

Missing Data, Statistical Power of Likelihood-Ratio Test and Differential Item Functioning in NECO Mathematics Examination in Nigeria

Nofisat Adeola OLANIGAN And Adeyemi Alaba ADEDIWURA

Department of Educational Foundations and Counselling, Faculty of Education, Obafemi Awolowo University, Ile-Ife

DOI: <https://dx.doi.org/10.47772/IJRISS.2023.701019>

Received: 27 August 2023; Revised: 11 September 2023; Accepted: 15 September 2023; Published: 26 October 2023

ABSTRACT

The study examined the percentage of missing data by persons and items, and the effect of missing data on statistical power of likelihood-ratio test across differential item functioning magnitude. The study adopted the ex-post facto research design. The population consisted of 1,034,629 candidates that sat for the June/July 2017 NECO mathematics examination. The study sample comprised all the 194,009 students that sat for the examination in the six Southwestern states of Nigeria. Data collected was analysed using frequency count, percentage, Likelihood-ratio Test and Multiple Imputation Chained Equation and T-test respectively. Results showed that 42.2% of examinees had one or more missing responses and that all the items of the 2017 SSCE Mathematics test attracted missing responses. The result also showed that 56 of the 60 items of the NECO Mathematics functioned differentially with respect to gender and that 55 of the 56 items displaying DIF flagged non-uniform DIF. Furthermore, results showed that likelihood ratio DIF test method identifies more differential functioning items when missing responses of examinees were replaced with Multiple Imputation Chain Equation and that there is no significant difference in the power of loglikelihood ratio test in detecting DIF items under traditional method of imputing missing data and the mice method. The study concluded that missing data had no significant influence on the statistical power of likelihood-ratio test for detecting differential item functioning in mathematics examination.

Keywords: Missing Data, Statistical Power, Likelihood-Ratio Test, Differential Item Functioning

INTRODUCTION

In educational measurement, test is a crucial instrument in determining students' academic achievement. Test has become one of the most important parameters by which a society adjudges the product of her educational system. The essence of testing is to reveal the latent ability of an examinee. Test is a mechanism or instrument commonly used for evaluation to measure the cognitive abilities an individual possesses or to determine the latent abilities of examinees. Test consists of a set of questions or task to which a student or testee responds to independently and the result of which can be treated in such a way as to provide a quantitative comparison in the performance of different students (Nworgu, 2011). Since test in education can be used for different purposes such as; selection, placement, diagnostic or certification, it should therefore meet specific standards in terms of validity, reliability and usability as one of the measurement tools. Even if the reliability of the measurements acquired with a measurement tool is investigated with different method, in some cases where the desired quality (latent trait) to be measured is mixed with other qualities, the individuals in different subgroups can be affected systematically from this situation. It is known as "bias" and causes negative effect on the validity and it decreases the reliability.

Bias that occurs as a systematic variation source and affects the validity is defined as "the difference

between the probabilities of correct answer of the individual within different subgroups with the same ability level. Hence, it is necessary to match the individuals in different subgroups regarding the ability levels and to examine statistically the item parameters of these individuals. This situation is defined as the examination of whether there is Differential Item Function (DIF) in the items or not.

Differential item functioning (DIF) can therefore be understood as a lack of conditional independence between an item response and group membership (often gender, location or ethnicity) given the equal latent ability or trait (Ajeigbe & Afolabi, 2014). It is required that the items with detected DIF should be checked by the experts and whether the DIF is due to another source rather than the desired measured quality. In cases that the DIF is detected to be caused by another source than the desired measured quality, it said that the related item(s) is/are biased. In order to provide validity of the items detected biased, it can be said that it is proper for them to be revised in possible cases, and in impossible cases to be removed completely from the test, after been described as one of the important threats that affect the objectivity and validity of the measurement tools (Kristanjansson, Aylesworth, McDowell & Zumbo, 2005). Thus, scientists have developed significantly extensive methods regarding the detection of DIF. As examples of some frequently used ones of these methods are Standardization (SPD-X), Mantel-Haenszel (M-H), Logistic Regression (LR) and Likelihood Ratio Test (LRT) methods. However, the detection of DIF can be complicated by the presence of many variables like number-ratio of items with DIF, test length, DIF level, sample size, DIF structure in items, and item scoring method etc. (Camili & Shepard, 1994; Padilla, Hidalgo, Benitez & Gomez-Benito, 2012; Selvi, 2013). Another variable that can change the findings acquired by the DIF detection methods is thought to be the problem of missing data or item non-response.

Missing data can be formed in cases like, for a performance test not reaching the item due to time limitations, accidentally omitting the item or leaving it empty due to not knowing the right answer (Banks, 2015); for a scale, accidentally omitting the related item or refusal to answer due to personal reasons or omit the questions they are not comfortable with (such as in the case of attitudinal measurement). Data are missing for some test items, and or for some examinees when an examinee do not answer items in a test because of his/her inability to respond to all questions. In other words, and in the most general sense, the missing data can be considered as an information loss (Alpar, 2011). Missing data occur when an examinee either does not respond to an item or question (i.e., item non-response) or does not respond to any question at all (i.e., unit non-response).

On a psychometric measure, there are multiple possible mechanisms to explain item responses that are unanswered. For example, the design of the administration may include planned missing items, in which individuals are deliberately not presented certain items. Alternatively, an examinee may decide not to answer an item because she is unsure of the correct response, or may not respond because she finds the item to be offensive, intrusive or embarrassing. The examinee may simply run out of time before reaching the item, or skip an item with the intention of returning to answer it later – only to run out of time, or forget that he skipped it (De Ayala, 2009). It is often difficult to ascertain why item responses are missing and to determine a fair way to account for them in scoring. As a result, several techniques have been proposed to deal with missing data, but no clear consensus has emerged as to the best approach to use. Various missing data handling methods and analysis were developed for the missing data mechanisms, with different assumptions about missing data. According to Rubin, there are three types of missing data mechanisms: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR) (Little and Rubin, 2002).

In addition, subject that deal with critical thinking, theories and its application such as Mathematics is likely to have high percentage/rate of missing data (for example in Mathematics test, students tends to leave items that seems difficult to them first while attending to items that are easy to them which at the end of the day may result into item non-response or missing data). Therefore, detecting DIF of such Mathematics items can

become complicated, since, most of the statistical approaches require full data such that missing values threatens the data analysis process.

Moreover, likelihood-ratio test in DIF detection has received considerable attention in literature (Finch, 2005; Bodner, 2006; Oshima & Morris, 2008). Relative to traditional approach such as the logistic regression and Mantel-Haenszel DIF detection which requires strict assumptions and which is prone to substantial bias, likelihood-ratio test is theoretically appealing because it require weaker assumption about the cause of missing data. From a practical standpoint, this means that the technique will produce parameter estimates with less bias and greater statistical power. The statistical power also known as power of a hypothesis test is the probability that the test correctly rejects the null hypothesis. Therefore, the statistical power of likelihood-ratio test is the probability that likelihood-ratio test will yield statistically significant results and correctly reject null hypothesis. It is also the probability of likelihood-ratio test to find effect if there is an effect to be found after setting certain standards which provides a basis for rejection.

The crucial question is then, should we care about item non-response or missing data while doing a DIF analysis? The answer is yes because there is the risk of potential statistical bias associated with valid inferences of test scores and their use. There is therefore the need for an IRT statistical method such as the likelihood-ratio test which is also robust to missing data in analysing item responses to evaluate items for DIF.

Despite that Mathematics is important for every student; there seems to be performance disparities among sub-group of examinees such that many see it as one of the highest hurdles to cross in their academic life (Adedayo, 2006). Also, it deals with critical thinking, theories and its application and because it also involves a lot of arithmetic and calculations, there has always been high rate of item non-response or missing data as compared to other subjects most especially at the senior school certificate examination (SSCE).

Missing data presents various problems such as the loss of information which can cause bias in the estimation of parameters, reduce the representativeness of sample and finally reduces statistical power of a test. missing data may also lead to problems like decrease of the power of the used statistical analyses, faulty estimate of standard error, increase in Type I error rate, not being able to estimate in quality the closed properties based on observation (Hohensinn & Kubinger, 2011; Molenberghs & Kenward, 2007). Thus, missing data may significantly affect the study outcome(s) due to the loss of information, thus complicating the interpretation of data analyses.

Various methods have been developed to solve the problem of missing data and they can have profoundly different effects on estimation. Literature review has also shown numerous missing data and missing data handling methods investigations in terms of combinations of factors like, sample size, proportion of missing data and method of analysis. However, there are limited empirical research on missing data on factors like significance levels, missing data mechanisms and magnitude of DIF, as well as senior school certificate mathematics examination where missing data is present. Hence, the need to investigate the possible effects of missing data on the statistical power of likelihood-ratio test for differential item functioning in senior school certificate mathematics examination in Southwestern Nigeria; hence the study.

Objectives of the Study

The specific objectives of this study are to:

1. examine the percentage of missing data by persons and items in the senior school certificate mathematics examination;
2. assess the magnitude and nature of differential item functioning of senior school certificate mathematics examination among southwestern students with respect to sex;

3. examine the effect of missing data on statistical power of likelihood-ratio test across differential item functioning magnitude with respect to sex; and
4. determine the consistency of the power of the likelihood-ratio across significance levels and across missing data mechanisms.

Research Questions

The following research questions were raised from the specific objectives.

1. What is the percentage of missing data by persons and items in the senior school certificate mathematics examination?
2. What is the magnitude and nature of differential item functioning in the senior school certificate mathematics examination among southwestern students with respect to sex?
3. What is the effect of missing data on statistical power of likelihood-ratio test across differential item functioning magnitude with respect to sex?
4. How consistent is the power of the likelihood-ratio across significance levels and across missing data mechanism?

METHODOLOGY

The study adopted the ex-post facto research design. It was considered appropriate for the study as it enabled the researchers the use of data form of candidates' responses to 2017 NECO Mathematics which already existed and allows impact analysis to be performed on this existing data without manipulation or control.

The population consisted of 1,034,629 candidates that sat for the June/July 2017 NECO mathematics examination. The 1,034,629 candidates were made of 595,120 males and 435,251 females. North West: 244,286, North East: 168,558, North Central: 212,702, South-South: 94,934, South West: 194,009, and South East: 78,256 (National Examination Council).

The study sample comprised all the 194,009 students that sat for the examination in the six Southwestern States. From each of the six states, an intact class of students (Oyo 52,353, Ekiti 11,426, Ogun 25,196, Osun 26,086, Lagos 52,407, and Ondo 26,541) who sat for the 2017 NECO Senior School Certificate Mathematics Examination was selected purposively because the data were readily available and not too large to be managed.

The research instrument used for the study was the secondary data that comprised records of candidates' responses and scoring contained in the scanned Optical Marks Record (OMR) sheets of the National Examination Council (NECO) June/July 2017 Mathematics objective items. The OMR sheets contained the responses of examinees to the 2017 NECO Mathematics objective items. The examination consists of 60 items in a multiple-choice format and scored dichotomously (responses of the examinees were scored 1 for correct response and scored 0 for incorrect option). It contained five response options ranging from A – E. the minimum score for an examinee from computation is zero (0) while the maximum score is sixty (60). The data was collected from the NECO office with the help of a letter of request written from the head of Department, Educational Foundations and Counselling to NECO office. Data collected were analysed using frequency count, percentage, Likelihood-ratio, Multiple Imputation Chained Equation and T-test.

RESULTS

Research Question One: What is the percentage of missing data by persons and items in the senior school certificate mathematics examination?

Table 1 (a) and Table 1 (b) shows the percentage of students with at least one missing response and the percentage of items with at least one missing response.

Table 1 (a): Missing responses in senior school certificate 2017 mathematics examination based on the items.

Item	Number of examinees	%MR	Item	Number of examinees	%MR
IT1	2103	1.1	IT31	4343	2.2
IT2	3301	1.7	IT32	3870	2.0
IT3	3930	2.0	IT33	4398	2.3
IT4	4402	2.3	IT34	3856	2.0
IT5	3684	1.9	IT35	3634	1.9
IT6	3321	1.7	IT36	3967	2.0
IT7	3201	1.6	IT37	4644	2.4
IT8	4039	2.1	IT38	5572	2.9
IT9	3862	2.0	IT39	4162	2.1
IT10	3898	2.0	IT40	5066	2.6
IT11	3471	1.8	IT41	4005	2.1
IT12	4240	2.2	IT42	6785	3.5
IT13	5187	2.7	IT43	5178	2.7
IT14	4546	2.3	IT44	4703	2.4
IT15	4552	2.3	IT45	5452	2.8
IT16	4229	2.2	IT46	4431	2.3
IT17	4984	2.6	IT47	5994	3.1
IT18	3961	2.0	IT48	4837	2.5
IT19	3224	1.7	IT49	4443	2.3
IT20	3714	1.9	IT50	5756	3.0
IT21	4896	2.5	IT51	3186	1.6
IT22	3765	1.9	IT52	4221	2.2
IT23	5215	2.7	IT53	6293	3.2
IT24	4591	2.4	IT54	5793	3.0
IT25	4475	2.3	IT55	6749	3.5
IT26	2990	1.5	IT56	6965	3.6
IT27	3961	2.0	IT57	6208	3.2
IT28	3549	1.8	IT58	6450	3.3
IT29	3686	1.9	IT59	7738	4.0
IT30	4269	2.2	IT60	9656	5.0

Table 1 (a) shows the distribution of items of 2017 SSCE Mathematics test with missing responses. The table shows that all the items of the 2017 SSCE Mathematics test attracted missing responses. For example, 2103 (representing 1.1%) of the examinees that sat for the test did not respond to item 1. In fact, 5% of the examinees did not respond to item 60. The result showed that all the items attracted missing responses. The implication of the finding is that all the items on the 2017 NECO Mathematics test restricted the examinees

from displaying what they know.

Table 1(b): Percentage of missing data by persons in the senior school certificate NECO 2017 mathematics examination

Missing Data	Frequency	Percent	Missing Data	Frequency	Percent
0	111500	57.472	31	51	0.026
1	37182	19.165	32	40	0.021
2	16301	8.402	33	41	0.021
3	8465	4.363	34	41	0.021
4	5186	2.673	35	26	0.013
5	3417	1.761	36	26	0.013
6	2259	1.164	37	33	0.017
7	1669	0.860	38	24	0.012
8	1251	0.645	39	26	0.013
9	960	0.495	40	22	0.011
10	876	0.452	41	13	0.007
11	699	0.360	42	11	0.006
12	525	0.271	43	13	0.007
13	454	0.234	44	10	0.005
14	377	0.194	45	9	0.005
15	330	0.170	46	8	0.004
16	292	0.151	47	7	0.004
17	244	0.126	48	10	0.005
18	215	0.111	49	3	0.002
19	208	0.107	50	4	0.002
20	169	0.087	51	2	0.001
21	150	0.077	52	4	0.002
22	124	0.064	53	5	0.003
23	111	0.057	54	3	0.002
24	119	0.061	55	2	0.001
25	96	0.049	56	3	0.002
26	89	0.046	57	6	0.003
27	86	0.044	58	1	0.001
28	70	0.036	59	2	0.001
29	63	0.032	60	16	0.008
30	60	0.031	Total	194009	100

Table 1 (b) shows the distribution of the missing responses of the examinees that took 2017 NECO Mathematics test. The table shows that about 57.5% of the examinees has no missing responses, while 42.2% had one or more missing responses. The result showed that quite a large number of the examinees had missing responses. The implication of the result is that about 50% of the examinees could not

demonstrate their proficiency completely.

Research Question Two: What is the magnitude and nature of differential item functioning in the senior school certificate mathematics examination among southwestern students with respect to sex?

To answer this research question, the responses of the students to the mathematics examination was subjected to differential item functioning and the assessment of the DIF was done using Likelihood Ratio Test (LRT) method of DIF assessment with the female students as the focal group. The result is presented in Table 2

Table 2: Magnitude of Differential item functioning of 2017 NECO with respect to gender among students from South-west Nigeria

Item	Gender	logLik	G2	Df	p		Item	Gender	logLik	G2	df	p	
1	Female	-6143870	92.765	3	0.000	DIF	31	Female	-6143900	151.959	3	0.000	DIF
	Male	-6143824						Male	-6143824				
2	Female	-6143864	79.89	3	0.000	DIF	32	Female	-6143904	159.55	3	0.000	DIF
	Male	-6143824						Male	-6143824				
3	Female	-6143897	145.459	3	0.000	DIF	33	Female	-6143853	58.26	3	0.000	DIF
	Male	-6143824						Male	-6143824				
4	Female	-6143949	250.313	3	0.000	DIF	34	Female	-6143887	126.716	3	0.000	DIF
	Male	-6143824						Male	-6143824				
5	Female	-6143849	49.876	3	0.000	DIF	35	Female	-6143892	135.892	3	0.000	DIF
	Male	-6143824						Male	-6143824				
6	Female	-6143846	44.586	3	0.000	DIF	36	Female	-6143900	151.785	3	0.000	DIF
	Male	-6143824						Male	-6143824				
7	Female	-6143991	334.093	3	0.000	DIF	37	Female	-6143868	88.954	3	0.000	DIF
	Male	-6143824						Male	-6143824				
8	Female	-6143947	246.639	3	0.000	DIF	38	Female	-6143860	71.768	3	0.000	DIF
	Male	-6143824						Male	-6143824				
9	Female	-6143897	146.748	3	0.000	DIF	39	Female	-6143862	75.375	3	0.000	DIF
	Male	-6143824						Male	-6143824				
10	Female	-6144050	451.474	3	0.000	DIF	40	Female	-6143930	212.923	3	0.000	DIF
	Male	-6143824						Male	-6143824				
11	Female	-6143979	310.83	3	0.000	DIF	41	Female	-6143870	92.964	3	0.000	DIF
	Male	-6143824						Male	-6143824				
12	Female	-6143830	12.317	3	0.002	DIF	42	Female	-6143845	41.489	3	0.000	DIF
	Male	-6143824						Male	-6143824				
13	Female	-6143835	2.226	3	0.329	NO DIF	43	Female	-6143851	53.876	3	0.000	DIF
	Male	-6143836						Male	-6143824				
14	Female	-6143869	89.598	3	0.000	DIF	44	Female	-6143826	4.117	3	0.128	NO DIF
	Male	-6143824						Male	-6143824				
15	Female	-6143865	82.148	3	0.000	DIF	45	Female	-6143878	107.584	3	0.000	DIF
	Male	-6143824						Male	-6143824				

16	Female	-6143896	144.914	3	0.000	DIF	46	Female	-6143845	41.581	3	0.000	DIF
	Male	-6143824						Male	-6143824				
17	Female	-6143851	54.434	3	0.000	DIF	47	Female	-6143825	2.226	3	0.329	NO DIF
	Male	-6143824						Male	-6143824				
18	Female	-6143842	35.139	3	0.000	DIF	48	Female	-6143927	205.544	3	0.000	DIF
	Male	-6143824						Male	-6143824				
19	Female	-6143868	87.563	3	0.000	DIF	49	Female	-6143857	66.529	3	0.000	DIF
	Male	-6143824						Male	-6143824				
20	Female	-6143873	97.696	3	0.000	DIF	50	Female	-6143838	28.049	3	0.000	DIF
	Male	-6143824						Male	-6143824				
21	Female	-6143840	32.078	3	0.000	DIF	51	Female	-6143851	53.505	3	0.000	DIF
	Male	-6143824						Male	-6143824				
22	Female	-6143836	23.91	3	0.000	DIF	52	Female	-6143843	38.639	3	0.000	DIF
	Male	-6143824						Male	-6143824				
23	Female	-6143832	16.089	3	0.000	DIF	53	Female	-6143883	118.191	3	0.000	DIF
	Male	-6143824						Male	-6143824				
24	Female	-6143924	200.391	3	0.000	DIF	54	Female	-6143849	49.805	3	0.000	DIF
	Male	-6143824						Male	-6143824				
25	Female	-6143848	48.741	3	0.000	DIF	55	Female	-6143833	17.642	3	0.000	DIF
	Male	-6143824						Male	-6143824				
26	Female	-6143917	186.108	3	0.000	DIF	56	Female	-6143874	99.501	3	0.000	DIF
	Male	-6143824						Male	-6143824				
27	Female	-6143921	193.706	3	0.000	DIF	57	Female	-6143842	36.742	3	0.000	DIF
	Male	-6143824						Male	-6143824				
28	Female	-6143860	71.207	3	0.000	DIF	58	Female	-6143870	91.153	3	0.000	DIF
	Male	-6143824						Male	-6143824				
29	Female	-6143829	9.43	3	0.009	DIF	59	Female	-6143859	69.792	3	0.000	DIF
	Male	-6143824						Male	-6143824				
30	Female	-6143875	101.979	3	0.000	DIF	60	Female	-6143825	1.205	3	0.548	NO DIF
	Male	-6143824						Male	-6143824				

Table 2 showed the comparison of the function of the item parameters of 2017 NECO Mathematics test among male and female students as well as the magnitude of the variation observed in the functioning of the item parameters. The table showed that 56 of the 60 items of the test functioned differentially with respect to gender. For example, item 1 functioned differently among female and male students (Loglikelihood statistics for male = -6143824, while for female = -6143870). The likelihood ratio test showed that the difference in the functionality of the item parameters was significant (diff loglikelihood = 2, χ^2 (df=3)= 92.765, p-value<0.05). Similar result as in item 1 was obtained for items 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 45, 46, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58 and 59. The table further showed that the difference observed in the function of item 13 (diff loglikelihood = 1, χ^2 (df=3)= 2.226, p-value>0.05); item 44 (diff loglikelihood = 2, χ^2 (df=3)= 4.117, p-value>0.05); 47 (diff loglikelihood = 1, χ^2 (df=3)= 2.226, p-

value > 0.05) and 60 (diff loglikelihood = 46, χ^2 (df=3) = 1.205, p-value > 0.05) was not significant. The result showed that 56 of the 60 items of the NECO Mathematics functioned differentially with respect to gender. The implication of the result is that the NECO test measured the Mathematics proficiency of male and female students differently.

Further results showed that 55 of the 56 items displaying DIF with respect to gender flagged non-uniform DIF, while only one, item 28 flagged uniform DIF (See Appendix). That is 2017 NECO Mathematics test items mostly flagged non-uniform DIF with respect to gender. The implication of the result is that the 2017 NECO Mathematics test items functioned differentially with respect to gender differently at low ability level and high ability level; the differential functioning of the NECO test items at lower ability level is different from the differential function at higher ability level.

Research question three: What is the effect of missing data on statistical power of likelihood-ratio test across differential item functioning magnitude with respect to sex?

To answer this research question, the responses the students would have made to the items they failed to respond to were determined using the Multiple Imputation Chained Equation (MICE). The analysis was conducted using mice package of R Language and environment for statistical computing. After the missing responses were replaced with the computed values, the whole data were subjected to DIF analysis under likelihood ratio test method. The obtained difference in the likelihood ratio for male and female were compared with the difference in the likelihood value of the data when the missing responses were scored zero. The result is presented in Table 3.

Table 3: Magnitude of Differential item functioning of 2017 NECO Mathematics test items under ignored missing responses and missing responses imputation based on multiple imputation chain equation

Item	Gender	With missing value scored zero			With missing imputation base on mice			Item	Gender	With missing value scored zero			With missing imputation base on mice		
		logLik	G2	Remark	logLik	G2	Remark			logLik	G2	Remark	logLik	G2	Remark
1	Female	-6143870	92.765	DIF	-6013763	102.87	DIF	31	Female	-6143900	151.959	DIF	-6013798	172.638	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
2	Female	-6143864	79.89	DIF	-6013759	94.644	DIF	32	Female	-6143904	159.55	DIF	-6013808	194.082	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
3	Female	-6143897	145.46	DIF	-6013797	171.08	DIF	33	Female	-6143853	58.26	DIF	-6013755	86.951	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
4	Female	-6143949	250.31	DIF	-6013857	291.42	DIF	34	Female	-6143887	126.716	DIF	-6013792	161.25	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
5	Female	-6143849	49.876	DIF	-6013742	61.582	DIF	35	Female	-6143892	135.892	DIF	-6013798	174.159	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
6	Female	-6143846	44.586	DIF	-6013745	68.009	DIF	36	Female	-6143900	151.785	DIF	-6013802	181.974	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
7	Female	-6143991	334.09	DIF	-6013903	383.77	DIF	37	Female	-6143868	88.954	DIF	-6013766	108.438	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
8	Female	-6143947	246.64	DIF	-6013863	303	DIF	38	Female	-6143860	71.768	DIF	-6013766	108.689	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		

	Male	-6143824			-6013711				Male	-6143824			-6013711		
9	Female	-6143897	146.75	DIF	-6013804	185.83	DIF	39	Female	-6143862	75.375	DIF	-6013763	103.87	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
10	Female	-6144050	451.47	DIF	-6013971	520.21	DIF	40	Female	-6143930	212.923	DIF	-6013842	260.884	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
11	Female	-6143979	310.83	DIF	-6013880	337.21	DIF	41	Female	-6143870	92.964	DIF	-6013773	123.489	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
12	Female	-6143830	12.317	DIF	-6013716	10.227	DIF	42	Female	-6143845	41.489	DIF	-6013744	66.189	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
13	Female	-6143835	2.226	NO DIF	-6013720	17.497	DIF	43	Female	-6143851	53.876	DIF	-6013758	93.624	DIF
	Male	-6143836			-6013711				Male	-6143824			-6013711		
14	Female	-6143869	89.598	DIF	-6013772	121.38	DIF	44	Female	-6143826	4.117	NO DIF	-6013718	13.457	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
15	Female	-6143865	82.148	DIF	-6013768	114.18	DIF	45	Female	-6143878	107.584	DIF	-6013788	153.068	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
16	Female	-6143896	144.91	DIF	-6013799	174.97	DIF	46	Female	-6143845	41.581	DIF	-6013745	68.239	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
17	Female	-6143851	54.434	DIF	-6013753	82.653	DIF	47	Female	-6143825	2.226	NO DIF	-6013712	1.589	NO DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
18	Female	-6143842	35.139	DIF	-6013726	29.032	DIF	48	Female	-6143927	205.544	DIF	-6013841	258.803	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
19	Female	-6143868	87.563	DIF	-6013777	131.12	DIF	49	Female	-6143857	66.529	DIF	-6013761	98.658	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
20	Female	-6143873	97.696	DIF	-6013778	133.28	DIF	50	Female	-6143838	28.049	DIF	-6013736	49.028	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
21	Female	-6143840	32.078	DIF	-6013731	39.504	DIF	51	Female	-6143851	53.505	DIF	-6013747	71.426	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
22	Female	-6143836	23.91	DIF	-6013735	46.476	DIF	52	Female	-6143843	38.639	DIF	-6013746	68.711	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
23	Female	-6143832	16.089	DIF	-6013722	22.279	DIF	53	Female	-6143883	118.191	DIF	-6013798	173.085	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
24	Female	-6143924	200.39	DIF	-6013832	241.52	DIF	54	Female	-6143849	49.805	DIF	-6013756	89.39	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		
25	Female	-6143848	48.741	DIF	-6013745	67.076	DIF	55	Female	-6143833	17.642	DIF	-6013726	28.528	DIF
	Male	-6143824			-6013711				Male	-6143824			-6013711		

26	Female	-6143917	186.11	DIF	-6013814	206.27	DIF	56	Female	-6143874	99.501	DIF	-6013800	176.925	DIF	
	Male	-6143824			-6013711				Male	-6143824			-6013711			
27	Female	-6143921	193.71	DIF	-6013824	225.33	DIF	57	Female	-6143842	36.742	DIF	-6013752	81.495	DIF	
	Male	-6143824			-6013711				Male	-6143824			-6013711			
28	Female	-6143860	71.207	DIF	-6013735	48.177	DIF	58	Female	-6143870	91.153	DIF	-6013780	136.804	DIF	
	Male	-6143824			-6013711				Male	-6143824			-6013711			
29	Female	-6143829	9.43	DIF	-6013716	10.349	DIF	59	Female	-6143859	69.792	DIF	-6013764	104.924	DIF	
	Male	-6143824			-6013711				Male	-6143824			-6013711			
30	Female	-6143875	101.98	DIF	-6013778	134.29	DIF	60	Female	-6143825	1.205	NO DIF	-6013716	9.099	DIF	
	Male	-6143824			-6013711				Male	-6143824			-6013711			
Statistical Power																
											0.12			0.13		

Table 3 showed the effect of missing data on the power of likelihood ratio test method of DIF in detecting DIF in 2017 NECO Mathematics test. The Table showed that likelihood ratio DIF test method identifies more differential functioning items when missing responses of examinees were replaced with Multiple Imputation Chain Equation (MICE) than when missing value was treated traditionally (i.e., replaced with zero). The result further showed that the statistical power of likelihood-ratio test across differential item functioning magnitude with respect to sex was higher when missing responses of examinees were replaced with multiple imputation chain equation (power = 0.13) than when missing value was treated traditionally (i.e., replaced with zero) (power = 0.12). The implication of the result is that replacing missing responses of examinees with zero reduced the statistical power of likelihood-ratio test in detecting DIF items.

Research Question Four: How consistent is the power of the likelihood-ratio across significance levels and across missing data mechanism?

To answer this research question, the p-values of the items under the two missing response mechanism was compared. The result is presented as follows

Table 4: Paired sampled t-test of the p-values of loglikelihood ratio test DIF method under missing value imputation using mice and traditional method of missing data imputation

			paired diff				
	Mean	STD	Mean	STD	T	df	p-value
Mice	0.007933	0.058327	-0.014483	0.083899	-1.33717	59	0.186301
missing	0.022417	0.092343					

Table 4 showed the consistency of the power of the likelihood-ratio across significance levels and across missing data mechanism. The Table showed that the mice method of missing data imputation was more significant in the detection of DIF item (mean = 0.008, STD = 0.058) than the traditional method of missing data imputation (mean = 0.022, STD = 0.092). paired sample t-test showed that the difference observed in the different missing data mechanism was not significant (t (59) = 1.337, p-value = 0.186). The result showed that there is no significant difference in the power of loglikelihood ratio test in detecting DIF items under traditional method of imputing missing data and the mice method. The implication of the result is that the power of the likelihood-ratio across significance levels and across missing data mechanism is consistent to a large extent.

DISCUSSION OF FINDINGS

The study examined the percentage of missing data by persons and items in the senior school certificate mathematics examination. It also assessed the magnitude and nature of differential item functioning of senior school certificate mathematics examination among Southwestern students with respect to sex. Furthermore, it examined the effect of missing data on statistical power of likelihood-ratio test across differential item functioning magnitude with respect to sex. It finally determined the consistency of the power of the likelihood-ratio across significance levels and across missing data mechanism. These were with a view to examining the effect of missing data on the statistical power of likelihood-ratio test for detecting differential item functioning in senior school certificate mathematics examination among southwestern students.

Findings from research question one showed that all the items in 2017 NECO Mathematics test attracted missing responses. It also showed that quite a large number of the examinees had missing responses. The implication of the finding is that all the items on the 2017 NECO Mathematics test did restricted the examinees from displaying what they know and that half of the examinees could not demonstrate their proficiency completely. The finding is in agreement with the report by Graham (2009) that missing data has long been a challenge for researchers in a range of different fields and become a pervasive problem in virtually any discipline or examination where examinees find it difficult to respond to the items or questions presented to them. Moreover, the prevalence of missing data in education research was illustrated most clearly by Peugh and Enders (2004) who examined leading education journals published in 1999 and 2003 where they identified 389 studies that were published with missing data.

Findings from research question two showed that almost all the items of the test functioned differentially with respect to gender (i.e., their likelihood ratio test showed that the difference in the functionality of the item parameters was significant). However, only three items of the test did not function differentially with respect to gender (i.e., their likelihood ratio test showed that the difference in the functionality of the item parameters was not significant). These implied that the 2017 NECO test measures the mathematics proficiency of male and female students differently. The finding supported the findings by Abedlaziz (2010) that females showed a statistically significant and consistent advantage over males on numerical ability while males showed a consistent advantage over females on spatial and deductive ability. Moreover, the study concurred with the work of Abba (2015) showed a significant gender difference exists in English Language multiple choice items set and administered by NECO SSCE 2010. Also, Madu (2012) in his study stated that male students have greater advantage over females in Mathematics multiple choice examination. However, the study opposes that of Nwargu and Odili (2005) who stated that gender and social-economic status are not indicators of differential item functioning in 1999 WAEC SSCE. Finding further supported the study of Oladele, Adegoke and LongJohn (2020) that both WAEC and NECO mathematics tests item exhibited DIF with respect to gender under CTT and IRT frameworks. It also agreed with the findings of Adedoyin (2010), who in his study investigated gender biased items in public examinations, and found that out of 16 items that fitted the 3PL item response theory statistical analysis, 5 items were gender biased. The implication of these findings is that the DIF tendency is not specific to questions or items used by NECO alone. This also agreed with the submission of Ogunsanmi (2021) in a study on the effect of language manipulation on the differential item functioning of WAEC's Physics multiple choice items, that items functioning differentially with respect to gender or school location is not specific to questions or items used by WAEC alone, as other public examinations contain test items with similar (DIF) characteristics.

Furthermore, the nature of the differential item functioning observed in the 2017 NECO Mathematics test showed that almost all the items that are functioning differentially with respect to gender displayed a non-uniform DIF (i.e the probability of a correct item response does not differ between groups of examinees, controlling for or matching on the measured ability), while only one item, displayed uniform DIF (i.e. the probability of a correct item response differs between groups of examinees, controlling for or matching on

the measured ability). The finding showed that 2017 NECO Mathematics test items showed non-uniform DIF with respect to gender. The implication of the finding is that the 2017 NECO Mathematics test items functioned differentially with respect to gender differently at low ability level and high ability level; the differential functioning of the NECO test items at lower ability level is different from the differential function at higher ability level. These findings corroborated with the results of Adediwura and Asowo (2022) that 2017 NECO mathematics multiple-choice items reflected DIF and that not only very difficult items are susceptible to DIF but with easier items as well.

Findings from research question three on the effect of missing data on statistical power of likelihood-ratio test across differential item functioning magnitude with respect to sex, showed that likelihood ratio DIF test method identifies more differential functioning items when missing responses of examinees were replaced with multiple imputation chain equation than when missing value was treated traditionally (i.e., replaced with zero). The finding further showed that the statistical power of likelihood-ratio test across differential item functioning magnitude with respect to sex was higher when missing responses of examinees were replaced with multiple imputation chain equation. However, the statistical power of likelihood-ratio test was lower when missing value was treated traditionally (i.e., replaced with zero). The implication of the finding is that replacing missing responses of examinees with zero reduced the statistical power of likelihood-ratio test in detecting DIF items. The finding supported that of Allison (2002) and Graham (2009) that traditional method such as listwise deletion decreases the effective sample size, thereby decreasing the statistical power of the analyses. The loss of power makes it more difficult to detect relatively small (but potentially important) effects or relationships between variables. These findings corroborated the conclusion of Croninger and Douglas (2005) that newer strategies for coping with missing data yield not only accurate but more precise parameter estimates than traditional strategies do. Also, Lee and Carlin, (2010) stated in their study that modern procedures of dealing with missing data yielded no biased parameter, but rather yields appropriate standard errors and retains much of the statistical power lost with other methods.

More so, findings on the consistency of the power of likelihood-ratio test across significance levels and across missing data mechanism showed that that the multiple imputation chain equation method of missing data imputation was more significant in the detection of DIF item than the traditional method of missing data imputation. Also, there is no significant difference in the power of loglikelihood ratio test in detecting DIF items under multiple imputation chain equation method and traditional method of imputing missing data. The implication of the finding is that the power of the likelihood-ratio across significance levels and across missing data mechanism is consistent to a large extent. This is in concord with the conclusion of Cox, McIntosh, Reason, and Terenzini (2013) in a study that traditional methods (e.g., listwise deletion, pairwise deletion, mean imputation, and dummy-variable adjustments) have provided relatively simple solutions, they likely have also contributed to biased statistical estimates and misleading or false findings of statistical significance.

CONCLUSION

The study concluded that missing data had no significant influence on the statistical power of likelihood-ratio test for detecting differential item functioning in senior school certificate mathematics examination among southwestern students.

RECOMMENDATIONS

Based on the findings of the study, the following recommendations were made:

1. Test experts and developers should consider the use of likelihood-ratio test in determining differential item functioning. This approach provides an intuitive and flexible methodology for detecting DIF.
2. Examination bodies should organize training for item developers on the construction of valid, reliable and fair test especially among sub-group of examinees.

3. NECO and other public examination bodies should subject test items to DIF analysis before final administration to the examinees.
4. Modern missing data methods such as multiple imputation method should be employed in cases of missing responses because of its robustness and statistical significance.

REFERENCES

1. Adedayo, O. A. (2006). Problems of teaching and learning Mathematics in secondary schools. Paper presented at workshop on effective teaching of Mathematics LSPSSDC. Magodo, 2006.
2. Adediwura, A. A. and Asowo A. P. (2022). Examining The Nature of Item Bias on Students' Performance in National Examinations Council (NECO) Mathematics Senior School Certificate Dichotomously Scored Items in Nigeria. *International Journal of Contemporary Education* 5(1), 16-28. <https://doi.org/10.11114/ijce.v5i1.5402>
3. Adedoyin, O. O. (2010). Investigating the invariance of person parameter estimates based on classical test and item response theories. *International Journal of Education Science*, 2(2), 107-113.
4. Abedalaziz, N. (2010). A gender-related differential item functioning of mathematics test items. *The International Journal of Educational and Psychological Assessment*, 5(2), 101-116
5. Abedalaziz, N. (2010). A gender-related differential item functioning of mathematics test items. *The International Journal of Educational and Psychological Assessment*, 5(2), 101-116
6. Ajeigbe, T. O. & Afolabi, E. R. I. (2014). Assessing unidimensionality and differential item functioning in qualifying examination for senior secondary school students, Osun State, Nigeria. *World Journal of Education*, 4(4), 30-37.
7. Allison, P. D. (2002). *Missing data*. Newbury Park, CA: Sage.
8. Alpar, R. (2011). *Uygulamalı çok değişkenli istatistiksel yöntemler*. Ankara: Detay Yayıncılık.
9. Banks, K. (2015). An introduction to missing data in the context of differential item functioning. *Practical Assessment, Research & Evaluation*. 20(12).
10. Bodner, T. E. (2006). Missing data: Prevalence and reporting practices. *Psychological Reports*, 99, 675-680.
11. Camilli, G., & Shepard, L. (1994). *Methods for identifying biased test items*. London: Sage Publications Ltd.
12. Cox, B. E., Mcintosh, K, Reason, R. D., & Terenzini, P. T., (2014). Working with Missing Data in Higher Education Research: A Primer in Real-World Example. *Review of Higher Education*, 37(3), 377-402. DOI:1353/rhe.2014.0026
13. Croninger, R. G., & Douglas, K. M. (2005). Missing data and institutional research. *New directions for institutional research*, 2005(127), 33-4
14. De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press
15. Finch, H. (2005). The MIMIC model as a method for detecting DIF: Comparison with Mantel-Haenszel, SIBTEST, and the IRT Likelihood Ratio. *Applied Psychological Measurement*, 29(4), 278-295. doi: 10.1177/0146621605275728.
16. Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual review of psychology*, 60:549-576.
17. Hohensinn, C. & Kubinger K. D. (2011). On the impact of missing values on item fit and the model validity of the Rasch model. *Psychological Test and Assessment Modeling*, 53, 380-393.
18. Lee K. J., Carlin J. B. (2010). Multiple imputation for missing data: Fully conditional specification versus multivariate normal imputation. *American Journal of Epidemiology*, 171, 624-632. doi:10.1093/aje/kwp425
19. Little, R. J. A. & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley.
20. Kristanjansson E. R., Aylesworth, I. M. & Zumbo, B. D. (2005). A Comparison of four methods for detecting differential item functioning in ordered response model. *Educational and Psychological*

Measurement. 65(6), 935-953.

21. Madu, B. C (2012). Analysis of gender-related differential item functioning in mathematics multiple choice items administered by West African Examination Council (WAEC). *Journal of Education and Practice*, 3(8), 71-79.
22. Molenberghs, G., & Kenward, M.G. (2007). *Missing data in clinical studie* (1st ed.). England: John Wiley & Sons
23. Nworgu, B. G. (2011). Differential item functioning: A critical issue in regional quality assurance. Paper presented in NAERA conference.
24. Oladele, B. K., Adegoke, B. A. & LongJohn, D. A., (2020). Assessment of differential item functioning in public examinations mathematics constructed-response tests. *Journal of Positive Psychology and Counselling*, 4(221-233)
25. Oshima, T. C. & Morris, S. B. (2008). An NCME instructional module on Raju's differential functioning of items and test (DFIT). *Educational measurement: issues and practice*. 43-50.
26. Padilla, J. L., Hidalgo, J. L., Benitez, I., & Gomez-Benito, J. (2012). Comparison of three software programs for evaluating DIF by means of the Mantel-Haenszel procedure; EASY DIF, DIFAS and EZDIF, *Psicologica*, 33,135-156.
27. Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, 74, 525–556.
28. Selvi, H. (2013). *Klasik test ve madde tepki kuramlarına dayalı değişen madde fonksiyonu belirleme tekniklerinin farklı puanlama durumlarında incelenmesi*. Yayınlanmamış Doktora Tezi. Mersin Üniversitesi Eğitim Bilimleri Enstitüsü.