UNIJERPS
Unizik Journal of Educational Research and Policy Studies VOL.19 (1);
January-June, 2025 https://unijerps.org
2023 Impact Factor:* 6.834;** 2022 ***Impact Factor: 5.156; ***2021 Impact Factor: 5.641;***
2020 Impact Factor: 5.355. *Indexing and Impact Factor URL: http://sjifactor.com/passport.php?id=21363*

DEVELOPMENT AND CALIBRATION OF DIAGNOSTIC CIVIC EDUCATION ENGAGEMENT SKILL TEST FOR SECONDARY SCHOOL STUDENTS USING ITEM RESPONSE THEORY

Nkechi P.M. Esomonu
Phone: +2347026422569 **Email:** npm.esomonu@unizik.edu.ng

Chiamaka MacLorita Ukaegbu
Phone: +2347041338745 **Email:** chiamakalorita4@gmail.com
&
Chukwuemeka Benedict Ikeanumba
Phone: +2347039370884 **Email:** chukwuemekaikeanumba@gmail.com
Department of Educational Foundations, Faculty of Education
Nnamdi Azikiwe University, Awka

ABSTRACT

Civic education is a vital component of junior secondary education, as it lays the foundation for students to become engaged, informed, and responsible citizens. This study constructed and calibrated Diagnostic Civic Education Engagement Skill Test (DCEEST) for junior secondary students using Item Response Theory (IRT). In carrying out the study, nine research questions were answered. The study used instrumentation research design. The population of the study was made up of 7,549 JS3 students offering Civic Education in 2023/2024 academic session in Abia State. This comprised 4256 females and 3293 males (Source: Secondary Education Management Board Headquarters, Umuahia, Umuahia, 2024). The sample for the study consisted of 2,150 JS3 Civic Education students drawn using simple random sampling technique. The initial instrument consisted of 90 multiple-choice items constructed by the researchers. The initial instrument was validated by six experts. After validation, the instrument was administered to the sampled students. The collected data were analyzed using DIMTEST statistics in DIMPACK 1.0 software, maximum likelihood estimation techniques in Mplus 7.0 software and Chi-square of goodness-of-fit in eirt software. The findings of the study showed among others that the Civic Education diagnostic test items were multidimensional as well as locally independent. The findings also showed that the 4-parameter logistics IRT model was the best fit for the Civic Education diagnostic test items. The study recommended among others that teachers should use diagnostic tests regularly to assess and monitor students' strength and weakness in Civic Education engagement skills. Also, test items developed at classroom level should have good discrimination indices, difficulty indices, guessing indices and function the same across the different genders.

Keywords: Instrumentation; Test; Diagnostic test; Item Response Theory; Civic Education.

Introduction

Civic Education is a vital component of junior secondary education, as it lays the foundation for students to become engaged, informed and responsible citizens. Moreover, Civic Education is a required component of primary, secondary and tertiary education in Nigeria (Nigerian Educational Research and Development Council, 2017). It is a prerequisite subject in junior secondary education. Civic Education also known as citizenship education or democracy education has long been considered by educators as the key framework for students to learn the basic principles of government (Barraco, 2020). It helps students in the understanding of democratic values, human rights and social responsibility. Civic Education is one of the core subjects in developing responsible and disciplined students for the society (Adeeduntan and Olatude, 2021). The subject is equipped with a lot of topics that have been designed to ensure the growth and development of every child. The National Council for Social Studies (2020) defined Civic Education as the study of citizenship and government, including the rights and responsibilities of citizens, the structure and function of government, and the political processes that shape public policy.

There is every need for the teaching and learning of Civic Education in Nigeria primary and secondary schools because Civic Education has the objective of exposing students and general public to various ways of achieving community and national development, such as right values, religious tolerance, right and duties of citizens, among other things. The benefits of teaching and learning Civic Education can be realized if the students who are taught Civic Education know and understand the basic needs and skills of the subject. These essential learning skills in Civic Education include civic engagement skill, interpersonal skill, communication skill, and critical thinking (Comber, 2005). For this study, the researcher will concentrate on civic engagement skill due to the alarming trend of student demonstrating a deficiency in civic engagement skill.

The students' deficiency in this learning skill in Civic Education was widely reported by Chief Examiners in the BECE report sheet of 2021, 2022, and 2023. The National Examination Council (NECO) chief examiners' report indicated that an area of weakness was inability of Civic Education candidate to answer correctly questions relating to civic activities, such as voting, volunteering and human rights, community participation, social problems. Also, the chief examiner report reveals that many students do not have clear understanding of the concepts in civic engagement skill (NECO, 2021, 2022, 2023). The Chief examiners' report suggested that candidates should be

encouraged to develop their skills in Civic Education engagement test before the examinations. The knowledge gap and skill deficit in Civic Education at junior secondary school certificate examination brought about the need to train students on civic engagement skill.

Civic engagement skill also known as participatory skill involves the capacity and knowledge for active participation in society, often through activities like advocacy, community organization, and policy involvement (Obiagu and Ajeps, 2022). Civic engagement skills as opined by Mihailidis and Thevenin (2013) are the abilities and knowledge that enable individuals to actively participate in their communities, influence change and contribute to democratic processes. These skills are essential for fostering democracy, social responsibility, and collective well-being. Consequently, educators and teachers need to seek a coherent instructional evaluation instrument (test) that can diagnose students' engagement skills in Civic Education. The essence of using tests and other evaluation instruments during the instructional process according to Esomonu and Erutujiro (2021) is to guide, direct and monitor students' learning and progress towards attainment of course objectives. Thus, the use of diagnostic Civic Education engagement skill test in teaching and learning process should be adopted to locate and identify students' civic education engagement weakness in the subject and proffer solutions adequately.

The quality of any test, for example diagnostic test, and the information the test generates is determined through item analysis of students' responses at any examinations. Item analysis according to Esomonu and Eleje (2017) is a process which examines students' responses to individual test items in order to assess the quality of those items and of the test as a whole. Classical Test Theory (CTT) and Item Response Theory (IRT) are the two approaches to test item analysis. One limitation with CTT approach to determine test quality through item analysis, according to Ikeanumba and Esomonu (2024) is that students with the same number of correct answers may have different response patterns, that is, correct answers on different items and, thus, may not have the same level of proficiency (ability) measured by the test. Also, the indexes of item difficulty and item discrimination depend on the group of examinees being tested and, therefore, do not adequately reflect the measurement quality of the test items (Esomonu and Eleje, 2017). The limitations of CTT approach to determine test quality through item analysis are successfully addressed in the framework of item response theory (IRT).

With regards to the use of IRT in test development process, Esomonu and Eleje (2017); Esomonu and Erutujiro (2021) and Esomonu and Okeke (2021) developed and

validated test items using IRT. Esomonu and Eleje used IRT in the development and validation of multiple-choice test in Economics while Esomonu and Erutujiro developed and validated Geography diagnostic test for senior school students using IRT. It is obvious that educational researchers have paid little attention to the use of IRT in test development, also little attention have been paid to the use of IRT in development of diagnostic tests in Civic Education.

Thus, there is lack of a diagnostic test on specific skills in Civic Education. In essence, the development and calibration of skill based diagnostic Civic Education test has not been researched on, thereby constituting an educational gap. The researchers hence deem it necessary to develop a diagnostic test on Civic Education engagement skill using item response theory (IRT). The paucity of research on diagnostic Civic Education engagement skill test to identify students' specific strengths and weaknesses in engagement skill of Civic Education necessitated this study.

The objective of this study therefore, is the development and calibration of a diagnostic Civic Education Engagement skill test (DCEEST) for secondary schools using IRT. Based on the objective stated above, the study sought answers to the following research questions:

1. What is the dimensionality of the diagnostic Civic Education engagement skill test?
2. What is the local dependence of the diagnostic Civic Education engagement skill test?
3. Which IRT model represents best fit for the diagnostic Civic Education engagement skill test?
4. What are the difficulty parameters of the items of the diagnostic Civic Education engagement skill test?
5. What are the discrimination parameters of the items of diagnostic Civic Education engagement skill test?
6. What are the guessing parameters of the test items of the diagnostic Civic Education engagement skill test?
7. What items function differently in the test with respect to gender?
8. What is the empirical reliability of diagnostic Civic Education engagement skill test?
9. What is the construct validity of the diagnostic Civic Education engagement skill test?

Research Method

The design of this study is instrumentation research design. The study was conducted in Abia State, Nigeria. The population of the study was made up of 7,549 JS3 students offering Civic Education in

2023/2024 academic session in Abia State. This comprised 4256 females and 3293 males (Source: Secondary Education Management Board Headquarters, Umuahia, Umuahia, 2024). The sample for the study consisted of 2,150 JS3 Civic Education students drawn using simple random sampling technique.

The instrument was constructed based on learning skills which Civic Education experts considered necessary for students understanding of Civic Education at the secondary school level. The skills are: knowledge and critical thinking, action and participation, attitudes and values, collaboration and networking, and technology and online engagement skills. The aforementioned skills were obtained from the Junior Secondary School Civic Education Curriculum and Chief examiners' NECO report sheets.

Items measuring the skills were generated from various sources, such as locally prepared past examinations and tests, Civic Education selection tests, standard achievement tests, textbooks, and from day-to-day experiences of Civic Education teachers. The initial instrument consisted of 90 multiple choice test items with 13 items measuring knowledge and critical thinking skill, 25 items for action and participation skill, 17 items for attitudes and values skill, 15 items for collaboration and networking skill, 20 items for technology and online engagement skill.

Based on the order of priority placed on them by the chief examiners in the BECE report sheets of 2021, 2022, and 2023 (27% of the students' weakness in Civic Education is in the area of knowledge and critical thinking skill, 24% in action and participation skill, 18% in attitudes and values skill, 19% in collaboration and networking skill, and 12% in technology and online engagement skill). In constructing the items of the instrument, a skill identification table was used. The initial instrument was presented to one expert in Measurement and Evaluation in a University as well as five subject experts at the secondary school level.

The instrument was administered to the sampled students. The difficulty level (b), the discrimination level (a), the guessing (c) parameters of each item were estimated using DIMTEST statistics in DIMPACK 1.0 software, information criteria statistics in Mplus 7.0 software, Chi-square goodness-of fit in eirt software, and maximum likelihood estimation techniques of Mplus 7.0 software. The following criteria guided the researchers in selection of final items:

1. In Principal Component Analysis (PCA) of standardized residuals, if the first eigen value is greater than 1.5, it suggests multidimensionality.
2. Items are locally dependent if the value of the residual correlations is 0.2 and above.
3. The smallest information criteria statistics in terms of Akaike information criteria (AIC) and -

2loglik value indicates better IRT model fit (Thorpe and Favia, 2012).

4. Item difficulty parameter ranges of -2 to +2 (Baker and Seock, 2017).
5. Item discrimination parameter:
   0.01-0.34: Low discrimination
   0.35-0.65: Moderate discrimination
   Above 0.65: High discrimination (Baker and Seock, 2017).
6. Item guessing parameter value above 0.4indicates significant contribution of guessing to the probability of a respondent to get an item correctly (Baker and Seock, 2017).
7. Mantel-Haenszel Chi-square above 6.63 across subgroup of testees indicates item that function differently (Meyer, 2014)
8. Empirical reliability above 0.89 indicates high reliability (Obinne, 2015)
9. Extracted communality value ≥ 0.4 indicates good construct validity of scale of the instrument.

Results and Discussion

Research Question 1: What is the dimensionality of the diagnostic Civic Education engagement skill test?

Table 1: *Dimensionality Assessment of Civic Education Diagnostic Test*

Model	Eigen Value	Variance	Proportion
F1	4.78	0.05	0.26
F2	4.01	0.04	0.22
F3	3.41	0.04	0.19
F4	3.08	0.03	0.17

The dimensionality of the Civic Education diagnostic test was assessed using the principal component analysis (PCA) of standardized residuals, and the Eigen values and the proportion of variance explained by each principal component were obtained. Table 1 indicates that the first Eigen value of the Civic Education diagnostic test is greater than 1.5, and this shows evidence of multidimensionality. This manifest multidimensionality implies that different aspects of Civic Education abilities were measured by the DCEEST items.

Research Question 2: What is the local dependence of the diagnostic Civic Education engagement skill test?

Table 2: *Local Dependence of Civic Education Diagnostic Test*

Item Pair	Index	Item Pair	Index
1-2	0.2385	47-52	0.3463
2-7	0.2889	53-54	0.3213
3-18	0.2717	54-55	0.2685
7-8	0.2005	57-58	0.2521
9-10	0.2324	57-59	0.2728
16-19	0.2029	58-59	0.3620
21-24	0.2013	58-60	0.2587

23-24	0.3383	59-60	0.3132
24-55	0.2221	81-82	0.2898
35-38	0.2404	81-83	0.2728
37-38	0.2380	82-83	0.2094
39-40	0.2580	88-89	0.2521
43-44	0.2108	88-90	0.2728
46-47	0.2627	89-90	0.2620

The assumption of local independence was investigated with the Yen's Q3 statistics. Based on this statistics, residuals for any pair of items should be uncorrelated, and generally close to zero. Residual correlations that are high indicate a violation of the local independence assumption, and this suggests that the pair of items have something more in common than the rest of the item set have in common with each other (Marais, 2013). For this study, using Yen's Q3 to screen items for local dependence, out of 4,005 residual correlations, only 28 pairs as shown in Table 2 (besides the default same item-item pair) exceeded the threshold of 0.20, indicating potential local dependence. On the other hand, 97% item residual correlations were below absolute value of 0.2. This indicates that the local independence assumption of the IRT was not violated. The implication of this is that the probability of a student getting the diagnostic Civic Education engagement skill test item correctly is unaffected by the answer given to other items in the test.

Research Question 3: Which IRT model represents best fit for the diagnostic Civic Education engagement skill test?

Table 3: *IRT Model Best fit for the Diagnostic Civic Education Engagement Skill Test*

Model	S-X2 (p<0.05)	AIC	-2Loglik
1PLM	29	133463.5	133321.4
2PLM	18	131633.5	131424.4
3PLM	17	130568.8	130518.8
4PLM	15	130256.4	130280.0

The analysis of model data fit was addressed by subjecting the models to the chi-squared item fit statistics in the Jmetrick software, and the responses to full information item factor analysis, and compared using Akaike Information Criterion (AIC), and Likelihood Ratio Test (LogLik). The essence of this analysis is to establish the best fit model that provided the information for the calibration of item parameters embedded in the test data, and this, according to Thorpe and Favia (2012) and Esomonu and Eleje (2017), is to ensure the validity of the test items. In terms of the chi-squared item fit statistics, if the p-value of an item is less than 0.05 ($p < 0.05$), the item did not fit the model, and in terms of the Akaike Information Criterion (AIC), and Likelihood Ratio Test (LogLik), the model with the smallest value has the best fit. Table 3 indicated that 29 items did not fit into the

1PLM, 18 items did not fit into the 2PLM, 17 items did not fit into the 3PLM, and 15 items did not fit into the 4PLM. The implication of this is that the 4PLM has the best fit of all the models. The Akaike information criteria also revealed that the 4PLM has the smallest information criteria, and hence, the best model fit for the data. Therefore, the 4PLM was used to calibrate the IRT-based parameters for the diagnostic Civic Education engagement skill test. This is consistent with Thorpe and Favia (2012) that the smallest AIC and -2loglik value is the best fit.

Research Question 4: What are the difficulty parameters of the items of the diagnostic Civic Education engagement skill test?

Table 4: *The Difficulty Parameters of the Diagnostic Civic Education Engagement Skill Test*

Items	B	Items	B	Items	B	Items	B	Items	b
1	-2.95	19	-0.16	37	-0.36	55	0.26	73	0.81
2	-2.97	20	-2.13	38	-0.30	56	0.03	74	0.04
3	3.93	21	-0.03	39	0.02	57	-0.08	75	0.66
4	0.95	22	3.21	40	0.23	58	-0.04	76	-0.09
5	-0.11	23	-2.74	41	-0.30	59	-0.35	77	0.17
6	0.13	24	-0.09	42	0.81	60	-1.30	78	0.17
7	-1.73	25	0.03	43	0.04	61	-0.10	79	0.21
8	0.04	26	-0.30	44	0.66	62	-0.31	80	0.55
9	-2.06	27	0.13	45	-0.09	63	-0.21	81	0.16
10	2.12	28	-0.06	46	0.17	64	-0.17	82	0.33
11	-0.14	29	0.15	47	0.17	65	0.11	83	0.51
12	-0.37	30	-0.10	48	0.21	66	-0.09	84	0.25
13	-0.12	31	-0.31	49	0.55	67	-0.07	85	0.02
14	-3.74	32	-0.21	50	0.16	68	-0.36	86	0.26
15	0.22	33	-0.17	51	0.33	69	-0.30	87	0.03
16	0.43	34	0.11	52	0.51	70	0.02	88	-0.08
17	0.14	35	-0.09	53	0.25	71	0.23	89	-0.04
18	-4.39	36	-0.07	54	0.02	72	-0.30	90	-0.35

The theoretical range of values of the item difficulty parameter based on the framework of the Item Response Theory ranges from -∞ to +∞, but the typical values range from -3 to +3. Values outside this usual typical range are rare to come by. Positive estimates of item difficulty progressively imply difficult items and negative estimates retrogressively imply easy items.

Based on the guideline provided by Baker and Seock (2017), any item whose difficulty index is between -2 to +2 is a good item. Table 4 indicates that eighty (80) items, that is, items 4, 5, 6, 7, 8, 11, 12, 13, 15, 16, 17, 19, 21, 24 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45,46 ,47, 48 , 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, and 90 have b values falling between -2 and +2 on the Logit scale, indicating moderate items based

on their difficulty level, while eight (8) items, that is, items 1, 2, 9, 10, 14, 18, 20, and 23 had values below -2 on the logit scale indicating easy items, and two (2) items, that is, items 3 and 22 had values above +2 on the logit scale indicating difficult items.

Research Question 5: What are the discrimination parameters of the items of the diagnostic Civic Education engagement skill test?

Table 5: *The Discrimination Parameters of the Diagnostic Civic Education Engagement Skill Test*

Items	A	Items	A	Items	A	Items	A	Items	A
1	1.30	19	1.26	37	1.65	55	2.47	73	0.96
2	0.68	20	0.15	38	1.50	56	1.80	74	1.61
3	2.63	21	0.62	39	1.88	57	1.61	75	0.89
4	0.67	22	1.14	40	1.14	58	0.99	76	0.81
5	1.32	23	0.32	41	1.49	59	1.07	77	2.46
6	1.08	24	0.72	42	0.96	60	0.82	78	2.58
7	0.41	25	1.00	43	1.61	61	1.25	79	2.04
8	1.01	26	1.15	44	0.89	62	1.40	80	2.63
9	1.01	27	1.07	45	0.81	63	2.74	81	1.73
10	1.36	28	1.29	46	2.46	64	2.06	82	1.60
11	1.10	29	1.41	47	2.58	65	0.13	83	2.58
12	1.08	30	1.25	48	2.04	66	1.64	84	1.66
13	0.65	31	1.40	49	2.63	67	0.32	85	2.07
14	0.26	32	1.74	50	1.73	68	1.65	86	2.47
15	1.28	33	2.06	51	1.60	69	1.50	87	1.80
16	0.75	34	1.30	52	2.58	70	1.88	88	1.61
17	1.21	35	1.64	53	1.66	71	1.14	89	0.99
18	0.20	36	2.01	54	2.07	72	1.49	90	1.07

Table 5 shows that six (6) items, that is, Items 14, 18, 20, 23, 65 and 67 are within the range of 0.00 to 0.34 indicating low discriminating power, three (3) items, that is, Items , 7, 13 and 21 are within the range of 0.35 to 0.65, indicating moderate discriminating power, while eighty one (81) items, that is, Items 1, 2, 4, 5, 6, 8, 9, 10, 11, 12, 13, 15, 16, 17, 19, 22, 24 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45,46 ,47, 48 , 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 66, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, and 90 are within the range of 0.65 and above, indicating high discriminating power.

Research Question 6: What are the guessing parameters of the items of the Diagnostic Civic Education Engagement Skill Test?

Table 6: *The Guessing Parameters of the Diagnostic Civic Education Engagement Skill Test*

Items	C	Items	C	Items	C	Items	C	Items	C
1	0.23	19	0.59	37	0.10	55	0.28	73	0.06
2	0.20	20	0.10	38	0.08	56	0.13	74	0.15
3	0.25	21	0.36	39	0.11	57	0.11	75	0.07
4	0.17	22	0.08	40	0.11	58	0.08	76	0.50
5	0.13	23	0.24	41	0.22	59	0.13	77	0.21
6	0.22	24	0.20	42	0.06	60	0.15	78	0.23
7	0.24	25	0.19	43	0.15	61	0.11	79	0.19
8	0.15	26	0.07	44	0.07	62	0.15	80	0.23
9	0.12	27	0.25	45	0.50	63	0.16	81	0.18
10	0.10	28	0.30	46	0.21	64	0.11	82	0.20
11	0.18	29	0.15	47	0.23	65	0.14	83	0.62
12	0.13	30	0.11	48	0.19	66	0.09	84	0.24
13	0.11	31	0.15	49	0.23	67	0.15	85	0.10
14	0.24	32	0.16	50	0.18	68	0.10	86	0.38
15	0.15	33	0.11	51	0.20	69	0.08	87	0.13
16	0.16	34	0.14	52	0.22	70	0.11	88	0.11
17	0.50	35	0.09	53	0.24	71	0.11	89	0.08
18	0.13	36	0.15	54	0.10	72	0.22	90	0.13

Table 6 indicates that fifty nine (59) items, that is, items 4, 5, 8, 9, 10, 11, 12, 13, 15, 16, 18, 20, 22, 25, 26, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 43, 44, 50, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 73, 74, 75, 79, 81, 85, 87, 88, 89, and 90 have c values within the range of $0 < c < 0.2$ indicating that the items are not easily guessed, twenty six (26) items, that is, items 1, 2, 3, 6, 7, 14, 21, 23, 24, 27, 28, 41, 46, 47, 49, 51, 52, 53, 55, 72, 77, 78, 80, 82, 84 and 86 have c values within the range of $0.2 < c < 0.4$ indicating that some degree of guessing might be possible, while five (5) items, that is, items 17, 19, 45, 76 and 83 have c values above 0.4 indicating that guessing plays a significant role in getting the items correctly.

Research Question 7: What items function differently in the test with respect to gender?

Table 7: *DIF Analysis of the Diagnostic Civic Education Engagement Skill Test*

Items	χ^2	p-value	ETS	Class	Items	χ^2	p-value	ETS	Class
1	0.45	0.50	0.30	A	46	1.35	0.25	0.37	A
2	1.87	0.17	0.52	A	47	5.14	0.05	-0.83	A
3	0.01	0.91	0.04	A	48	0.07	0.79	-0.08	A
4	4.02	0.05	-0.55	A	49	0.71	0.40	0.25	A
5	0.82	0.36	0.27	A	50	4.36	0.05	0.63	A
6	0.01	0.92	-0.03	A	51	2.27	0.13	-0.44	A
7	5.46	0.05	0.88	A	52	2.59	0.11	0.52	A
8	0.55	0.46	0.22	A	53	3.16	0.08	-0.53	A
9	0.03	0.86	0.05	A	54	0.60	0.44	0.25	A
10	0.65	0.42	-0.24	A	55	5.73	0.05	0.77	A
11	0.94	0.33	0.28	A	56	5.54	0.05	-0.86	A

12	0.67	0.41	-0.24	A	57	0.22	0.64	0.14	A
13	2.64	0.10	0.46	A	58	0.51	0.47	-0.21	A
14	0.14	0.70	0.12	A	59	0.57	0.45	-0.22	A
15	0.64	0.43	-0.23	A	60	0.32	0.57	-0.16	A
16	0.28	0.59	-0.15	A	61	3.45	0.06	0.55	A
17	5.00	0.05	0.76	A	62	0.00	0.96	-0.02	A
18	2.27	0.13	0.42	A	63	1.05	0.30	-0.32	A
19	0.41	0.52	-0.19	A	64	0.10	0.76	0.10	A
20	0.06	0.80	0.07	A	65	0.29	0.59	-0.16	A
21	0.02	0.88	0.05	A	66	2.48	0.12	-0.48	A
22	2,96	0.09	0.50	A	67	0.10	0.75	0.10	A
23	0.07	0.79	0.08	A	68	0.78	0.38	-0.26	A
24	3.86	0.05	0.61	A	69	1.39	0.24	0.36	A
25	1.35	0.24	-0.33	A	70	0.57	0.45	0.23	A
26	3.91	0.05	0.57	A	71	0.05	0.83	-0.06	A
27	3.44	0.06	0.56	A	72	0.00	0.95	-0.02	A
28	0.14	0.71	0.11	A	73	5.41	0.05	-0.87	A
29	3.48	0.06	0.55	A	74	0.37	0.54	-0.18	A
30	3.45	0.06	0.55	A	75	5.73	0.05	-0.88	A
31	0.00	0.96	-0.02	A	76	0.02	0.25	-0.04	A
32	1.05	0.30	-0.32	A	77	1.35	0.05	0.37	A
33	0.10	0.76	0.10	A	78	5.14	0.79	-0.83	A
34	0.29	0.59	-0.16	A	79	0.07	0.40	-0.08	A
35	2.48	0.12	-0.48	A	80	0.71	0.05	0.25	A
36	0.10	0.75	0.10	A	81	4.36	0.13	0.63	A
37	0.78	0.38	-0.26	A	82	2.27	0.11	-0.44	A
38	1.39	0.24	0.36	A	83	2.59	0.08	0.52	A
39	0.57	0.45	0.23	A	84	3.16	0.08	-0.53	A
40	0.05	0.83	-0.06	A	85	0.60	0.44	0.25	A
41	0.00	0.95	-0.02	A	86	5.73	0.05	0.77	A
42	5.41	0.05	-0.87	A	87	5.54	0.05	-0.86	A
43	0.37	0.54	-0.18	A	88	0.22	0.64	0.14	A
44	5.73	0.05	-0.88	A	89	0.51	0.47	-0.21	A
45	0.02	0.90	-0.04	A	90	0.57	0.45	-0.22	A

Table 7 shows that all the items were classified as A indicating no evidence of DIF. Additionally, none of the p-values were statistically significant and the chi-square values were all below the threshold value of 6.63, supporting the result that all the items did not function differently in terms of gender. The implication of this is that all the items functioned equally across gender groups, and there is no evidence of bias in the test items with respect to gender.

Research Question 8: What is the empirical reliability of diagnostic Civic Education engagement skill test?

Table 8: *Empirical Reliability of Diagnostic Civic Education Engagement Skill Test*

Reliability Coefficient	Estimate
Cronbach Alpha	0.94
Guttman's L2	0.94
Feldt-Gilmer	0.94
Feldt-Brennan	0.94
Raju's Beta	0.94

The reliability of the Civic Education diagnostic test was evaluated using several reliability coefficients as computed with the JMetrik software. As noted by Obinne (2015), empirical reliability above 0.89 indicates high reliability. Table 8 shows that the diagnostic Civic Education engagement skill test has high reliability. This implies that the items have good internal consistency and can reliably measure the intended construct.

Research Question 9: What is the construct validity of the diagnostic Civic Education engagement skill test?

Table 9: *Communality Values of the Diagnostic Civic Education Engagement Skill Test*

Items	C	Items	C	Items	C	Items	C	Items	C
1	.472	19	.501	37	.984	55	.949	73	.900
2	.549	20	.501	38	.922	56	.890	74	.968
3	.655	21	.504	39	.905	57	.948	75	.945
4	.450	22	.456	40	.980	58	.902	76	.949
5	.498	23	.518	41	.955	59	.943	77	.913
6	.453	24	.512	42	.900	60	.523	78	.962
7	.480	25	.416	43	.968	61	.945	79	.976
8	.520	26	.526	44	.945	62	.962	80	.924
9	.520	27	.591	45	.949	63	.955	81	.929
10	.502	28	.509	46	.913	64	.919	82	.971
11	.487	29	.447	47	.962	65	.975	83	.935
12	.406	30	.945	48	.976	66	.885	84	.968
13	.387	31	.962	49	.924	67	.924	85	.884
14	.443	32	.955	50	.929	68	.984	86	.949
15	.494	33	.919	51	.971	69	.922	87	.890
16	.555	34	.975	52	.935	70	.905	88	.948
17	.483	35	.885	53	.968	71	.980	89	.902
18	.477	36	.924	54	.884	72	.955	90	.943

The construct validity of the diagnostic Civic Education engagement skill test was assessed using the extracted communalities from Factor Analysis. Communality values represent the proportion of variance in each item explained by the underlying factors. A threshold of 0.40 was used to determine whether an item was adequately explained by the factors. Items with extracted communalities ≥ 0.40 were retained, as they contributed significantly to the measured construct, while items with communalities < 0.40 were removed as a result of insufficient contribution to the construct. Table 9 shows that only one item, that is, item 13 falls below the communality threshold.

Discussion of Findings

The result of the research question 1 of the study on the dimensionality of the diagnostic Civic Education engagement skill test revealed that the tests were multidimensional. This manifest multidimensionality implies that different aspects of Civic Education abilities were measured by the DCEEST items. This is in line with the findings of Esomonu and Eleje (2017) whose study found out that Economics diagnostic test was multidimensional. This finding agreed with that of Anyawale, Isaac-Oloniyo and Abayomi (2020) who ascertained the dimensionality of Osun State Unified Mathematics Achievement Test items, and found that they were multidimensional in nature. Findings from the work of Oguoma, Metibemu and Okoye (2016) on dimensionality assumption test on 2014 Mathematics achievement items of West African Senior Secondary Certificate Examination (WASSCE) also concluded that the test items of WASSCE mathematics were inherently multidimensional in nature, which means that more than one dimension accounted for the variation observed in examinees to the mathematics test items. The finding is also in conformity with the result of Ibrahim (2023) which revealed multidimensional indices of Mathematics West African Senior School Certificate Examination (WASSCE) from 2020 to 2022 in Dala education zone, Kano state, Nigeria. This could be as a result of the fact that the diagnostic test considered more than one construct/dimension.

The result of the research question 2 of the study on the local dependence of the diagnostic Civic Education engagement skill test revealed that out of 4,005 residual correlations, only 28 pairs representing 3% of the item residual correlations were locally dependent, while 97% item residual correlations were locally independent. The implication of this is that the probability of a student getting the diagnostic Civic Education engagement skill test item correctly is unaffected by the answer given to other items in the test. This according to Ojerinde (2013) does not mean that items do not correlate with each other, but that performance on different items is independent but conditional on the student's ability. That is, items are not related and may not have acted as a clue to one another during the testing session. This finding agreed with Ubi (2006), who assessed 800 scripts of candidates that wrote Mathematics in the University Matriculation Examination in Cross River State, Nigeria and discovered that a good number of items in the examination were locally independent.

The result of the research question 3 of the study on the IRT model best fit for the diagnostic Civic Education engagement skill test revealed that the four-parameter logistic model (4-PLM) has the best chi-squared item fit statistics, and the smallest Akaike Information Criterion (AIC), and Likelihood Ratio Test (LogLik). This is consistent with Thorpe and Favia (2012) that the smallest AIC and -2loglik value is the best fit. This finding agreed with that of Okoye and Nduka (2022) who assessed the IRT best model for the calibration of the 2020 NECO

mathematics multiple-choice questions and discovered that the 4PLM provided the best fit of all the models. This disagrees with the result of Esomonu and Erutujiro (2021) whose study found out that 3PLM provided the best fit of all the models. This could be as a result of the nature of the instrument and the construct it measured.

Findings from research question 4 of the study on item difficulty indices of the diagnostic Civic Education engagement skill test showed that 8 items which represents 8.88% of the items were easy items, 80 items which represents 88.88% of the items were moderate based on their difficulty level, while 2 items which represents 2.22% of the items were difficult items. This finding agreed with that of Galle, Alaku and Samuel (2020) who investigated the psychometric properties of English Language West African Senior School Examination in Borno State, and revealed that most of the items had moderate difficulty.

Findings from research question 5 of the study on item discrimination indices of the Civic Education diagnostic test showed that 6 items which represents 6.66% of the items had low discrimination power, 3 items which represents 3.33% of the items had moderate discrimination power, while 81 items which represents 90% of the items had high discriminating power. This finding equally agreed with that of Galle, Alaku and Samuel (2020) who investigated the psychometric properties of English Language West African Senior School Examination in Borno State, and revealed that most of the items had high discriminating power.

Findings from research question 6 of the study on guessing indices of the diagnostic Civic Education engagement skill test showed that 59 items are not easily guessed, which means the chances of the respondents getting the items correctly depend on their ability rather than on guess. 26 items indicated some degree of guessing might be possible, while 5 items indicated that guessing plays a significant role in getting the items correctly. This finding agreed with that of Jimoh, Daramola, Oladele, and Sheu, (2022) whose study on estimation of item parameter indices of NECO mathematics multiple choice test items among Nigerian students revealed that the NECO mathematics multiple test items were not vulnerable to guessing as 86.7% of the items were good in terms of guessing.

The findings of research question 7 showed that all the items did not exhibit differential item functioning based on gender. This means that all the items functioned equally across gender groups, and there is no evidence of bias in the test items with respect to gender. In a related study on sex differential item functioning in the National Business and Technical Examination Board (NABTEB) 2015 Mathematics Multiple choice test items (Dichotomous) Examination in Nigeria, Omorogiuwa and Iro-Aghedo (2023) revealed that male and female testees do not significantly function differentially in 33 items (66%). This finding disagrees with the study of Ohiri and Okoye (2023) on gender-related differential item functioning of the WAEC May/June mathematics multiple-choice items in Owerri Education Zone 1

who revealed that twelve items (24%) in 2019, fifteen items (30%) in 2020 and eleven items (22%) in 2021 respectively, functioned differentially based on gender of the students.

Conclusion

Based on the findings of the study, a total of 18 items, that is items, 1, 2, 3, 9, 10, 13, 14, 17, 18, 19, 20, 22, 23, 45, 65, 67, 76 and 83 were dropped from the 90 multiple choice items of the civics diagnostic test, and the remaining surviving 72 items will form the final instrument. It is therefore concluded that the diagnostic Civic Education engagement skill test is a valid and reliable test tool for assessing junior secondary school 2 students (JS2) in Civic Education.

Recommendations

Based on the findings of this study, the following recommendations were made.

1. Teachers should use diagnostic tests regularly to assess and monitor students' the strength and weakness in Civic Education engagement skills.
2. Test items developed at classroom level should have good discrimination indices, difficulty indices, guessing indices and function the same across the different genders.
3. The education ministry should train teachers on how to develop test items with good psychometric properties.

References

Adeduntan, R.I. and Olatunde, F. (2021). Civic Education and wealth of generating: the point of convergence. *Journal of African Social Studies* (JASS), *2*(2).

Anyanwale, M. A., Isaac-Oloniyo and Abayomi, F. R. (2020). Dimensionality assessment of binary response test items. A non-parametric approach of Bayesian Item Response Theory measurement. *International Journal of Evaluation and Research in Education, 9(2), 385-393.*

Baker, F. B. and Seock, I. (2017). *The basics of Item Response Theory*. New York: ERIC.

Barraco, M. (2020).*Civic Education: why it's important.* Retrieved from http://www.linkedin.com., on December 20, 2023.

Comber, M. K. (2005). *Civic skills and Civic Education. An empirical assessment* (Doctoral destination, University of Maryland). ProQuest Dissertations and Thesis Global.

Esomonu, N.P.M and Eleje, L.I. (2017). Diagnostic quantitative economics skill test for secondary schools: Development and validation using Item Response Theory. *Journal of Education and Practice*, *8*(22), 110-125.

Esomonu, N.P.M. and Erutujiro, G. (2021). Development and validation of geography diagnostic test using Item Response Theory. *Journal of Humanities and Social Science*, 26(11), 1-9.

Esomonu, N.P.M. and Okeke, O.J. (2021). French Language Diagnostic Writing skill Test for Junior Secondary School Students: Construction and Validation Using Item Response Theory. *International Journal of*

Education and Social Science Research.

Galle, S. A., Alaku, E. M. & Samuel, J. (2020). Analysis of psychometric properties of English language West African Senior School Certificate Examination in Borno state, Nigeria. *World Journal of Innovative Research (WJIR)*, 9(6), 01-07

Ikeanumba, C.B. and Esomonu, N.P.M. (2024). Estimating the effects of the instrumentation facets and their interactions on score dependability in examinations using generalizability theory. *International Journal of Advanced Education and Research*, 9(2), 12-15

Jimoh, M.I., Daramola, D.S., Oladele, J.I., and Sheu, A.L. (2022). Assessment of items prone to guessing in SSCE Economics multiple-choice tests among students in Kwara State, Nigeria. *Anatolian Journal of Education*, 5(1),17-28.

Meyer, J. P. (2014). *Applied measurement with jmetrik*.Do-10.4324/9780203115190. New York: Guilford

Mihailidis, P. and Thevenin, B. (2013). Media literacy as a core competency for engaged citizenship in participatory democracy. *American Behavioral Scientist, 57*(11), 1611- 1622.

Thorpe, G. L., and Favia, A. (2012). *Data analysis using Item Response Theory methodology: An introduction to selected programs and applications.* Retrieved from http://digitalcommons.lidrary.umaine .edu/psy_facpub/20

Nigerian Educational Research and Development Council (NERDC) (2017). *Civic Education curriculum for primary, secondary and tertiary education.* Abuja, Nigeria: NERDC.

Obiagu, A. N., and Ajaps., S. O. (2022). Civic engagement in Nigeria: A new sociopolitical development perspective. *International Journal of Social Science and Educational Studies*, 9(1),128-147

Obinne, A. D. E. (2015). Test item validity: Item Response Theory (IRT) perspective for Nigeria. *Research Journal in Organizational Psychology and Educational Studies*, 2(1).

Oguoma, C. C., Metibemu, M.A. and Okoye, R.O. (2016). An assessment of the dimensionality of 2014 West African secondary school examination mathematics objective test scores in Imo State, Nigeria, *African J. Theory Pract. Educ. Assessment*, 4, 18-33.

Ohiri, S.C. & Okoye, R.O. (2023). Gender-related differential item functioning of the WAEC May/June mathematics multiple-choice items in Owerri education zone 1. *International Research Journal of Modernization in Engineering Technology and Science*, 5(10), 2241- 2250.

Ojerinde, D. (2013). *Classical test theory Item Response Theory: An evaluation of the comparability of item analysis results*. A paper presented at the Institute of Education University Ibadan.

Okoye, R.O. and Nduka, S.V. (2022). Application of Item Response Theory in calibrating 2020 NECO Mathematics multiple-choice questions. *International Journal of Advanced Education and Research*, 7(1), 1-5.

Omorogiuwa, K.O. & Iro-Aghedo, E.P. (2016). Determination of differential item functioning by gender in the National Business and Technical Examinations Board 2015 Mathematics multiple choice examination. *International Journal of Education, Learning and Development*, 4(10), 25-35.