session. A total of 1,200 students were selected using a multistage sampling procedure. For NECO, the 3-parameter logistic model produced the lowest values for all information criteria (AIC = 82986.44, SABIC = 83330.90, HQ = 83331.57, BIC = 83902.65). For WAEC, the 4-parameter logistic model had the lowest (AIC = 70591.57, SABIC = 70974.30, HQ = 70975.04, BIC = 71609.58). WAEC items were more difficult (mean = 0.62, STD = 1.246) than NECO (mean = 0.52, STD = 1.553). Examinees performed slightly better in WAEC (mean = 0.004, STD = 0.935) than in NECO (mean = 0.002, STD = 0.948). The study recommended continuous monitoring and improvement of test validity and reliability by both examination bodies.

Introduction
Public examinations are conducted to determine the abilities of learners after they have come in contact with some volume of learning and learning activities. In Nigeria, one of the public examinations that learners take is the West African Examination Council (WAEC), which means they get the Senior School Certificate (SSC). Another body that offers public examinations to learners in Nigeria is the National Examination Council (NECO). The questions that are developed in these examinations are in tune with the National Educational Research and Development Council (NERDC). Based on the effectiveness of these examination bodies, they aid in testing all that students have learned and their academic skills. In the same vein, when constructing the tests, different test construction principles are engaged, such as validity, reliability, fairness, and effectiveness in assessing the intended learning outcomes, which align with clearly defined curriculum objectives to ensure content relevance and coverage. In all of these, there is always the need to investigate test difficulty, which is a major focus in the present study.

Physics has been a required subject for all students in Nigeria ever since the Senior Secondary Certificate Examination (SSCE) was introduced. The objective was to encourage the expansion of the country's technological infrastructure. One of the most significant scientific fields is physics. The two sections that make up the physics section of the Secondary School Certificate Examination (SSCE) are called paper I and paper II. Paper I is the exam's practical component, and Paper II is broken up into two subtests: multiple-choice questions and an essay. This means that students studying physics will need to pass a variety of tests to receive their Senior Secondary Certificate (SSCE). Since physics offers the foundation for the development of higher-order thinking skills, students who are most interested in the subject stand to benefit from a rigorous physics education.

According to Jegede and Adebayo, (2013) teaching physics in secondary schools motivates students to seek higher education in science-related subjects and helps them acquire the information and abilities needed for scientific studies. This helps students acquire basic scientific concepts, which in turn help create new technologies that are intended to enhance people's quality of life. the fact that physics is essential to improving a country's technical

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

capabilities. Young scientists who want to further their studies in the field have several options on where to concentrate their efforts. Astrophysics, particle physics, geophysics, engineering physics, medical physics, nuclear physics, biotechnology, nanotechnology, aerospace dynamics, atomic and laser physics, atmospheric, oceanic, and planetary physics, and more are some of the numerous fields of physics that fall under this general heading. Applicants must frequently have at least a credit level pass in Physics at the SSCE level to be admitted to higher education institutions. It is anticipated that students who sign up for the course will be more enthusiastic and perform better academically. This may be due to the common perception that physics is the foundation of all other scientific fields, which has significantly facilitated the global adoption of innovative technologies in both developed and developing nations. This perspective may be explained by the fact that physics is widely acknowledged as the foundational field on which all technological developments are based.

The comparability of item parameter indices of Physics on the 2019 SSCE, with a focus on the multiple-choice tests administered by WAEC and NECO, is investigated in this study, which is of utmost importance. The goal of this research is to assess the comparison that exists between item difficulty and students' ability based on Physics examination items.

Item Response Theory (IRT) provides a strong statistical framework for assessing the dependability of item responses in psychological and educational tests (as cited in Osterlind 2012). The approach is based on a cognitive model that enables accurate measurement of latent traits like ability and aligns assessment with mental processes. The possibility for students to also provide wrong or right answers to some given questions. In another study, Magno (2009) asserted that response accuracy is also measured, which is an important aspect of the Item Characteristic Curve (ICC). This is why Ojerinde et al. (2011) asserted that IRT parameters fall into two categories: those related to the examinee and those linked to the test items. Each examinee is assumed to possess a level of underlying ability, represented by the Greek letter θ, which influences their probability of answering items correctly. Under IRT, this probability is denoted as $P(\theta)$. Items are characterized by three indices: discrimination (a), difficulty (b), and guessing (c), which are the core parameters of the three-parameter logistic (3PL) model used in Item Response Theory (as cited in Hambleton et al. 1991).

Nguyen et al. (2014) identified four fundamental assumptions underlying IRT. It is assumed that it is easy for students to answer questions correctly when they understand the questions and know the answers, as this forms a basic principle in test development and validity (as cited in Haladyna and Rodriguez 2013). Aside from this, there are unidimensionality, local independence, and item response function. For the effective application of latent trait models, all these assumptions play important roles. Unidimensionality posits that test items measure a single domain of knowledge, allowing for the prediction of responses based on the examinee's ability. Local independence illustrates that getting a key right cannot be affected by the responses of other items in a given examination or test, which is conditional on the examinee's ability. This does not invalidate the correlations among items but asserts that performance on each item is independently influenced by students' abilities. The local independence guarantees that observed item responses rely solely on the latent trait as asserted by Liu and Maydeu-Olivares (2012). The item response function also referred to as the item characteristic curve implies the correlation that exists between an item's success and capabilities measured in an examination or test. This curve serves as the fundamentals of IRT, enhancing the characterization of the correlation between students' ability and item response, as explained by Baker and Kim (2017).

IRT gives numerous advantages, which include its approach to reliability and measurement error through the item information function; this assesses the effectiveness of items across different ability levels (as cited in Nguyen et al. 2014). These item information functions help in choosing optimal items during test construction. One of the prominent benefits of IRT is the

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

invariance of item parameters, seeking and enhancing the results to remain consistent across groups with varying or different abilities. This approach allows the development of a standardized measurement scale that applies to different groups in ensuring the comparability of scores derived from varying sets of items tailored to particular ability levels, as stated by Magno (2009). Given IRT's advantages in guaranteeing score comparability and reliability, it becomes especially pertinent when assessing the alleged discrepancies and objections to the validity and quality of tests administered by different examination bodies.

Researchers have different perspectives as to the quality and validity of tests conducted by these examination bodies. Comparative analyses often focus on aspects such as content, difficulty, and psychometric properties, exploring statistical characteristics of examination results, student performance, and the extent to which scores predict academic success. However, for some time, examination bodies in Nigeria have faced criticism from certain institutions and employers. A preference has emerged for candidates with credit to pass senior school certificate examinations, largely due to misconceptions about the quality of NECO examinations.

Despite these concerns, there are limited research on a direct comparison of item difficulty and students' abilities between the two examination items. This study explores students' abilities between the two examination items based on some Physics questions (multiple-choice) among some students in secondary schools in Osun State, focusing on item difficulty and student ability parameters.

The purpose of this study is to investigate the differences in average item difficulty parameters using WAEC and NECO that were conducted in 2019 and comparatively explore two Physics examination items based on item parameters (item difficulty and students' ability). Furthermore, examine the standard error of measurement in the examinations.

Research Questions

i. What are the differences in average item difficulty parameters of the 2019 SSCE (WAEC and NECO) Physics multiple-choice items among Osun State secondary school students?
ii. What is the comparison between item difficulty and students' ability based on Physics examination items?
iii. What are standard errors of measurements that exist in the WAEC and NECO?

METHOD

In the study, the survey research design is adopted as it helps in the gathering of information from participants in their environment. All senior secondary school students from all the senatorial districts in Osun State made up the study population. Osun State has 1,494 senior secondary schools of which 1,100 and 394 schools are private and public respectively. The population of Senior Secondary Three (SSS III) students consisted of a total number of 17,784 as of the 2022/2023 session. From the population, a total of 1200 SSS III were selected and engaged in the study with the aid of the multistage sampling procedure. Since there are three senatorial districts in Osun State, through the aid of the simple random sampling technique, the local government areas selected were four. Spanning from the local government, two secondary schools each were randomly selected making 24 secondary schools in all. Also, using a simple random sampling technique, 50 SS III students were selected from each school.

If a school with less than 50 candidates in SS III were selected, such a school was dropped and a fresh selection was made until the required 24 schools were selected. The research instruments used for data collection in this study were adopted from 2019 (WAEC and NECO) Physics multiple-choice examinations consisting of 50 and 60 items for WAEC and NECO respectively and were administered to the same group of students. Data were collected for the study after the students gave responses to the items in the adopted 2019 WAEC and NECO Physics multiple-choice examinations. Data collected

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

were analysed using the Multidimensional Item Response Theory (MIRT) of the R language package environment for statistical packages.

Result And Discussion

Result

Research Question One: What are the differences in average item difficulty parameters of the 2019 SSCE (WAEC and NECO) Physics multiple-choice items among Osun State secondary school students??

As a way of providing an answer to the research question above, the responses of the sampled examinees were subjected to two levels of analysis. The first was the model-data fit assessment. The analysis was done to determine the best IRT model that should be used to calibrate the 2019 NECO and 2019 WAEC data respectively. The analytical tools used to conduct the analysis are the MIRT analytical package and the item response theory model. The variety of logistic models provided by item response theory (IRT) is determined by the number of parameters used to describe item characteristics. Item difficulty is taken into consideration in the 1-Parameter Logistic Model (1PL or Rasch model); item discrimination is added in the 2-Parameter Logistic Model (2PL); a guessing parameter is introduced in the 3-Parameter Logistic Model (3PL); and an upper asymptote is incorporated in the 4-Parameter Logistic Model (4PL) to account for the potential for high-ability students to occasionally provide an incorrect response. The following are the results of the analysis.

Table 1: Model-data fit of the 2019 Physics Examination

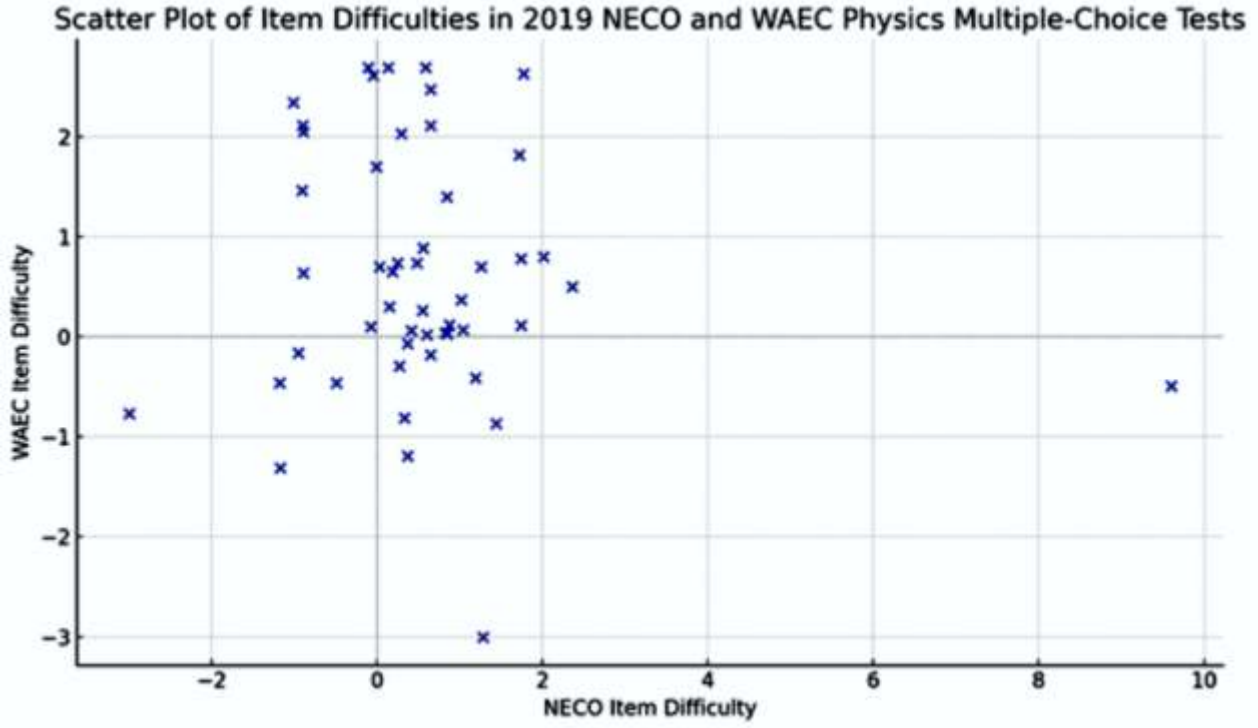
	AIC	SABIC	HQ	BIC
	NECO			
RASCH	88106.29	88223.03	88223.26	88416.79
1PL	88106.30	88223.04	88223.26	88416.79
2PL	85008.01	85237.65	85238.10	85618.82
3PL	82986.44	83330.90	83331.57	83902.65
4PL	83126.37	83585.65	83586.54	84347.98
	WAEC			
RASCH	72623.19	72720.79	72720.98	72882.78
1PL	72623.21	72720.81	72721.00	72882.80
2PL	71241.39	71432.76	71433.13	71750.39
3PL	70931.46	71218.51	71219.06	71694.97
4PL	70591.57	70974.30	70975.04	71609.58

In Table 1 (above), the 2019 NECO Physics multiple-choice test and 2019 WAEC Physics multiple-choice test were tested using the model-data fit assessment. In the table, the 3-parameter logistic model as presented has the lowest values for all of the information criteria (AIC = 82986.44, SABIC = 83330.90, HQ = 83331.57, BIC = 83902.65). Therefore, among the models the 3PL model appears to be the best choice as it offers a good balance between model fit and complexity. However, for the WAEC data, the table shows that the 4-parameter logistic (4PL) model has the lowest values for all of the information criteria (AIC = 70591.57, SABIC = 70974.30, HQ = 70975.04, BIC = 71609.58). Therefore, among the models the 4PL model appears to be the best choice as it offers a good balance between model fit and complexity. Consequently, the NECO test was calibrated with 3PL, while the WAEC test was calibrated with using 4PL. The following figure and table shows the results, a scatter plot comparing item difficulties between the 2019 NECO and WAEC Physics multiple-choice

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

items is presented in Figure 1 below, which also shows the relationship between the difficulty indices across the shared items. The full numerical values for all 60 NECO and 50 WAEC items are provided in Table 2 below, and the scatter plot does not include the 10 additional NECO items that were not included in WAEC because they do not have comparison values.

Figure 1



The scatter plot in figure 1 shows how the item difficulty of 50 common items on the 2019 NECO and WAEC Physics multiple-choice tests relate to one another. In both examinations, each point denotes the degree of difficulty of a specific item. Moderate item difficulty in both tests is indicated by the clustering around the center. Although not exactly aligned, a positive linear trend indicates that items that are challenging in NECO are typically also reasonably challenging in WAEC, and vice versa.

Table 2: 2019 NECO and WAEC multiple-choice tests' Item difficulty

Item	NECO	WAEC	Item	NECO	WAEC
item1	-0.49	-0.46	item32	1.26	0.70
item2	0.82	0.04	item33	0.15	0.30
item3	9.60	-0.49	item34	1.02	0.37
item4	0.65	-0.18	item35	0.55	0.26
item5	1.04	0.07	item36	-0.95	-0.16
item6	1.74	0.11	item37	0.85	0.03
item7	0.87	0.12	item38	1.74	0.78
item8	0.25	0.74	item39	0.19	0.65
item9	2.36	0.50	item40	-3.00	-0.77
item10	-0.08	0.10	item41	1.19	-0.41
item11	-0.05	2.61	item42	-0.89	2.05
item12	-1.18	-0.46	item43	0.37	-1.19

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

item13	0.65	2.47	item44	-1.17	-1.31
item14	0.29	2.03	item45	-0.90	2.11
item15	0.59	2.69	item46	0.37	-0.07
item16	0.84	1.40	item47	-0.11	2.69
item17	0.03	0.70	item48	0.33	-0.81
item18	0.65	2.11	item49	0.27	-0.29
item19	0.56	0.89	item50	1.77	2.63
item20	1.72	1.82	item51	0.04	
item21	1.28	-3.00	item52	0.66	
item22	-1.01	2.34	item53	-1.06	
item23	0.41	0.06	item54	0.95	
item24	0.13	2.69	item55	2.23	
item25	1.44	-0.87	item56	0.27	
item26	-0.91	1.46	item57	-0.53	
item27	0.48	0.74	item58	0.91	
item28	-0.01	1.70	item59	-1.23	
item29	0.60	0.02	item60	1.38	
item30	2.01	0.80	**Mean**	**0.52**	**0.62**
item31	-0.89	0.64	**STD**	**1.553**	**1.246**

In Table 2, 2019 NECO and WAEC multiple-choice test items based on their level of difficulty.

Table 2 shows the difficulty of the 2019 NECO and WAEC multiple-choice test items. More so, WAEC items were more difficult ($\bar{x}_b$ = 0.62, STD=1.246) than the item difficulties of NECO ($\bar{x}_b$ = 0.52, STD =1.553) as shown in the table. The result showed that the WAEC multiple-choice items of 2019 were more difficult than the NECO multiple-choice test of the year. Direct comparison was not possible for the ten extra items in the NECO test that did not have matching items in the WAEC test. Nonetheless, the overall NECO item difficulty statistics (mean = 0.52, STD = 1.553) show its impact. The t-test analysis was used to explore the difference and difficulty observed in the two examinations as presented in Table 3 (below).

Table 3: t-Test Comparison of NECO and WAEC Item Difficulties

Exam	N	Mean	STD	Std. Err Mean	T	Df	P-value
NECO	60	0.52	1.554	0.201	0.373	108	0.71
WAEC	50	0.62	1.246	0.176			

In Table 3, as shown above, the difference and difficulty observed in the two examinations have been presented. In the result, there were no significant differences in the difficulty levels of the examinations (NECO and WAEC) (t (df = 108) = 0.373, $p > 0.05$). This means that there was a similar level of difficulty in the two examinations.

Research Question 2: What is the comparison that exists between item difficulty and students' ability based on Physics examination items?

As a way of providing an answer to the research question above, the independent t-test was carried out using the students' ability estimates, which are revealed in Table 4.

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

Table 4: Ability estimates of examinees in 2019 NECO and WAEC tests

EXAM	N	Mean	STD	Std. Error Mean	t	df	P-value
NECO	1200	0.002	0.948	0.027	0.051	2398	0.959
WAEC	1200	0.004	0.935	0.027			

Table 5 shows the standard error of measurement of students' ability estimates in 2019 NECO and WAEC. This estimate shows the precision with which the ability estimates of the students were estimated. The table shows that the ability estimates of the students" standard error of measurement in NECO were better (mean = 0.25, STD = 0.065) than WAEC (mean = 0.33, STD = 0.126). Using the independent sample t-test, there was a significant difference in the standard error of measurement of the ability estimates in both examinations ($t (df = 2398) = 19.493$, $p < 0.05$). The result showed that NEC had a better standard of error of measurement in the estimation of examinees' ability than the WAEC test does. The results imply that the 2019 NECO Physics multiple-choice test measured examinees' abilities with more precision than the WAEC test does.

Discussion

This study investigated the comparability of the 2019 Senior School Certificate Physics multiple-choice examinations in NECO and WAEC. It was revealed that WAEC's multiple-choice items were generally more difficult than NECO's for that year. However, item difficulty between the two examination bodies did not yield any significant difference, which implies that they were of comparable difficulty. In line with the findings of the present study, the response items that were designed by WAEC were difficult compared to those in NECO. The implications for this fall back to the essence of curriculum alignment, student preparation, and readiness to answer WAEC questions.

The results of this study are in alignment with the submission of Udofia and Udoh (2017) where it was found that there was a similarity in the difficulty rate of NECO and WAEC questions. This shows that different factors determine test difficulties such as the process of item construction, standardization, and content selection. In the same vein, the perception of teachers, their quality, and preparation play important roles in the difficulty of a test. In another study that corroborates the present study, Nweze and Obu (2022) asserted that the item distribution of Chemistry multiple-choice examination in 2021 NECO and WAEC Chemistry was significant with their level of difficulty. On this note, establishing whether a significant or non-significant relationship exists in a test difficulty based on teacher quality and item constructions.

In the present study, a significant difference was found in the performance of students in both examinations (WAEC and NECO). This implies that the majority of the students had an equal performance in the two examinations. Although it has been observed that Physics is a generally difficult subject, the students had relatively equal performance in WAEC and NECO of the subject. Some of the areas in which the students had similarities in performance were physics curriculum, comparable levels of cognitive demands in test items, and the use of identical multiple-choice formats. Additionally, similarities in school environments, instructional methods, and common test preparation practices. Unlike the findings of Utibe and Agah (2015) in a study that was done between 2009 to 2012, the performance of students in WAEC was better compared to NECO. This shows that while the method of supervision might facilitate differences in test difficulty, this factor was not considered in the present study.

From the study, it was found that NECO's examination had a lower standard error of measurement (SEM) (mean = 0.25) compared to WAEC (mean = 0.33). By implication, NECO ensures more reliable and precise student abilities estimates. Also, there was a significant difference in SEM ($t = 19.493$, $p < 0.05$), which means more accuracy in the abilities of students in NECO compared to WAEC. Although, having more data points offer a more reliable estimate of the examinee's true score, test reliability rises and SEM decreases, respectively (Crocker & Algina, 2006; Lord, 1980). By inference, there is always the need to go through viable statistical

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

analyses, content alignment, careful item construction, and discrimination in the designing of test items as also stressed by Adewuni and Busari (2021). In addition to this, more components in test construction like reliability and precision can also be investigated in further studies.

Conclusion And Recommendation

As shown in the study, WAEC items were more difficult than the NECO items. However, there was no significant difference in students' performance in both examinations. Also, student ability estimates were better in NECO using the standard error of measurement, and this difference was statistically significant, indicating more precise ability estimation.

As a way of recommendation, test experts for WAEC should consider adjusting the difficulty level of their questions to ensure a fair assessment of student's knowledge and abilities which could involve reviewing and revising the question items to make them more aligned with the expected difficulty level for the students and their scoring system to ensure fairness and accuracy; School administrators and teachers should use these findings useful in tailoring their teaching and assessment methods to better prepare students for the exams in which seems more challenging. In the same vein, test experts for NECO and WAEC should continuously monitor and improve the validity and reliability of tests.

References

Adewuni, D. A., & Busari, Y. T. (2021). Analysis of standard error of measurement (SEM) of2019 West African Senior School Certificate Examination multiple-choice objective tests in Economics. *Kashere Journal of Education, 2*(2), 25–33.

Baker, F., & Kim, S. (2017). *The basics of item response theory using R*. https://doi.org/10.1007/978-3-319-54205-8

Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory* (2nd ed.). Cengage Learning.

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items* (3rd ed.). Routledge.

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

Jegede, S. A., & Adedayo, J. O. (2013). Enriching physics education in Nigeria towards enhancing a sustainable technological development. Department of Curriculum Studies, Faculty of Education, Ekiti State University.

Liu, Y., & Maydeu-Olivares, A. (2012). Local dependence diagnostics in IRT modeling of binary data. *Educational and Psychological Measurement, 73*(2), 254–274. https://doi.org/10.1177/0013164412451900

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

Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *International Journal of Educational and Psychological Assessment, 1*(1), 1–11.

Nguyen, T. H., Han, H. R., Kim, M. T., & Chan, K. S. (2014). An introduction to item response theory for patient-reported outcome measurement. *Patient, 7*(1), 23–35. https://doi.org/10.1007/s40271-013-0041-0

Nweze, B. N., & Obu, A. U. (2022). Comparative analysis of 2021 and 2022 West Africa Examination Council (WAEC) and National Examination Council (NECO) Chemistry multiple choice questions in Enugu State, Nigeria. *British Journal of Education, 10*(14), 7–14.

Ojerinde, D., Popoola, K., Ojo, F., & Onyeneho, O. P. (2012). *Introduction to item response theory: Parameter models, estimation and application*. Goshen Print Media Ltd.

Osterlind, S. J. (2012). Item response theory. *Journals of Home School and Academic Learning*.

Udofia, N. A., & Udoh, K. I. (2017). Comparative analysis of WAEC and NECO senior secondary school mathematics examination. Department of Educational Foundations, Guidance and Counselling, Faculty of Education, University of Uyo.

Utibe, U. J., & Agah, J. J. (2015). Comparative analyses of physics candidates' scores in West African and National Examinations Councils. *Journal of Education and Practice, 6*(25), 100–107.