AJSTME, Volume. 12, Issue 2. 17-25, March, 2026. https://www.ajstme.com.ng
ISSN: 2251-0141 (Print), 2971-6233 (Online)

Estimation of the Item Characteristics of Basic Science Achievement Test Item Formats Using Three Parameter Logistic Model

Agah, John Joseph, Ali, Obioma & Nwani, Samuel Uchenna
Department of Science Education, University of Nigeria, Nsukka
Corresponding Author: nwanisamueluchenna@gmail.com

Abstract

The study estimated the item characteristics of Basic Science achievement test item formats using three parameter logistic model. The study adopted descriptive research design and was carried out in Enugu state. The population of the study was21,215 JSS3 students. A sample of 1,500 was obtained through multistage sampling procedure. The instrument used for data collection was Basic Science achievement test which was prepared in three different formats namely; Alternate Item Format (AIF), Matching Item Format (MIF) and Multiple-choice Item Format (MIF). Reliability coefficients of 0.96, 0.98 and 0.99 were obtained, for the three item formats using Kuder-Richardson 20. The data were collected through direct delivery technique. The research questions were answered using the item difficulty, discrimination, guessing parameter indices using JMetrik statistical Package. The findings of the study among others revealed that there is no significant difference in the item facility indices of the three item formats in Basic science achievement test, there is no significant difference in the item discrimination indices of the three item formats in Basic science achievement test, and there is no significant difference in the guessing parameter indices of the three item formats in Basic science achievement test. The study among others recommends that test developers should use the 3PL model to carefully select test items for Basic Science Achievement Tests.

Keywords: Basic science, item formats, item difficulty, item discrimination and guessing parameter indices.

Introduction

Increase in demand for precise and reliable assessment tools in educational assessment, particularly with Basic Science achievement, has triggered research on advanced methodologies for estimating item characteristics. Basic science prepare students for learning of science subjects such as chemistry, physics, biology, agricultural sciences among others in higher secondary school (Oguguo & Lotobi (2019). In spite of all this enormous benefit of Basic science for national development students' achievement is still below expectations. This is evident in the reports of 2016 to 2018 results from the Examination Development Centre of the Enugu State Ministry of Education that are in charge of Basic Education Certificate Examination has shown that students' achievement in Basic science is on the down ward trend

This issue bothers educators, stakeholders and researchers. Studies have shown that there are various factors that contribute to the downward trend in student achievement. Adamoah and Acquah (2016); Asare-Bediako and Aboagye, (2016); Onanuga and Saka, (2018) reported that the factors that the factors that contribute to the downward trend in student achievement include; teachers, the student, home environment and teaching methodology. It would appear that little or no attention was paid on the ways questions are generated and presented in a test to the students. Every test is a collection of items and the format in which items are presented matters a lot. Item format and item characteristics play important role in students' academic achievement. In teaching and learning processes, methods of setting and standardizing questions are important to both teachers and students.

A standard test is designed to measure the degree or level of proficiency of students. The objectives of achievement tests according to Nworgu (2018) include; to measure whether students possess the

pre-requisite skills needed to succeed in any unit or whether the students have achieved the objectives of the planned instruction, to monitor students' learning and to provide on-going feedback to both students and teachers during the teaching-learning process, to identify the students' learning difficulties-whether persistent or recurring, to assign grades.Despite the usefulness of achievement test and the considerable efforts made by educators, policy makers, governments, researchers, society and parents at various levels to see that student's achievement is improved, little or no improvement has been recorded on the student achievement in basic science as earlier reported. According to the report and available data from examination bodies especially the National Examination Council (2018) in their appraised report on SSCE Indicated a downward trend in student's achievement in science subjects in Nigeria (Chief Examiners Report, 2018).

It is based on the persistent failure of students in basic science and technology that made the researcher to further presume item format and item characteristics as a probable factor responsible or contributing to this problem. The reason may be that some of examination item format may be poorly constructed. Item formats and item characteristics may lack psychometric properties that may be expected of them. This problem has engaged the attention of early scholars. For instance, Oguguo (2016) reported that the parameters of Basic science test items of 2011 Basic Education Certificate Examinations has only few items that satisfied all three Item Response Theory (IRT) parameter estimates in. Orluwene (2017) maintained that item format has effect on students' achievement. From the ongoing, it was observed that no evidence show that any research has been carried out to investigate the effect of item format on item characteristics of achievement test in basic science and technology. Although a lot researcher has done so many works on the possible causes of student poor achievement in basic science. Most of the finding show that students' achievement in basic science and technology are still below the expected range.

Each of the item formats differ in terms of characteristics of the items. The item characteristics are the validity indices. These indices guide test item development and can also be statistically determined (Meyer & Zhu, 2013). De Ayala (2009) called these indices "item parameters". Item parameters in this study are; item difficulty indices, item discrimination indices and guessing factor indices. Item Difficulty (*b*) is the parameter that determines the manner of which the item behaves along the ability scale. It is a measure of proportion of examinees who answered the item correctly (Maydeu-Olivares, 2015).According to Mozaffer and Farhan (2012) it is determined at the point of median probability, that is, the ability at which 50% of respondents endorse the correct answer. The more difficult it is for a student to have a 50% chance of correctly answering an item, the higher the ability level needed to achieve this goal (Yang & Kao, 2014) especially when juxtaposed with the discrimination index of the items.

The discrimination index (*a*) of an item is the ability to distinguish high and low scoring learners. The closer this value is to 1, the better the item distinguishes the learners who get a high score from those who get a low score (Yang & Kao, 2014). Discrimination index contributes to increasing the validity and reliability of the tests by revealing whether the items are working well or not. It determines the rate at which the probability of endorsing a correct item changes given ability levels. It is imperative in differentiating between examinees possessing similar levels of the latent construct of interest (Maydeu-Olivares, 2015). According to Ziegler and Hagemann (2015), the ultimate purpose, for designing a precise measure is to include, items with high discrimination, in order to be able to map individuals along the continuum of the latent trait. The item discrimination index is a measure of how well an item is able to distinguish between examinees who are knowledgeable and those who are not (Miller, Linn & Gronlund, 2009).

The analysis of each item by calculating difficulty and discrimination indices provides feedback on what the learners have learned and enables instructors to determine and correct the faulty items. The implication of not carrying out item analysis to determine the difficulty indices and discrimination indices of Basic science and technology examination is that some items may be faulty especially in classifying the examinees based on the abilities. Guessing is giving an answer or making judgment about something without being sure of all the facts. According to Obinne (2012), guessing is a standard test-taking strategy presented by examinees taking multiple-choice examination. It provides an opportunity to have an item counted correct even when the examinee does not have sufficient knowledge of the subject matter. If test scores are based on the number of questions answered correctly, then a random guess increases the chance of a higher score. Delavar and Zahrakar (2013) stated that

guessing is a strategy employed by examinees to earn more marks. The way an item is framed exposes the item to guessing.

Guessing factor index has serious assessment implications because the more the items in the examination are prone to guessing, the more difficult it becomes to determine the true abilities of the examinees. One major effect of guessing is that it has the capacity to affect the psychometric properties of the test. Guessed answers increase the variance error of test scores and reduce the reliability of the test (Yoella, Joachim & Oded, 2002). Guessing raises measurement error. It also results in a structural error variance and as such considered as a major threat toconstruct validity. It is based on the increase in demand for precise and reliable assessment tools in educational assessment, particularly with Basic Science achievement, that has triggered research on advanced methodologies for estimating item characteristics. Recognizing the limitations of other item response theory models in capturing the complexities that are inherent in Basic Science achievement item formats, the researcher posed the question; what are the item characteristics of three different Basic Science Achievement test formats using the 3-parameter logistics model of item response theory? This study specifically determined the;

1. item difficulty indices of three different item formats of Basic Science achievement test.
2. item discrimination indices of three different item formats of Basic Science achievement test.
3. guessing parameters of three different item formats of Basic Science achievement test.

Research Questions

The researcher sought answers to the following research questions

1. What are the item difficulty indices of three different item format of Basic Science achievement test?
2. What are the discrimination indices of three different item formats of Basic Science achievement test?
3. What are the guessing parameters of three different item formats of Basic Science achievement test?

The following null hypotheses were formulated for the study and were tested at 0.05 level of significance:

Ho_1: There is no significant difference in the item difficulty indices of three item formats of Basic Science achievement test.

Ho_2: There is no significant difference in the discrimination indices of three item formats of Basic Science achievement test.

Ho_3: There is no significant difference in the guessing parameter indices of three item formats of Basic Science achievement test.

Method

The study adopted instrumentation research design. Instrumentation research design is used for development and validation of a test by establishing unquestionable reliability and validity of the test for effective, efficient, accurate and feasible measurement of a particular variable, construct, trait, attribute, concept or phenomenon in the social or behavioural science (Leavy, 2017). The design is suitable for the study because the study sought to empirically determine the validity indices of three item formats of Basic Science Achievement test. The study was carried out in Enugu State. This area was chosen because there is evidence of inconsistency in the performance of students in Basic Science over the years Enugu state is made up of 287 public schools in the six Education Zones of the state. The Population of the study was 21,215 JSS 3 students in the 287 public secondary schools in Enugu State. JSS 3 students were used for the study because they have been exposed to most of the topics contained in the Biology Achievement Test and they are fully prepared for their Basic Education Certificate Examination (BECE).

The sample of the study was 1,500 JSS 3 students, which was obtained through multi-stage sampling procedure. The first stage is the of two education zones out of the six Education Zone. The second stage was the selection of two local governments from each of the Education Zones that emerged from the first stage. In the third stage, simple random sampling technique was used to draw five schools from each of the four local governments that emerged. The total of twenty schools emerged. The JSS 3 students from the twenty schools form the sample for the study. The researcher developed three instruments (three different Basic Science Achievement Test formats) namely; Alternate Item Format (AIF), Matching Item Format (MIF) and Multiple-Choice Item Format (MIF). In other to ensure content validity of the instruments, the researcher developed a test blue print using the revised Bloom taxonomy of educational measures. For face validation, the instruments were face validated by three experts from the Department of Science Education, two from Measurement and

AJSTME, Volume. 12, Issue 2. 17-25, March, 2026. https://www.ajstme.com.ng
ISSN: 2251-0141 (Print), 2971-6233 (Online)

Evaluation Unit and one from Basic Science Unit, all from the Faculty of Education, University of Nigeria Nsukka.

The validators were given the purpose of the study, research question, hypotheses as well as the copies of the three item formats and were requested to vet the instrument in terms of clarity, simplicity of vocabulary and relevance of the items to the purpose of the study. The instruments were trial tested on a sample of thirty (30) students who are not part of the sample for the study. The students' responses were collated and subjected to Kuder-Richardson 20 (K – R 20) reliability method. Reliability coefficients of 0.96, 0.98 and 0.99 were obtained, for the alternate test format, matching test format and multiple choice test item format respectively. The researcher administered the instrument to the students in the selected schools through Direct Delivery Technique (DDT) with the help of the school Basic Science teachers and three research assistants. "The data collected using the instruments were analyzed using JMetrik statistical software with respect to the research questions and hypotheses formulated for the study. The research questions were answered using the difficulty, discrimination guessing parameter indices obtained using the three parameter logistic model of item response theory. The hypotheses were tested using analysis of variance (ANOVA) at 0.05 level of significance.

Results

Table 1: Difficulty indices of three different item formatsof Basic Science achievement test

S/N	ALT	MAT	MCQ	S/N	ALT	MAT	MCQ	S/N	ALT	MAT	MCQ
1	0.52	0.79	0.78	**28**	0.33	0.45	0.50	**55**	-0.53	-0.24	-0.24
2	0.68	0.86	0.83	**29**	1.32	0.92	0.96	**56**	-0.48	-0.25	-0.25
3	-24.23	-4.72	-2.70	**30**	-0.14	-0.03	-0.02	**57**	1.66	2.01	1.45
4	0.04	0.14	0.13	**31**	-0.33	-0.18	-0.18	**58**	64.17	42.21	26.57
5	0.62	-0.61	-0.50	**32**	0.16	0.33	0.32	**59**	1.63	1.50	1.56
6	-0.30	-0.05	-0.07	**33**	0.09	0.08	0.07	**60**	0.06	0.13	0.13
7	-0.30	-0.20	-0.20	**34**	0.27	0.06	0.13	**61**	0.52	-0.29	-0.18
8	0.27	0.31	0.30	**35**	0.81	0.65	0.54	**62**	0.68	0.82	0.32
9	-0.24	-0.16	-0.17	**36**	0.87	0.72	0.70	**63**	-24.23	0.64	0.07
10	0.18	0.36	0.35	**37**	27.55	5.97	5.97	**64**	0.04	0.57	-0.13
11	0.57	0.60	0.58	**38**	0.03	-0.04	-0.03	**65**	0.62	0.56	0.54
12	5.97	5.97	5.97	**39**	0.68	0.77	0.71	**66**	-0.30	-0.10	0.70
13	-0.25	-0.22	-0.20	**40**	-0.85	-0.77	-0.84	**67**	-0.30	0.19	5.97
14	0.62	0.50	0.42	**41**	0.28	0.40	0.25	**68**	0.27	0.45	-0.03
`15	20.19	5.97	27.23	**42**	0.50	0.47	0.41	**69**	-0.24	0.92	0.71
16	0.20	0.09	0.12	**43**	1.12	0.98	0.82	**70**	0.18	-0.03	-0.84
17	-0.69	-0.62	-0.61	**44**	0.31	0.45	0.28	**71**	0.57	-0.18	0.25
18	0.34	0.11	0.05	**45**	0.61	0.49	0.39	**72**	5.97	0.33	0.41
19	0.23	0.22	0.25	**46**	0.57	0.43	0.30	**73**	-0.25	0.08	0.82
20	1.23	1.12	1.38	**47**	0.58	0.77	0.67	**74**	0.62	0.06	0.28
21	-0.32	-0.29	-0.24	**48**	-0.35	-0.46	-0.70	**75**	20.19	0.65	0.39
22	0.79	0.82	0.84	**49**	-0.62	-0.44	-0.62	**76**	0.20	0.72	0.30
23	0.55	0.64	0.68	**50**	0.22	0.14	0.07	**77**	0.69	5.97	0.67
24	0.40	0.57	0.58	**51**	0.09	0.37	0.36	**78**	0.34	-0.04	-0.70
25	0.28	0.56	0.57	**52**	-0.83	-0.61	-0.62	**79**	0.23	0.77	-0.62
26	0.05	-0.10	-0.11	**53**	5.95	5.90	5.90	**80**	1.23	-0.77	0.07
27	0.13	0.19	0.19	**54**	0.22	0.46	0.45				

ALT = Alternate Test Items, MAT = Matching Test Items, MCQ = Multiple Choice Questions

Table 1 shows the item difficulty indices of the three different items formats of Basic science and technology achievement test. The result shows that 71 items (88.75%) of the alternate test items have good difficulty

indices because the item indices fall within the acceptable range of item difficulty. 2 items (2.5%) of the alternate test items are easy items because the indices fall below the acceptable range of item difficulty. 7 items (8.75%) of the alternate are difficult items because the indices were above the acceptable range of item difficulty.

The result in Table 1 further revealed that 73 items (91.25%) of the matching item format have good difficulty indices because the item indices fall within the acceptable range of item difficulty. 1 item (1.25%) of the matching item format is an easy item because the index fall below the acceptable range of item difficulty. 6 items (7.50%) of the matching item format are difficult items because the indices were above the acceptable range of item difficulty. Furthermore, the result also revealed that 74 items (92.50%) of the multiple choice item format have good difficulty indices because the item indices fall within the acceptable range of item difficulty. None of the items the multiple choice item format were easy items. 7 items (7.50%) of the multiple choice item format were difficult items because the indices were above the acceptable range of item difficulty.

Table 2: Analysis of variance result on the difference in the item facility indices of three item formats of Basic science achievement test

Test Item Formats	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	8.027	2	4.014	.093	.911
Within Groups	10240.594	237	43.209		
Total	10248.621	239			

Table 2 reveals that there is no significant difference in the item facility indices of the three item formats in Basic Science, F (2, 239) = 0.093, p = .911. Thus, the null hypothesis is not rejected, since p > 0.05.

Table 2: Discrimination indices of three different item formats of Basic Science achievement test

S/N	ALT	MAT	MCQ	S/N	ALT	MAT	MCQ	S/N	ALT	MAT	MCQ
1	2.24	1.72	1.54	**28**	1.14	1.53	1.47	**55**	0.80	0.96	0.96
2	2.78	2.16	2.03	**29**	0.69	0.83	0.82	**56**	1.75	1.65	1.73
3	0.03	0.09	0.16	**30**	1.60	2.05	1.88	**57**	0.26	0.24	0.27
4	1.98	1.65	1.64	**31**	1.89	1.91	1.91	**58**	0.01	0.02	0.03
5	2.70	0.70	0.69	**32**	1.84	1.83	1.93	**59**	1.27	1.68	1.54
6	2.15	2.08	2.06	**33**	1.72	1.64	1.61	**60**	1.71	1.85	1.92
7	2.19	2.04	2.07	**34**	1.54	1.87	1.63	**61**	2.24	1.59	1.91
8	2.25	1.54	1.53	**35**	1.52	2.24	1.70	**62**	2.78	2.70	1.93
9	2.34	2.29	2.29	**36**	0.83	0.86	0.89	**63**	0.03	1.44	1.61
10	1.59	1.35	1.35	**37**	0.82	1.53	1.53	**64**	1.98	2.47	1.63
11	2.25	1.69	1.63	**38**	1.69	1.82	1.84	**65**	2.70	2.17	1.70
12	1.56	1.53	1.54	**39**	2.66	1.99	1.89	**66**	2.15	1.83	0.89
13	1.81	1.52	1.55	**40**	1.24	1.30	1.38	**67**	2.19	2.30	1.53
14	2.69	2.18	1.73	**41**	1.03	1.37	1.37	**68**	2.25	1.53	1.84
`15	0.82	1.64	0.82	**42**	2.75	2.59	2.64	**69**	2.34	0.83	1.89
16	1.71	1.51	1.59	**43**	1.23	1.06	0.99	**70**	1.59	2.05	1.38
17	2.24	1.78	1.76	**44**	1.62	1.29	1.24	**71**	2.25	1.91	1.37
18	2.60	2.62	2.56	**45**	1.26	1.34	1.45	**72**	1.56	1.83	2.64
19	1.81	2.03	2.11	**46**	1.57	1.37	1.38	**73**	1.81	1.64	0.99
20	2.22	2.08	0.84	**47**	0.78	0.90	1.01	**74**	2.69	1.87	1.24
21	1.70	1.59	1.55	**48**	1.41	1.19	1.23	**75**	0.82	2.24	1.45
22	2.66	2.70	2.51	**49**	1.46	1.52	1.62	**76**	1.71	0.86	1.38
23	1.34	1.44	1.27	**50**	1.90	2.09	2.30	**77**	2.24	1.53	1.01
24	1.83	2.47	2.07	**51**	1.21	1.13	1.11	**78**	2.60	1.82	1.23

25	1.91	2.17	1.84	**52**	1.71	1.58	1.61	**79**	1.81	1.99	1.62
26	1.57	1.83	1.64	**53**	1.34	1.10	1.11	**80**	2.22	1.30	2.30
27	1.88	2.30	2.03	**54**	1.43	1.40	1.42				

ALT = Alternate Test Items, MAT = Matching Test Items, MCQ = Multiple Choice Questions

The result in Table 3 shows the discrimination indices of the three different item format of Basic science achievement test. The result revealed that 54 items (67.50%) of the alternate item formats have good discrimination indices because the indices were within the acceptable range of item discrimination whereas 26 items (32.50%) of the alternate item formats do not have good item discrimination indices because the indices were above the acceptable range of item discrimination index. The result also revealed that 58 items (72.50%) of the matching item formats have good discrimination indices because the indices were within the acceptable range of item discrimination whereas 22 items (27.5%) of the matching item formats do not have good item discrimination indices because the indices were above the acceptable range of item discrimination index.

Furthermore, the result it Table 3 also revealed that 67 items (83.75%) of the multiple choice item formats have good discrimination indices because the indices were within the acceptable range of item discrimination whereas 13 items (16.25%) of the multiple choice item formats do not have good item discrimination indices because the indices were above the acceptable range of item discrimination index.

Table 4: Analysis of variance on the difference in the discrimination indices of three item formats of Basic Science achievement test

Test Item Formats	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	1.449	2	.725	2.163	.117
Within Groups	79.385	237	.335		
Total	80.834	239			

The result in Table 4 shows the result for the difference in the item discrimination indices ofthree item formatsof Basic Science achievement test. The result revealed that difference in the item discrimination indices of the three item formatsin Basic science achievement test yielded an F ratio of 2.163 with an associated probability value of .117. Since the associated probability level of .117 was greater than the 0.05 level of significance, the null hypothesis which states that there is no significant difference in the item discrimination indices of thethree item formatsof Basic science achievement test was not rejected.

Table 5: Guessing parameters of three different item formats of Basic Science and achievement test

S/N	ALT	MAT	MCQ	S/N	ALT	MAT	MCQ	S/N	ALT	MAT	MCQ
1	0.21	0.24	0.23	**28**	0.02	0.03	0.05	**55**	0.03	0.03	0.04
2	0.62	0.60	0.60	**29**	0.03	0.02	0.03	**56**	0.02	0.03	0.03
3	0.23	0.25	0.26	**30**	0.08	0.03	0.04	**57**	0.22	0.19	0.17
4	0.02	0.04	0.04	**31**	0.02	0.03	0.04	**58**	0.25	0.25	0.25
5	0.69	0.50	0.43	**32**	0.03	0.02	0.03	**59**	0.05	0.09	0.09
6	0.02	0.03	0.03	**33**	0.09	0.02	0.03	**60**	0.02	0.04	0.05
7	0.02	0.02	0.03	**34**	0.05	0.36	0.03	**61**	0.21	0.03	0.04
8	0.05	0.02	0.03	**35**	0.05	0.42	0.39	**62**	0.62	0.57	0.03
9	0.03	0.03	0.03	**36**	0.02	0.02	0.03	**63**	0.23	0.04	0.03
10	0.02	0.03	0.03	**37**	0.11	0.18	0.18	**64**	0.02	0.11	0.30
11	0.03	0.04	0.04	**38**	0.02	0.02	0.03	**65**	0.69	0.06	0.39
12	0.10	0.09	0.09	**39**	0.51	0.50	0.50	**66**	0.02	0.02	0.03
13	0.02	0.03	0.05	**40**	0.18	0.12	0.07	**67**	0.02	0.02	0.18
14	0.43	0.37	0.34	**41**	0.25	0.29	0.25	**68**	0.05	0.03	0.03
`15	0.16	0.16	0.16	**42**	0.34	0.33	0.31	**69**	0.03	0.02	0.50

16	0.11	0.10	0.12	**43**	0.28	0.15	0.11	**70**	0.02	0.03	0.07
17	0.02	0.05	0.06	**44**	0.19	0.13	0.08	**71**	0.03	0.03	0.25
18	0.67	0.58	0.57	**45**	0.31	0.24	0.21	**72**	0.10	0.02	0.31
19	0.02	0.05	0.07	**46**	0.23	0.17	0.12	**73**	0.02	0.02	0.11
20	0.53	0.51	0.50	**47**	0.04	0.04	0.03	**74**	0.43	0.36	0.08
21	0.03	0.03	0.07	**48**	0.34	0.30	0.20	**75**	0.16	0.42	0.21
22	0.59	0.57	0.57	**49**	0.18	0.17	0.05	**76**	0.11	0.02	0.12
23	0.02	0.04	0.04	**50**	0.31	0.26	0.23	**77**	0.02	0.18	0.03
24	0.04	0.11	0.11	**51**	0.06	0.02	0.03	**78**	0.67	0.02	0.20
25	0.04	0.06	0.06	**52**	0.05	0.03	0.03	**79**	0.02	0.50	0.05
26	0.03	0.02	0.03	**53**	0.11	0.12	0.12	**80**	0.53	0.12	0.23
27	0.01	0.02	0.03	**54**	0.02	0.03	0.04				

ALT = Alternate Test Items, MAT = Matching Test Items, MCQ = Multiple Choice Questions

The result in Table 5 shows the guessing parameter indices of three different item formatsof Basic science and technology achievement test. The result revealed that all the 80 items (100%) of the alternate item formats have good guessing parameter indices because the indices were within the acceptable range of guessing parameter index. All the 80 items (100%) of the matching item formats have good discrimination indices because the indices were within the acceptable range of guessing parameter index. Furthermore, the result also revealed that all the 80 items (100%) of the multiple choice item formats have good guessing parameter indices because the indices were within the acceptable range of guessing parameter index.

Table 6: Analysis of variance on the difference in the guessing parameter indices of three item formats in Basic Science achievement test

Test Item Formats	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.018	2	.009	.299	.742
Within Groups	7.276	237	.031		
Total	7.294	239			

The result in Table 6 shows the result for the difference in the guessing parameter indices of the three item formatsof Basic Science achievement test. The result revealed that difference in the guessing parameter indices of the three item formatsin Basic science achievement test yielded an F ratio of .299 with an associated probability value of .742. Since the associated probability level of .742 was greater than the 0.05 level of significance, the null hypothesis which states that there is no significant difference in the guessing parameter indices of the three item formatsof Basic science achievement test was not rejected.

Discussion of Findings

The study showed that 71 items out of the 80 items of the alternate test items have good difficulty indices, 73 items out of the 80 items of the matching item format have good difficulty indices and 74 items out of the 80 items of the multiple choice item format have good difficulty indices. The result further revealed that there is no significant difference in the item facility indices of the three item formats in Basic Science achievement test. It is possible that the result of the three different item formats could have been like this due to suspected ambiguity in language which must have led to multiple interpretations of the item stem or options, making it difficult for examinees to discern the intended meaning. This is because complex language requires additional cognitive effort to parse and understand which can distract examinees from focusing on the content being assessed. Ambiguous or complex wordings may led examinees to focus on irrelevant information or misconstrue the key concept being tested, leading to incorrect responses.

The study is in agreement with the findings of Obinne (2008)that revealed that in 2000 and 2001 (NECO); 82% of biology test items were difficult, in 2001, while 75% items were difficult in 2002. For WAEC; 34% of biology test items were difficult in both 2000 and 2001 but 36% of the items are difficult in

2002. The study concludes that NECO Biology items are more difficult than the WAEC biology items.The finding of the study is in tandem with the report of Agah (2013) that studied the relative efficiency of test scores equating methods in the comparison of students' continuous assessment measures. The difficulty indices (b parameters) for state A range from -0.16 to 1.91, for state B the b-value ranges from 0.21 to 2.00 hence the assessment measures have good difficulty indices. The study reported that the parameters were good because they fall within the same acceptable range with the present study. The finding of the study was also in agreement with the finding of Ani (2014) that revealed that 66% of the items were easy while 34% of the items were difficult following the same acceptable range of the present study.

The result showed that 54 items out of the 80 items of the alternate item formats have good discrimination indices, 58 items out of the 80 items of the matching item formats have good discrimination indices and 67 items out of the 80 items of the multiple choice item formats have good discrimination indices. The result further revealed that there is no significant difference in the item discrimination indices of the three item formats in Basic Science achievement test. It could be possible that the items of the three item formats in Basic Science achievement test are too similar to one another which may lower the discrimination indices. When items are too similar, it could be challenging for examinees to distinguish between them. This lack of discrimination can lead to confusion or difficulty in making decisions. The findings of this study is in agreement with the findings of Obinne (2008) that reported that the items have varying levels of discrimination ranging from high, moderate, low and very low with respect to the acceptable range used for the study.The findings of the study also agree with the findings of Agah (2013) that reported varying levels of discrimination within the acceptable range. The findings of the study is also in line with the report of Ani (2014) that reported that 20% of the items had very low discrimination index, 36% had low discrimination index, 40% has high while only 4% had very high discrimination index with respect to the acceptable range of item discrimination.

The study showed that all the 80 items of the alternate item formats have good guessing parameter indices, all the 80 items of the matching item formats have good discrimination indices and all the 80 items of the multiple choice item formats have good guessing parameter indices. The result further revealed that there is no significant difference in the guessing parameter indices of the three item formats in Basic Science achievement test. Items with plausible distracters especially in multiple-choice examination like the items of the three item formats in Basic Science achievement test are designed to assess not just factual recall but also deeper understanding and critical thinking skills. Plausible distracters require examinees to apply their understanding of the material to eliminate incorrect options. This process tests comprehension rather than mere recognition. Examinees who guess may select an incorrect option that seems plausible on the surface, but those who truly understand the material are more likely to discern the correct answers.

The findings of the study is in agreement with the findings of Obinne (2008) that reported that all the items fall within the same acceptable range of guessing parameter as such they are good items. In the same vein, the findings of this study is in tandem with the findings of Agah (2013) that reported that the guessing factor indices for state A ranges from 0.08 to 0.40, for state B, c-parameter ranges from 0.11 to 0.38, the study reported that the parameters were good because they fall within the acceptable range. The findings was also in consonance with the findings of Ani (2014) that reported that 90% of the items fall within the guessing values of 0.02 to 0.20 (good items) and 10% of the items fall within 0.21 to 0.30 (bad items).

Conclusion

Based on the findings, the study concluded that a good number of the items of the three different test formats have item characteristics indices, and there is no significant difference in the item characteristics indices of the three item format

Recommendations

Based on the findings of the study, the study recommends that:

1. Test developers should use the three-parameter logistic (3PL) model to carefully select or refine test items for basic science achievement tests. This involves identifying items with appropriate difficulty levels, discrimination parameters (how well items distinguish between high and low performers), and guessing parameters (likelihood of guessing correct answers).
2. Researchers should validate test items by analyzing their characteristics using the 3PL model. This ensures that items are measuring

what they are intended to measure (construct validity) and are appropriate for the intended population of test-takers.

3. The application of three-parameter logistic (3PL) model in improving the overall validity (accuracy of the test in measuring what it intends to measure) and reliability (consistency of results) of basic science achievement tests. This can lead to more dependable assessments of student knowledge and skills in basic science.
4. Training test developers, educators, and policymakers in the use and interpretation of item characteristics estimated by the 3PL model. This promotes better understanding and application of psychometric principles in educational assessment practices.

References

Adamoah, Y. K. & Acquah, J. (2016). Determinants of students' performance in the Basic Education Certificate Examination (BECE) in upper Eastern region of Ghana: A case study of KASSENA-NANKANA West District (KNWD). *American Journal of research communication*. 4 (10), 91 – 107.

De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*; London: The Guilford press.

Delavar, A., & Zahrakar, K., (2013). *Assessing and measuring in psychology and educational sciences.* Tehran; Arasbaran.

Leavy, P. (Ed.). (2017). *Handbook of arts-based research*. London: Guilford Publications

Maydeu-Olivares, A. (2015). Item diagnostics in multivariate discrete data. *Psychological Methods*, *20*(2), 276–287

Meyer, J. P., and Zhu, S. (2013). Fair and equitable measurement of student learning in MOOCs: An introduction to item response theory, scale linking, and score equating. *Research & Practice in Assessment*, *8*, 26-39.

Miller, M. D. Linn, R., & Gronlund, N. (2019). *Measurement and evaluation in teaching (10th ed)*. Upper saddle River, NJ: Meririll, Prentice hall

Mozaffer, R., Hingorjo, F. J. (2012). Analysis of one-best MCQs: The difficulty index, discrimination index and distracter efficiency. Journal of Pakistan Medical Association, 62(1), 142-146.

Nworgu, B.G. (2018). Introduction to Educational Measurement and evaluation: theory and practice (2nd ed.). Nsukka: Hallman Publisher.

Obinne, A. D. E. (2012). *Comparison of the psychometric properties of West African Examinations Council and National Examination Council test items under Item Response Theory*. Unpublished PhD thesis submitted to the department of science education, University of Nigeria, Nsukka.

Onuanuga, P. A. and Saka, A. O. (2018). Trend Analysis of students' academic performance in selected Science, Technology, Engineering and Mathematics Subjects in Ogun State Basic Education Certificate Examination (OG-BECE), Nigeria, from 2011 to 2015: A projection for five years. *International Council of Association for science Education*. 29, (2), 110 – 116.

Orluwene, G. W. & Otuata, E. A. (2017) Gender influence on the efficacy of multiple-choice, alternate response and completion objective test formats on students' achievement in economics. Department of Educational Psychology, Guidance and Counseling Faculty of Education, University of Port Harcourt, Nigeria. *European Journal of Research and Reflection in Educational Sciences, 5(1),71-89.*

Yang, F. M., & Kao, S. T. (2014). Item response theory for measurement validity. *Shanghai archives of Psychiatry*, *26*(3), 171 – 177. doi:10.3969/j.issn.1002-0829.2014.03.010

Yoella, B. M., Joachim, M., & Oded, M. F. (2002). Prospect theory analysis of guessing in multiple choice test. *Journal of behavioral decision marking, 5*(14), 21 – 34.

Ziegler, M., & Hagemann, D. (2015). Testing the unidimensionality of items. *European Journal of Psychological Assessment, 31*(4).*https://doi.org/10.1027/1015-5759/a000309*