JOURNAL OF RESEARCH IN SCIENCE AND VOCATIONAL EDUCATION (JRSVE)
Volume 3; Issue 1, 2023
Official Publication of Science and Vocational Department
Godfrey Okoye University, Ugwuomu- Nike, Enugu-Nigeria
Published by K-Injo | Kobia Data Limited

PARAMETERISATION OF PHYSICS MULTIPLE CHOICE ITEMS FOR STUDENTS OF TECHNICAL COLLEGES IN ENUGU STATE, NIGERIA

Mark Eze Ugwoke
Department of Science and Computer Education, Godfrey Okoye University, Enugu, Nigeria.
markugwoke@gouni.edu.ng

Abstract: The study parameterised physics multiple choice items towards achieving validity of the items for students in technical colleges in Enugu State. The design adopted for the study was instrumentation research design. All year two students of physics in technical colleges in Enugu State in the 2022/2023 academic session formed the population of the study. Purposive sampling technique was used to select 175 students from six technical colleges across the six educational zones of the state. The instrument used for data collection was Physics Achievement Test which was validated by experts. Kuder-Richardson's formular 20 reliability coefficient of 0.83 was established. The instrument was administered to the students and their responses were used to prepare a-person by item matrix. The Binary Logistic Regression was used for data analyses. The results of the data analyses yielded the answers to the three research questions. It was revealed that while some items of PAT did not fit the three-parameter logistic model of Item Response Theory. The result equally revealed that some items of the PAT were too difficult while some discriminated negatively. One of the major recommendations is that physics test items should be parameterised towards enhancing the overall validity of the test**.**

Keywords: Construction of items, parameterisation of items, test items, test validity.

Introduction

Technical college is one of the educational institutions at the post- basic–education level where Technical and Vocational Education (TVE) is comprehensively offered in Nigeria. The TVE programme in the colleges runs for a period of three years (F.R.N., 2014). The Technical and Vocational Education is used as a comprehensive term referring to "those aspects of educational process involving, in addition to general education, the study of technologies and related sciences and acquisition of practical skills, attitudes, understanding and knowledge relating to occupation in various sectors of economic and social life" (F.R.N., 2014:22). Admissions into technical colleges are trade-based and passing Basic Education Certificate Examination (BECE). The BECE is therefore one of the entry requirements. Each trade consists of Trade-Related subjects, Core-Component subjects and General-Education subjects including Physics (F.R.N., 2014).

Physics is one the basic and compulsory General-Education science subjects for most students in technical colleges in Nigeria. Physics is the study of matter, energy, and the interactions between them. It provides a comprehensive understanding of the natural world through observation, experimentation, and mathematical modeling. Physics is useful towards national development and can be applied in automobiles, space exploration, aeronautics, electronics, communication, medicine, teaching and warfare. Specifically, Physics is offered by all students in Engineering and Construction trades. Enrolment figures for the subject In the National Business/Technical Certificate (NBC/NTC) examinations in Enugu state, in recent years, have been declining while students' performance on the subject in the NBC/NTC examinations of the National Business and Technical Examinations Board (NABTEB) is persistently poor (NABTEB, 2021). In the colleges, teaching and assessment of the subject are based on national curriculum.

Assessment is one of the crucial components of teaching and learning activities; it is the major drive of students' learning. Assessment drives the curriculum. In technical colleges many assessment tools are deployed to evaluate what learners have been able to achieve. Multiple choice items (MCIs) are among the tools used to evaluate learners' achievements. A multiple choice item consists of a problem, known as the stem, and a list of suggested solutions, known as alternatives (Ebuoh 2016; Anigbo, 2014). The alternatives consist of one correct or best alternative, which is the answer, and incorrect or inferior alternatives, known as distractors (Anigbo, 2014). According to Abdalla, Gaffer, and Suliman (2011), MCIs are popular in educational assessment because they are easy to administer and mark even to large number of students. The MCIs have testing efficiency as well as objectivity. Each test item must be valid to function effectively in a test (Chakrabartty, 2020).

Validity of test items is essential to ascertain the level of achievement of the learners. Validity is the most important feature while considering selection of test items. Asuquo et al., (2022) remarked that validity refers to a particular construct or characteristic the test measures and how well the test measures that construct or characteristic. Validity gives desired information on the characteristic being measured by a test in relation to job qualifications and requirements. Validity evidence indicates that there is linkage between test performance and job performance, which is, giving meaning to the test scores as well as the usefulness of the test. If a test has been demonstrated to be a valid predictor of performance on a specific job, it can be concluded that persons scoring high on the test are more likely to perform well on the job than persons who score low on the test, all else being equal. Different types of validity include content validity which ensures that the test adequately samples the content it is supposed to measure and it is assessed by examining the test items and comparing them to the content domain. Construct Validity on the other hand focuses on the underlying theoretical construct that the test is designed to measure and involves establishing relationships between the test scores and the theoretical construct while the criterion-related validity involves comparing the test scores to an external criterion to determine how well the test predicts or correlates with the criterion.

In test development and validation processes, the actual values of the parameters of the items in a test are not known, therefore, it is necessary to analyse or parameterise the items to estimate the values for these parameters. The obtained item parameter estimates provide information on the technical properties of the test items. Parameterisation of test items is the process of estimating unknown values of test items' parameters by providing information on the technical properties of the items (Baker & Kim 2017). It is the process of defining and setting parameters for each item in a test and involves systematically analysing and adjusting test items to achieve validity. The parameters of a test item include various characteristics that influence its difficulty, discriminatory power, and overall effectiveness in measuring the construct of interest. Proper parameterization of test items is essential for creating a valid and reliable assessment tool. Baker and Kim (2017) remarked that parameterisation involves combining the characteristic of an item and a person to yield a probability of correct response from that particular test taker. It is a process that involves systematically analyzing and adjusting test items to achieve validity of the test. In summary, parameterisation of test items is a critical assessment tool for achieving test fairness, test equating and adaptive testing by tailoring the difficulty of items based on a test-taker's performance.

Test developers always place emphasis on the quality of test items – the validity of the items. Traditionally, the development, scoring and analysis of test items have been carried out using classical test theory (CTT). The Classical Test Theory despite its popularity among testing boards in Nigeria has inherent weaknesses which include the facts that CTT measures performance of an examine at test level instead of measuring the same at item level; the item parameters(p) and reliability estimate obtained using CTT depend on examinee sample. However, a more modern and model-based theory called item response theory (IRT) has been introduced into the assessment world and offers and has the capacity of addressing practical measurement problems found difficult to solve through classical methods. Used extensively in educational testing applications (Ojerinde, et al 2014) opined that IRT is a set of mathematical models that describe the relationship between an individual's 'ability' or 'trait' and how they respond to items on a scale. This relationship is depicted by an item characteristic curve (ICC). The ICC is a probability curve that is monotonic, or continuously increasing in nature (Ojerinde, et al 2014). As an individual's trait level increases, the probability of endorsing an item also increases.

The IRT has the property in which ability scores remain constant, despite the administration of different items which is the driving force behind computer-based testing (CAT) that selects specific items for each individual based on their prior response to an item with the goal of maximizing precision. In IRT the estimated item parameters are constant across different samples (i.e. measure equivalence). IRT facilitates the estimation of measurement bias that occurs when individuals from different groups respond differently to an item, resulting in estimated parameter(s) that vary even after controlling for the underlying trait level. According to Tschering (2006) cited in Ugwoke (2016), item parameters estimates of the three – parameter (3p) logistic model consists of the item difficulty (b-parameter), item discrimination power (a-parameter) and the guessing factor (c- parameter) . Nwogu (2015) stated that items like every other measuring devices, have indices for determining their (items') quality. The indices are called the psychometric qualities or characteristics of the items (Nworgu, 2015). Items with unsound qualities would lead to poor achievement whenever a test consisting of such unsound items are administered to a group of examinees. Therefore, an investigation of the item - parameter estimates as intended in this study is desirable as it will shed more light into the issue. There are basically three parameters or factors that determine the nature of IRT model, these are: Item difficulty (b-parameter), Item discrimination (a-parameter), and Pseudo guessing factor (c-parameter) (Mahjabeen, et al., 2018 & Mukherjee, 2015). The fourth parameter – carelessness, has not been fully developed (Adegoke, 2013). Thus, the probability of correct response includes a small component that is due to guessing and the equation for the three parameter logistic IRT model (Nataranja, 2009) is

$$P(\Theta) = C + (1 - c)\frac{1}{1 + e^{-a(\theta - b)}} \qquad (1)$$

where e is the constant of value 2.718.
b is the difficulty parameter,
a is the discrimination parameter (power);
c is the guessing parameter; and
Ɵ is the ability level of a particular examinee.

Assessment of goodness of fit of responses to the 3 – parameter logistic model is a prerequisite activity to item parameterisatrion. Guler, Uyanik and Teker (2013) remarked that items with misfit statistics to a particular model of IRT will yield wrong parameters. The item difficulty (b-parameter) is the most central of all the three parameters. Ojerinde (2013) stated that item difficulty parameter is a location on the ability continuum where an examiner has fifty percent chance of getting the item correct and fifty percent chance of getting the item incorrect. The item difficulty is useful in interpreting how the examinees whose abilities fall before the ability that corresponds to the item difficulty parameter on the ability continuum would score the item incorrectly (Ojerinde, et al., 2012). The examinees whose ability exactly fall on the location of the ability that corresponds to the item difficulty parameter on the ability continuum have fifty percent chance of scoring the item correctly and fifty percent chance of scoring the item incorrectly (Siwi et al., 2020). The examinees whose abilities fall after the ability that corresponds to the item difficulty parameter on the ability continuum are expected to score the item correctly (Ojerinde, et al., 2012). It can be referred to as the difficulty parameters or difficulty index of the Item Response Function (Si, 2002). The b-parameter of an item is determined by finding the point on the ability scale (θ) at which the probability of correct response P(θ), to the item is 0.5, when the Item Characteristic Curve starts from zero or is approximately so The more difficult an item is, the larger (in positive direction) the b-value of the item response functions. The b-parameter has typical values of range from -3.0 to +3.0 (Adegoke, 2013). This implies that items with -3.0 b – value are very easy while items with b-value of 3.0 are very difficult and items of b-value of 0.0 indicates/shows moderately difficult items. Items that are too easy or too difficult should be revised before inclusion in the test (Ebuoh, 2004).

Item discrimination otherwise called a-parameter is another quality of a valid test item should possess. Lord cited in Ekwonye & Eguzo (2011) defined the item discrimination (a-parameter) as the degree to which item response varies with ability. It is the discrimination index of the item response function (Item Characteristic Curve) which describes the slope of the logistic curve. It is determined from the Item Characteristic Curve (ICC) by taking the slope of the line tangent to the Item Characteristic Curve at 'b'. The slope, m, is approximately equal to a(1 – c)/4 (Baker, 2001 cited in Ugwoke, 2016). However, a-parameter is related to the slope of the ogive at the inflection point or at the b-value. For normal ogive with c=0.0; a = 4m, where m = slope of the Item Characteristic Curve. The greater the value of a (a-value) the better the item discriminates or distinguishes among examinees. On the other hand, item of low a–value discriminates poorly among examinees. An item with high discrimination parameter is more discriminative than an item with low discrimination parameter (Chakrabartty, 2020). In other words, item discrimination parameter measures the worth of an item in terms of its usefulness in separating high ability examinee and low ability examinee (Ojerinde, 2013). It is in the light of this property of the discrimination parameter that a psychometrician gets delighted by an item with high discrimination value and concerned by an item with low discrimination value Warm cited in Ekwonye and Eguzo stated that the a–value of typical items vary from 0.5 to 2.5. While most test items will discriminate in positive manner (i.e. the probability of correct response increases as the ability level increases), some items have negative discrimination. In such items, the probability of correct response decreases as the ability level increases from low to high.

The c-parameter is the guessing parameter. It is the lower asymptote of the logistic curve, i.e. the low point of the curve as it moves to a negative infinity on the horizontal axis. The c-parameter can be used to model guessing in multiple choice items. The c-values range from 0.00 to 0.4. Items with c-values 0.3 or greater are not very good items. It is desirable to have c-values at 0.2 or less; the lower the value of c, the better the items while a zero c-value is ideal (Ekwonye and Eguzo, 2011). Siwi et al. (2020)

developed an online-based English proficiency test called the English Proficiency Online Test (EPOT) and used a descriptive quantitative approach to describing the characteristics of EPOT test items in terms of item difficulty index, item discrimination index, test information's function, test measurement's errors and found from analysis of responses of 2,652 online test-takers from 20 provinces in Indonesia using 3-PL model of IRT that most of EPOT's test items had a good range of difficulty index and discrimination index. Uyar and Öztürk-Gübeş (2020) who studied item parameter estimation for dichotomous items based on IRT while citing Lord (1968) stated that, at least 50 items and 1000 sample sizes were required to estimate the discriminant parameter (a parameter) accurately for the 3PL model and further remarked on the contrary that Swaminathan and Gifford (1983) investigated the effect of sample size, test length, and the ability distribution on the estimation of item and ability parameters using the 3-PL model got results that showed that the condition in which sample size was 1000 and test length was 20 produced more accurate estimates of the difficulty and guessing parameters, and fairly good estimates of the item discrimination parameters than the conditions in which sample size was 50 and test lengths were 10 or 15 and sample size was 200 and test lengths were 10 and 15.

Statement of Problem

The persistent poor performance of students of Enugu State in the National Business Certificate (NBC) and National Technical Certificate physics examinations of the National Business and Technical Certifications Board (NABTEB) is a source of serious concern to educators in the state. Little attention is paid to the critical issue of the quality of the test items used in measuring the level of achievement of students in physics. Test items which parameters were not critically parameterised may yield undesirable responses which could lead to the persistence poor achievement of students in physics tests at this level. To the best of the researcher's knowledge, most of the studies on parameterisation of Physics items were conducted at secondary school level thus, neglecting the technical colleges. The researcher views this as a serious gap that could lead to the persistent poor achievement of students in physics test in technical colleges in Enugu State. This study, therefore parameterised Physics multiple choice items using of three-parameter logistic model of the IRT for students in technical colleges in Enugu state.

Purpose of the Study

The purpose of the study is to parametise physics multiple choice items for students of technical colleges in Enugu State. Specifically, the study intends to find estimate the:

(i) difficulty (b) parameter of the Physics' Achievement Test

(ii) discrimination (a) parameter estimates of the Physics' Achievement Test, and

(iii) guessing factor (c) parameter of the Physics' Achievement Test;

Research Questions

To guide the research, the following research questions were posed:

(i) What are the difficulty (b) parameter estimates of the Physics Achievement Test?

(ii) What are the discrimination (a) parameter estimates of the Physics Achievement Test?

(iii) What are the guessing factor (c) parameter estimates of the Physics Achievement Test?

Research Method

Instrumentation research design was used. The study was conducted in Enugu state which has large number of accredited technical colleges spread in six educational zones of the state namely: Enugu, Nsukka , Obollo-Afor, Awgu, Udi and Agbani zones. The population for the study consisted of 1025 year two (NTC 2) Physics students of all accredited technical colleges in Enugu State for the 2022/2023 academic session (Science, Technical and

Vocational Schools' Management Board, Enugu). The students were presumed to have covered the topics as contained in the curriculum used for the development of PAT. Multistage sampling technique was employed in selecting 175 students using six colleges, one from each of the three educational zones. A table of specification was prepared to guide the development of the test called Physics Achievement Test (PAT). It comprised 40 – multiple - choice item drawn from Mechanics, Heat, Optics, and Electricity. The instrument (PAT) was face- and content- validated by three experts and corrections were effected. The instrument was trial tested on 30 students (year II - NTC 2) of Government Technical College, Abakaliki, Ebonyi State. The college and the students were not part of the sample used for the study. The college is also located in the urban area and the students are in the same class (NTC 2) with those used in the study; therefore, they are presumed to have covered all the topics contained in the test. The reliability coefficient of 0.83 was established. using the Kuder-Richardson formular 20 (K-R 20) method. The Kuder-Richardson 20 (K-R 20) because PAT was dichotomously scored (Idoko, 2011; Obodo, 2014). Two research assistants who were trained on test administration, scoring, recording of the examinees' responses in EXCEL package were used. They were used for administration of the instrument and retrieval of scripts on the spot. The students' responses in PAT were scored 1 for correct answers and 0 for wrong answers while item not responded to by any examinee was represented by 9. The data were analysed and parameterised by using Binary Logistic Regression.

Results

Research question one

What are the difficulty (b) parameter estimates of the Physics Achievement Test?

Table 1: Item difficulty (b) statistics of the Physics Achievement Test using three – parameter (3p) model of the Item Response Theory.

Item ID	Goodness of fit	b	Flag(s)
1	1.492	0.473	
2	1.430	0.898	
3	0.880	3.000	
4	1.099	**4.000**	Flagged - too difficult
5	1.227	0.358	
6	1.508	1.696	
7	1.537	**3.626**	Flagged - too difficult
8	1.819	**3.900**	Flagged - too difficult
9	1.418	0.306	
10	0.799	**3.693**	Flagged - too difficult
11	0.402	2.503	
12	1.231	0.464	
13	1.270	3.000	
14	0.271	3.000	
15	0.497	2.674	
16	1.846	-0.664	
17	1.228	**3.543**	Flagged - too difficult
18	0.826	2.644	
19	1.278	3.000	
20	0.980	2.664	

Item ID	Goodness of fit	b	Flag(s)
21	1.324	-1.455	
22	1.115	**3.709**	Flagged - too difficult
23	1.879	**3.472**	Flagged - too difficult
24	1.722	**3.637**	Flagged - too difficult
25	0.806	2.921	
26	0.427	**3.676**	Flagged - too difficult
27	0.445	2.249	
28	1.073	2.841	
29	1.857	**3.851**	Flagged - too difficult
30	1.843	**3.978**	Flagged - too difficult
31	1.296	0.476	
32	0.861	**3.675**	Flagged - too difficult
33	0.606	**3.645**	Flagged - too difficult
34	1.623	**3.569**	Flagged - too difficult
35	1.339	3.000	
36	1.348	1.702	
37	0.929	**3.727**	Flagged - too difficult
38	0.409	3.000	
39	0.457	3.000	
40	0.724	2.921	

Table 1 depicts item difficulty (b) statistics of the Physics Achievement Test using three – parameter (3p) model of the Item Response Theory. The benchmark for inclusion of an item in the PAT was $-3.0 \leq b \leq +3.0$, therefore, items numbers 4, 7, 8, 10, 17, 22, 23, 24, 26, 29, 30, 32, 33, 34 and 37 were too difficult and flagged for revision. F=misfit items (z-residual > 2.00)

Research question two: What are the discrimination (a) parameter estimates of the Physics Achievement Test?

Table 2: Item discrimination (a) statistics of the Physics Achievement Test using three – parameter (3p) model of the Item Response Theory.

Item ID	Goodness of fit	a	Flag(s)
1	1.492	0.454	
2	1.430	**-0.459**	Flagged – Negative discrimination
3	0.880	0.719	
4	1.099	1.018	
5	1.227	0.410	
6	1.508	0.578	
7	1.537	1.103	
8	1.819	1.148	
9	1.418	0.374	
10	0.799	1.036	

Item ID	Goodness of fit	a	Flag(s)
11	0.402	0.735	
12	1.231	0.503	
13	1.270	0.787	
14	0.271	0.919	
15	0.497	0.727	
16	1.846	0.592	
17	1.228	0.940	
18	0.826	0.969	
19	1.278	0.911	
20	0.980	0.678	
21	1.324	0.400	
22	1.115	0.980	
23	1.879	0.881	
24	1.722	0.946	
25	0.806	1.084	
26	0.427	0.992	
27	0.445	0.793	
28	1.073	0.959	
29	1.857	1.117	
30	1.843	1.068	
31	1.296	**- 0.484**	Flagged – Negative discrimination
32	0.861	1.039	
33	0.606	1.034	
34	1.623	0.866	
35	1.339	0.907	
36	1.348	**-0.408**	Flagged – Negative discrimination
37	0.929	1.026	
38	0.409	0.833	
39	0.457	0.968	
40	0.724	1.185	

Table 2 depicts item discrimination (a) statistics of the Physics Achievement Test using three – parameter (3p) model of the Item Response Theory. Benchmark for inclusion of an item in the PAT $+0.3 \le a \le +4.0$ with table 2 indicating that items numbers 2, 31 and 36 discriminated negatively and were flagged for revision. F=misfit items (z-residual > 2.00)

Research question three: What are the guessing factor parameter (c) estimates of the Physics Achievement Test?

Table 3: Item guessing factor (c) statistics of the Physics Achievement Test using three – parameter (3p) model of the Item Response Theory.

Item ID	Goodness of fit	c	Flag(s)
1	1.492	0.292	
2	1.430	0.273	
3	0.880	0.329	
4	1.099	0.187	
5	1.227	0.255	
6	1.508	0.244	
7	1.537	0.188	
8	1.819	0.171	
9	1.418	0.259	
10	0.799	0.202	
11	0.402	0.237	
12	1.231	0.251	
13	1.270	0.270	
14	0.271	0.220	
15	0.497	0.247	
16	1.846	0.248	
17	1.228	0.262	
18	0.826	0.237	
19	1.278	0.270	
20	0.980	0.285	
21	1.324	0.255	
22	1.115	0.249	
23	1.879	0.265	
24	1.722	0.272	
25	0.806	0.209	
26	0.427	0.193	
27	0.445	0.266	
28	1.073	0.273	
29	1.857	0.205	
30	1.843	0.221	
31	1.296	0.259	
32	0.861	0.180	
33	0.606	0.188	
34	1.623	0.311	
35	1.339	0.299	
36	1.348	0.276	
37	0.929	0.209	
38	0.409	0.253	
39	0.457	0.218	
40	0.724	0.172	

Table 3 depicts guessing factor (c) statistics of the Physics Achievement Test using three – parameter (3p) model of the Item Response Theory. Benchmark for inclusion of an item in a the PAT $0.00 \leq c \leq +0.40$ with table 3 indicating that none of the items PAT was flagged based on guessing factor.
F=misfit items (ƶ-residual > 2.00)

Discussion

The results of the item parameter calibration of the PAT indicate that some of the items were too difficult. Items that were identified to be too difficult imply that it only the group of students with high abilities will be able to respond correctly to the answer. Such items are not supposed to be included in the test; instead they are flagged for revision. This result is corroborated by Ebuoh (2004) who stated that test items should be critical examined or parameterised to ensure that they are neither too difficult nor too easy. According Ebouh (2004), items that are too difficult are failed by majority of the test takers while those that too easy will be passed by almost all the test takes.

Item discrimination parameter is one of the essential characteristics of a test. Quality test items should discriminate the good students from the bad students or students with high abilities from those with low abilities. The result of this study indicates that some of the items discriminated negatively. Items with negative discrimination value imply that students with low abilities will respond correctly to the items while students with high abilities on the construct being measured will fail such items. The items were flagged for revision. This finding was corroborated by Chakrabartty (2020), Ebouh, (2004), Ekwonye and Eguzo (2011) who remarked that items with negative discrimination values should be removed from inclusion in the test or better still, revised. However, Uyar and Öztürk-Gübeş (2020) who studied item parameter estimation for dichotomous items based on IRT pointed out that while discriminant parameter (a parameter) 3PL model depends on sample size and test length.

Furthermore, the result of the study indicated that none of the items of PAT was flagged based on guessing c-parameter. In actual testing situation using MCIs, there is the chance for test takers of low abilities getting an item correctly by guessing. Such items when identified should be removed from the test. A test taker should possess the ability on a construct being measured by an item to be able to respond correctly to the item.

Conclusion and recommendation

From the results of the study it is concluded that parameterization is essential to enhancing test quality and validity. It is therefore recommended that:

1. Physics teachers particularly those in technical colleges in Enugu state should parameterise their test to enhance its quality and validity.
2. The State Ministry of Education should organize regular workshops for teachers to develop skills on parameterisation of test items to facilitate effective teaching, learning, assessment and evaluation activities in the colleges in the state.

References

Abdalla, M. E., Gaffer A. M. & Suliman R.A. (2011). Construction of a multiple choice type of questions (MCQs): Step - step

https://www.academia.edu/1215065/Constructing_A_Type_Multiple_Choice_Questions_MCQs_Step_By_Step_Manual

Adegoke, B.A. (2013). Comparison of item statistics of Physics achievement test using classical test theory and item response theory frameworks: *Journal of Education and Practice*. (4) 22, 87.

Anigbo, L.C. (2014). *Teachers' handbook on* *measurement and evaluation for effective teaching and learning.* Enugu: Executive Press Limited.

Asuquo U. B. 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. & Kim, S. H. (2017). Estimating item parameters. In: The Basics of Item Response Theory Using R. *Statistics for Social and Behavioral Sciences*. Springer, Cham. https://doi.org/10.1007/978-3-319-54205-8_3

Bergan J. R. (2010). Assessing the relations fit of alternative item response theory model to data. *Research Paper*. Tucson, Arizona, USA. Assessment Technology, Incorporated Publishers.

Chakrabartty, N. S. (2020). Improved quality: Item and test parameters. *Health Sciences*. I(ID) DOI: 10.15342/hs.2020.267

Ebuoh, C. N. (2004). *Educational measurement and evaluation for effective teaching and learning*. Enugu: Sky Printing Press.

Ekwonye, E. C. & Eguzo, G. O. (2011). *Basic test theories in measurement and evaluation.* Owerri: Joe Mankpa Publishers.

Guler, N; Uyanik, G. K. & Teker (2012). Comparison of classical test theory and item response theory in terms of item parameters. *European Journal of Research on Education*, 2(1), 1 – 6.

Idoko, C. .E. (2011). *Research in education and social sciences (practitioners' companion).* Enugu: Our Saviour Press Limited.

Mahjabeen, W., Alam, S., Hussan, U., Zafar, T.,Butt, R., Konain, S. & Rizvi, M. (2018). Difficulty index, discrimination index, and distractor efficiency in multiple choice questions. *Annuals of Pakistan Institute of Medical Sciences.*4:310-315.

Mukherjee P., & Lahiri S.K. (2015). Analysis of multiple-choice questions (MCQs): Item and test statistics from an assessment in a medical college of Kolkata, West Bengal. *Journal of Dental and Medical Sciences.* 14(12):47-52.

Nworgu, B. G. (2010). Differential item functioning: A critical issue in regional quality assurance. *The Journal of the Association for Educational Assessment in Africa*, 6. pp. 274-284.

Nworgu, B. G. (ed) (2015). *Educational measurement and evaluation: theory and practice*. Nsukka: University Trust Publishers.

Obodo, G. C. (2014). *Ordinary and advance level statistics for undergraduates and postgraduates.* Enugu: Celex Printers (Nig).

Ojerinde, D.; Popoola, K.; Ojo, F.; & Onyeneho, O. P. (2012). *Introduction to item response theory parameter models, estimation and application (2nd Edition). Ogbomosho: Marvellous Mike Press Ltd*

Ojerinde, D. (2013). *Classical test theory vs item response theory: An evaluation of the comparability of item analysis results.* A paper presented at the Institute of Education, University of Ibadan on May, 23.

Olufemi, A. S. (2013). Item response theory as a basis for measuring latent trait of interest. *Greener Journal of Social Science.* 3(7) pp. 378-382

Siwi, E., Anindyarini, R., & Nahar, S. (2020). Item parameters of Yureka Education Center (YEC) English Proficiency Online Test (EPOT) instrument. REiD *Research and Evaluation in Education*, 6(1), 51-65. doi.org/10.21831/reid.v6i1.31013.

Ugwoke, M. E. (2016). Application of item response theory in the construction of physics items for item banking for technical colleges in Enugu State. Unpublished *M.Sc. Thesis*, Enugu State University of Science and Technology.

Uyar, Ş. & Öztürk-Gübeş, N. (2020). The importance of sample weights and plausible values in large-scale assessments. *Journal of Measurement and Evaluation in Education and Psychology*, 11(1), 27-42. doi: 10.21031/epod.591415